Activity Report 2011

Section New Results

Edition: 2012-03-22
### AUDIO, SPEECH, AND LANGUAGE PROCESSING

<table>
<thead>
<tr>
<th>Project-Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPAGE Project-Team</td>
</tr>
<tr>
<td>METISS Project-Team</td>
</tr>
<tr>
<td>PAROLE Project-Team</td>
</tr>
<tr>
<td>Sémagramme Team</td>
</tr>
<tr>
<td>TALARIS Project-Team</td>
</tr>
</tbody>
</table>

### INTERACTION AND VISUALIZATION

<table>
<thead>
<tr>
<th>Project-Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALICE Project-Team</td>
</tr>
<tr>
<td>ARTIS Project-Team</td>
</tr>
<tr>
<td>AVIZ Project-Team</td>
</tr>
<tr>
<td>EVASION Project-Team</td>
</tr>
<tr>
<td>IN-SITU Project-Team</td>
</tr>
<tr>
<td>IPARLA Project-Team</td>
</tr>
<tr>
<td>MIMETIC Team</td>
</tr>
<tr>
<td>MINT Team</td>
</tr>
<tr>
<td>REVES Project-Team</td>
</tr>
<tr>
<td>VR4I Team</td>
</tr>
</tbody>
</table>

### KNOWLEDGE AND DATA REPRESENTATION AND MANAGEMENT

<table>
<thead>
<tr>
<th>Project-Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXIS Project-Team</td>
</tr>
<tr>
<td>DAHU Project-Team</td>
</tr>
<tr>
<td>DREAM Project-Team</td>
</tr>
<tr>
<td>EDELWEISS Project-Team</td>
</tr>
<tr>
<td>EXMO Project-Team</td>
</tr>
<tr>
<td>GRAPHIK Project-Team</td>
</tr>
<tr>
<td>LEO Team</td>
</tr>
<tr>
<td>MAIA Project-Team</td>
</tr>
<tr>
<td>MOSTRARE Project-Team</td>
</tr>
<tr>
<td>ORPAILLEUR Project-Team</td>
</tr>
<tr>
<td>SMIS Project-Team</td>
</tr>
<tr>
<td>WAM Project-Team</td>
</tr>
<tr>
<td>ZENITH Team</td>
</tr>
</tbody>
</table>

### ROBOTICS

<table>
<thead>
<tr>
<th>Project-Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>AROBAS Project-Team</td>
</tr>
<tr>
<td>COPRIN Project-Team</td>
</tr>
<tr>
<td>E-MOTION Project-Team</td>
</tr>
<tr>
<td>FLOWERS Project-Team</td>
</tr>
<tr>
<td>IMARA Project-Team</td>
</tr>
<tr>
<td>LAGADIC Project-Team</td>
</tr>
</tbody>
</table>

### VISION, PERCEPTION AND MULTIMEDIA UNDERSTANDING

<table>
<thead>
<tr>
<th>Project-Team</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIANA Project-Team</td>
</tr>
<tr>
<td>IMEDIA Project-Team</td>
</tr>
</tbody>
</table>
37. LEAR Project-Team ......................................................... 258
38. MAGRIT Project-Team .................................................... 264
39. Morphee Team .............................................................. 268
40. PERCEPTION Project-Team ............................................. 274
41. PRIMA Project-Team ...................................................... 276
42. PULSAR Project-Team .................................................... 277
43. TEMICS Project-Team .................................................... 304
44. TEXMEX Project-Team ................................................... 316
45. WILLOW Project-Team ................................................... 326
ALPAGE Project-Team

6. New Results

6.1. Advances in symbolic parsing with DyALog/FRMG

Participant: Éric Villemonte de La Clergerie.

Within the team is developed a wide-coverage French meta-grammar (FRMG) and an efficient hybrid TAG/TIG parser based on the DYALOG logic programming environment [127] and on the Leff morphological and syntactic lexicon [118]. It relies on the notion of factorized grammar, themselves generated from a representation that lies at a higher level of abstraction, named Meta-Grammars [129]. At that level, linguistic generalizations can be expressed, which in turn makes it possible to transfer meta-grammars from one language to a closely related one. The hybrid TAG/TIG parser generator itself implements all kinds of parsing optimizations: lexicalization (in particular via hypertags), left-corner guiding, top/bottom feature analysis, TIG analysis (with multiple adjoining), and others.

Éric de La Clergerie has continued to improve the coverage, quality and efficiency of the French meta-grammar FRMG. On the EasyDev corpus (around 4000 sentences), parsing times have improved over 2011 from an average of 1.03s per sentence to 0.28s, coverage (in terms of sentences with full parses) has improved from 72.5% to 82.60%, and accuracy (in terms of f-measure over relations) from 64.54% to 68.28%.

A part of the accuracy gains comes from the addition of a new output format for FRMG, namely the CONLL format, allowing us to use the CONLL-based dependency version of the French Treebank (around 12K sentences) for training and evaluation. We also used new machine learning techniques to improve FRMG’s disambiguation algorithm, allowing us to combine heuristic based disambiguation rules (with manually provided weights) with more standard parsing features associated with automatically learned weights. More precisely, the idea was to study the efficiency of the disambiguation rules over the French treebank and to favor (resp. penalize) well-working (resp. bad working) rules by adjusting their weight, taking into account additional (and more standard) features. Using these techniques, on ftb6_3 test part, FRMG improved from a base accuracy of 82.31% (in terms of CONLL Labeled Attachment Score) to 84.54%. These gains resulting from a training over the French TreeBank have also been observed (with however a lesser impact) on the EasyDev corpus (using a different format and using a different evaluation metric).

6.2. Task-based evaluation of syntactic lexica: coupling FRMG with various resources

Participants: Éric Villemonte de La Clergerie, Benoît Sagot.

The FRMG symbolic parser was used for comparing the performances of various syntactic lexicons as sources of information for parsing. The idea is to convert syntactic lexica other than the Leff into the Leff’s format, i.e., turn them into Alexina lexicons, and then use the resulting lexica together with the FRMG grammar for producing several parsers. These parsers only differ by the lexical information they rely on. Preliminary results had already been obtained in 2009 [119], but were restricted to one external lexicon, namely Lexicon-Grammar tables, and only to verbal entries (other entries were gathered from the Leff when using Lexicon-Grammar-based verbal entries). However, conversion tools for other resources, such as Dicovalence [136], had already been developed, in the context of the development and improvement of the Leff. Moreover, the development of a new version of the Leff verbal entries
Task-based evaluation results have been obtained on parsing with FRMG, showing that the Lefff performs better than both Lexique-Grammaire and DICOVALENCE (after conversion to the Alexina formalism) [48], [49]. The new version of the Lefff, mentioned above, leads for now to lower results than the current version, but its results are better than with Lexique-Grammaire or DICOVALENCE data, despite a significant increase of the average amount of entries per lemma. These results are satisfying both because they show that the Lefff is a useful resource for symbolic parsing, but also because they illustrate the relevance of converting other resources into the Alexina formalism, in order to merge the valuable linguistic information they contain — as done in the last years for improving the Lefff [85], [84], [86], [111], [87], [112].

6.3. Information extraction from corpora parsed with FRMG

Participants: Yayoi Nakamura-Delloye, Rosa Stern, Éric Villemonte de La Clergerie, Benoît Sagot.

Following previous experiments, in particular in the context of the FUI-funded project Scribo that ended in 2010⁴, work has been achieved for extracting information from corpora parsed with FRMG.

In the context of the EDyLex project, we have proposed two pattern-based named entity extraction methods for ontology enrichment [36], [35]. The proposed methods are characterized by the use of entity relation patterns obtained by our unsupervised extraction method. These patterns correspond to syntactic paths that connect two named entities in dependency trees produced by FRMG. This work aims to take advantage of parsing benefits and also offers solutions for parsing disadvantage. The proposed methods are characterized by the use of entity relation patterns obtained by our unsupervised extraction method. These patterns correspond to syntactic paths that connect two named entities in dependency trees. This work aims to take advantage of parsing benefits and also offers solutions for parsing disadvantage.

We also developed a mechanism for integrating the results into a domain ontology, namely the ontology under deployment at the Agence France-Presse [37].

6.4. Advances in statistical parsing

Participants: Marie Candito, Benoît Crabbé, Djamé Seddah, Enrique Henestroza Anguiano.

6.4.1. Improving statistical dependency parsing

Alpage has provided state-of-the-art results for French statistical Parsing, adapting existing techniques for French, a richer morphological language than English, either for constituency parsing or dependency parsing. The Bonsai tool (see section 5.4) is available, that gathers preprocessing tools and models for dependency parsing French. We have innovated in the tuning of tagsets and the handling of unknown words. In the last years, Alpage has then contributed on four main points:

- conversion of the French Treebank [59] used as constituency training data into dependencies [72], the resulting treebank being used by several teams for dependency parsing;
- an original method to reduce lexical data sparseness and include coverage and robustness by replacing tokens by unsupervised word clusters or morphological clusters [69], [121], [73]; all of our morphological clustering approaches were integrated into our parsing chains; data driven lemmatization required the adaptation of a state-of-the-art part-of-speech tagger and lemmatizer (Morfette [77]) based on a data-driven joint model benefiting of the inclusion of external lexica such as the Lefff [121].
- a parser-agnostic postprocessing step, developed this year, which uses specialized models for dependency parse correction [30]: dependencies in an input parse tree are revised by selecting, for a given dependent, the best governor from within a small set of candidates, using a discriminative linear ranking model that includes a rich feature set that encodes syntactic structure in the input parse tree; the parse correction framework can correct attachments using either a generic model or specialized models tailored to difficult attachment types like coordination and pp-attachment; our

experiments have shown that parse correction, combining a generic model with specialized models for difficult attachment types, can successfully improve the quality of predicted parse trees output by several representative state-of-the-art dependency parsers for French.

- an adaptation of the above-mentioned technique of word clustering to the problem of adapting statistical parsers to different text domains [25]. We show that in order to parse texts from a different domain than the one a statistical parser is trained on (namely to parse target domain text using a parser trained on indomain treebank), word clusters computed over a bridge corpus that couples indomain an target domain raw texts do improve parsing performance on target domain, without degrade performance on indomain texts (contrary to previous domain adaptation techniques). To evaluate these experiments, we use as target domain biomedical texts. We have supervised the manual syntactic annotation of a test corpus from the biomedical domain (European Public Assessment Reports concerning the marketing authorization of medicinal products).

Besides this line of work, it should be noted that two parsing models built around Stochastic Tree Insertion Grammars are currently under investigation: experiments have been conducted on Spinal TIGs [122]. Moreover, we are still improving the TIG-based dependency parser MICA, developed in collaboration with University of Marseilles, Columbia university and AT&T [61] (see section 5.5).

6.4.2. Functional labelling

Alpage worked towards the improvement of a functional labeller to be used as a post-parsing tool on an unfolded parse forest (as outputted e.g. by the Berkeley parser in the Bonsai architecture) using CRF models of various orders thereby extending the previous maximum entropy labeller designed in the team. The use of CRFs for modelling triggered a collaboration with Isabelle Tellier and JP Prost (LIFO, Orleans). The labeller implementation has been considerably improved and the accuracy of the labeller has improved as well on correct treebank trees. However we found out that the feature engineering work outweighs the formal improvements since we were able to show that the use of higher order graphical models were not contributing significantly to improve an unstructured model. Our modest gains come mostly from feature engineering. Moreover we notice that combined with a constituent parser the labeller does not improve at all on constituent parsing output. The reason being that our current architecture for the Bonsai parser is sequential (which is unsatisfactory). Following experiments on n-best parsing outputs, we observe that the labeller can drastically improve on better parses where its input is indeed correct. This suggest investigating formulating constituent parsing and functional labelling as a joint task requiring to address serious efficiency issues. We intend to tackle the two drawbacks of our current architecture (sequential process, parse forest unfolding) by formulating constituent parsing as a joint task with functional labelling in the next few months.

6.4.3. Parsing spontaneous oral text

Alpage also got involved in parsing spontaneous oral text taken from ESTER 3 data (with overlaps) generated in the ANR ETAPE project in collaboration with A. Abeillé (LLF) with the aim of preannotating a seed for a future treebank of oral French which would considerably support work in experimental linguistics led in the Labex. He has also a collaboration set up with A. Abeillé, C. Gardent and C. Cerisara for ensuring interoperability across ongoing efforts for producing oral treebanks for French. The way to carry out the task was by using a form of preprocessing of oral text to simulate a written entry to the Bonsai parser trained on written text. In the next few months we intend to test semi-supervised learning techniques to speed up the annotation process made by the LLF lab.

6.5. Named Entity Recognition and Entity Linking

Participants: Rosa Stern, Benoît Sagot.
Identifying named entities is a widely studied issue in Natural Language Processing, because named entities are crucial targets in information extraction or retrieval tasks, but also for preparing further NLP tasks (e.g., parsing). Therefore a vast amount of work has been published that is dedicated to named entity recognition, i.e., the task of identification of named entity mentions (spans of text denoting a named entity), and sometimes types. However, real-life applications need not only identify named entity mentions, but also know which real entity they refer to; this issue is addressed in tasks such as knowledge base population with entity resolution and linking, which require an inventory of entities is required prior to those tasks in order to constitute a reference.

6.5.1. Improvements of the Aleda entity database

Within the Alexina framework, we develop since 2012 the entity database Aleda [124], aimed at constituting such a reference. Aleda was first developed for French but is under development for English. Aleda is extracted automatically from Wikipedia and Geonames. It is used among others in the SXPipe processing chain and its NP named entity recognition, as well as in the NOMOS named entity linking system.

In 2011, major efforts have been made for improving the coverage, precision and richness of the French Aleda: improvements in the tool for creating an XML almost-raw-text version of the wikipedia, new method for identifying and typing entities among wikipedia articles, based on infoboxes and wikipedia categories, richer database structure for storing more detailed information about each entity, and many other improvements. A paper about these advances has been submitted to LREC 2012.

6.5.2. Cooperation of symbolic and statistical methods for named entity recognition and typing

Named entity recognition and typing is achieved both by symbolic and probabilistic systems. We have performed an experiment [24] for making the rule-based system NP, SXPipe’s high-precision named entity recognition system developed at Alpage on AFP news corpora and which relies on the Aleda named entity database, interact with LIANE, a high-recall probabilistic system developed by Frédéric Béchet and trained on oral transcriptions from the ESTER corpus. We have shown that a probabilistic system such as LIANE can be adapted to a new type of corpus in a non-supervised way thanks to large-scale corpora automatically annotated by NP. This adaptation does not require any additional manual annotation and illustrates the complementarity between numeric and symbolic techniques for tackling linguistic tasks.

6.5.3. Nomos, a statistical named entity linking system

For information extraction from news wires, entities such as persons, locations or organizations are especially relevant in a knowledge acquisition context. Through a process of named entity recognition and entity linking applied jointly, we aim at the extraction and complete identification of these relevant entities, which are meant to enrich textual content in the form of metadata. In order to store and access extracted knowledge in a structured and coherent way, we aim at populating an ontological reference base with these metadata. We have pursued our efforts in this direction, using an approach where NLP tools have early access to Linked Data resources and thus have the ability to produce metadata integrated in the Linked Data framework. In particular, we have studied how the entity linking process in this task must deal with noisy data, as opposed to the general case where only correct entity identification is provided.

We use the symbolic named entity recognition system NP, a component of SXPipe, and use it as a mention detection module. Its output is then processed through our entity linking system, which is based on a supervised model learnt from examples of linkings. Since our named entity recognition is not deterministic, as opposed to other entity linking tasks where the gold named entity recognition results are provided, it is configured to remain ambiguous and non-deterministic, i.e., its output preserves a number of ambiguities which are usually resolved at this level. In particular, no disambiguation is made in the cases of multiple possible mentions boundaries (e.g., [Paris]+[Hilton] vs. [Paris Hilton]). In order to cope with possible false mention matches, which should be discarded as linking queries, the named entity recognition output is made more ambiguous by adding a not-an-entity alternative to each mention detection. The entity linking module’s input therefore consists in multiple possible readings of sentences. For each reading, this module must perform entity linking on every possible entity mention by selecting their most probable matching entity. Competing readings are
then ranked according to the score of entities (or sequence of entities) ranked first in each of them. The reading with no entity should also receive a score in order to be included in the ranking. The motivation for this joint task lies in the frequent necessity of accessing contextual and referential information in order to complete an accurate named entity recognition; thus the part where named entity recognition usually resolves a number of ambiguities is left for the entity linking module, which uses contextual and referential information about entities.

We have realized a first implementation of our system, as well as experiments and evaluation results. In particular, when using knowledge about entities to perform entity linking, we discuss the usefulness of domain specific knowledge and the problem of domain adaptation.

6.6. Extending wordnets

**Participants:** Benoît Sagot, Marianna Apidianaki, Valérie Hanoka.

The WOLF (see section 5.9) is a freely available, automatically created wordnet for French, the biggest drawback of which has until now been the lack of general concepts that are typically expressed with highly polysemous vocabulary that is on the one hand the most valuable for applications in human language technologies but also the most difficult to add to wordnet accurately with automatic methods on the other. In collaboration with Darja Fišer (University of Ljubljana), we have developed a self-training-like technique for acquiring a classifier that is able to assign appropriate synset ids (i.e., senses) to new words, extracted from non-disambituated multilingual sources of lexical knowledge, such as Wiktionaries and Wikipedia [39], [40]. Automatic and manual evaluation shows high coverage as well as high quality of the resulting lexico-semantic repository. Another important advantage of the approach is that it is fully automatic and language-independent and can therefore be applied to any other language still lacking a wordnet. Indeed, it was applied to Slovene as well.

Other techniques were used as well and are the basis of various submitted conference papers. They rely, among others, on morphological derivation, on graph-based representation of highly multilingual lexicons extracted from numerous wiktionaries, and on automatically induced sense clusters.

6.7. Unsupervised lexical semantics

**Participant:** Marianna Apidianaki.

6.7.1. Unsupervised word sense induction and disambiguation

Word sense induction (WSI) is the task aimed at automatically identifying the senses of words in texts, without the need for handcrafted resources or annotated data. Up till now, most WSI algorithms extract the different senses of a word ‘locally’ on a per-word basis, i.e. the different senses for each word are determined separately. In collaboration with Tim van de Cruys, at Alpage in 2010, now at University of Cambridge [19], [50], we have compared the performance of such algorithms to a new algorithm that uses a ‘global’ approach, i.e. the different senses of a particular word are determined by comparing them to, and demarcating them from, the senses of other words in a full-blown word space model. The induction step and the disambiguation step are based on the same principle: words and contexts are mapped to a limited number of topical dimensions in a latent semantic word space. The intuition is that a particular sense is associated with a particular topic, so that different senses can be discriminated through their association with particular topical dimensions; in a similar vein, a particular instance of a word can be disambiguated by determining its most important topical dimensions. We evaluated our model on the SemEval-2010 word sense induction and disambiguation task. All systems that participated in this task use a local scheme for determining the different senses of a word. We obtain state-of-the-art results.
6.7.2. Unsupervised cross-lingual lexical substitution

Cross-Lingual Lexical Substitution (CLLS) is the task that aims at providing for a target word in context several alternative substitute words in another language. The proposed sets of translations may come from external resources or be extracted from textual data. In 2011, we have introduced a new approach for this task [18], namely the use of an unsupervised cross-lingual word-sense induction method. This method identifies the senses of words by clustering their translations according to their semantic similarity. We evaluated the impact of using clustering information for CLLS on the SemEval-2010 CLLS data set. Our system performs better on the ‘out-of-ten’ measure than the systems that participated in the SemEval task.


Participants: Pierre Magistry, Benoît Sagot.

For most languages using the Latin alphabet, tokenizing a text on spaces and punctuation marks is a good approximation of a segmentation into lexical units. Although this approximation hides many difficulties, they do not compare with those arising when dealing with languages that do not use spaces, such as Mandarin Chinese. Many segmentation systems have been proposed, some of them use linguistically motivated unsupervised algorithms. However, standard evaluation practices fail to account for some properties of such systems. New results [33] have shown that a simple model, based on an entropy-based reformulation of a language-independent hypothesis put forward by Harris in 1955, allows for segmenting a corpus and extracting a lexicon from the results. Tested on the Academia Sinica Corpus, our system allows for inducing a segmentation and a lexicon with good intrinsic properties and whose characteristics are similar to those of the lexicon underlying the manually-segmented corpus. Recent unpublished work using a slightly different model have improved these results. In parallel, preliminary experiments on other languages (Hindi, Sinhalese, Tamil, French) and original visualization techniques have already led to promising results.

6.9. Computational morphology

Participants: Benoît Sagot, Géraldine Walther.

Although computational morphology has been a topic of interest for Alpage for several years now, several new research topics have received attention in 2011, often in collaboration with morphologists from the Laboratoire de Linguistique Formelle (University Paris 7).

6.9.1. Inflectional morphology

Non-canonical inflection (suppletion, deponency, heteroclisis...) is extensively studied in theoretical approaches to morphology. However, these studies often lack practical implementations associated with large-scale lexica. Yet these are precisely the requirements for objective comparative studies on the complexity of morphological descriptions. We have shown [16], [43] how the Parsli model of inflectional morphology [132], which can represent many non-canonical phenomena, as well as a formalisation and an implementation thereof can be used to evaluate the complexity of competing morphological descriptions. After illustrating the properties of the model with data about French, Latin, Italian, Persian and Sorani Kurdish verbs and about noun classes from Croatian and Slovak we have conducted experiments on the complexity of four competing descriptions of French verbal inflection. The complexity is evaluated using the information-theoretic concept of description length. We show that the new concepts introduced in the model by the Parsli model enable reducing the complexity of morphological descriptions w.r.t. both traditional or more recent models.

6.9.2. Derivational morphology

This year, in relation with the ANR project EDyLex (see section 8.2.2), work has started targeted towards the acquisition of lexical information at the level of derivational morphology, both using semi- and non-supervised techniques.
Semi-supervised techniques have been used in a work dedicated to French denominal adjectives, for which we have implemented an automatic technique based on large-scale lexicons and corpora for extracting derivation links between base nouns and derived adjectives based on the same stem [46]. The resulting derivational lexicon, which is freely available, has already been partially manually validated. Future work include a full validation and adding denominal adjectives with a suppletive base.

Unsupervised techniques have been used for extraction of derivational links that appear more systematically, although their definition is less linguistically motivated as such [51].

### 6.9.3. Morphological issues concerning loan words

Also in the context of the ANR project EDyLex (see section 8.2.2), we have carried out a preliminary study on the morphological issues raised by borrowing phenomena, concerning in particular French nouns and verbs borrowed from English [52]. Using techniques that are similar to those used on derivational morphology, we have extracted a significant amount of loan words from a large raw corpus. We have proposed a model of the borrowing phenomenon, that takes into account graphemic (spelling), phonetic and morphological variability.

### 6.10. Allophony and word segmentation in language acquisition models

**Participants:** Luc Boruta, Benoît Crabbé.

Allophonic rules are responsible for the great variety in phoneme realizations. Infants can not reliably infer abstract word representations without knowledge of their native allophonic grammar. We have explored the hypothesis that some properties of infants’ input, referred to as indicators, are correlated with allophony. First, we provide an extensive evaluation of individual indicators that rely on distributional or lexical information. This evaluation relies on a phonetically transcribed corpus, generated automatically from a phonemically transcribed English, French and Japanese child-directed corpus. As such corpora do not exist as such, we used automatically extracted allophonic grammars of various sizes leading to various granularity levels, using our own allophonic rule extraction algorithm [57]. Then, we present a first evaluation of the combination of indicators of different types, considering both logical and numerical combinations schemes [23]. Though distributional and lexical indicators are not redundant, straightforward combinations do not outperform individual indicators.

Models of the acquisition of word segmentation are typically evaluated using phonemically transcribed corpora. Accordingly, they implicitly assume that children know how to undo phonetic variation when they learn to extract words from speech. Moreover, whereas models of language acquisition should perform similarly across languages, evaluation is often limited to English samples. Using the phonetically annotated corpora described above, that cover three typologically different languages, we evaluated the performance of state-of-the-art statistical models given inputs where phonetic variation has not been reduced. We have measured segmentation robustness across different levels of segmental variation, simulating systematic allophonic variation or errors in phoneme recognition. We have shown that these models do not resist an increase in such variations and do not generalize to typologically different languages. From the perspective of early language acquisition, the results strengthen the hypothesis according to which phonological knowledge is acquired in large part before the construction of a lexicon.

### 6.11. Modelling the acquisition of syntactic categories by children

**Participant:** Benoît Crabbé.

B. Crabbé co-supervised A. Gutman for an M2 thesis (MPRI) in collaboration with A. Christophe (LSCP/ENS) in the domain of psycholinguistic modelling. The topic was concerned with modelling and implementing psychologically motivated models of language treatment and acquisition. Contrary to classical Natural Language Processing applications, the main aim was not to create engineering solutions to language related tasks, but rather to test and develop psycholinguistic theories. In this context, the study was concerned with the question of learning word categories, such as the categories of Noun and Verb. It is established experimentally that 2-year-old children can identify novel nouns and verbs. It has been suggested that this can be done
using distributional cues as well as prosodic cues. While the plain distributional hypothesis had been tested quite extensively, the importance of prosodic cues had not been addressed in a computational simulation. We provided a formulation for modelling this hypothesis using unsupervised and semi-supervised forms of bayesian learning (EM) both offline and online.

6.12. Modelling and extracting discourse structures

Participants: Laurence Danlos, Charlotte Roze.

6.12.1. Cross-lingual lexical semantics of discourse connectives

Discourse connectives are words or phrases that indicate senses holding between two spans of text. The theoretical approaches accounting for these senses, such as text coherence, cohesion, or rhetorical structure theory, share at least one common feature: they acknowledge that many connectives can indicate different senses depending on their context. Depending on its sense, the translation of a connective into another language can vary greatly, either using an equivalent connective, or using a different construction or even no explicit connective at all.

On the basis of data provided by the bilingual concordancer TransSearch which propose statistical word alignment [64], [53] made a semi-manual annotation of the English translation of two French connectives ("en effet" and "alors que"). The results of this annotation show that the translations of these connectives do not correspond to the “transpots” identified by TransSearch and even less to the translations proposed in bilingual dictionaries.

The conclusions of this work were presented at an European workshop organized by the project COMTIS\(^5\), and some members decide to use our technic for other connectives and other aligned corpora (e.g. Europarl).

6.12.2. Discourse relations inference rules

In 2011 we have developed a new methodology for building discourse relations inference rules, to be integrated into an algebra of these relations [54], [38]. The construction of such an algebra has as main objective the improvement of the comparison of discourse structures within the evaluation of discourse annotations and the creation of a gold-standard corpus. The inference rules can also help detecting inconsistencies in discourse structures, in order to improve human or machine annotation. The premises of rules already studied lead to the formulation of inference rules, established by the theoretical definition of discourse relations, manually constructed data and extracted data. By manually annotating discourses, we also compute inference probabilities. We have illustrated the adopted methodology taking as theoretical background the Segmented Discourse Representation Theory [60].

6.12.3. Discourse structure and factivity

Discursive annotations proposed in theories of discourse such as RST (Rhetorical Structure Theory) or SDRT (Segmented Representation Theory Discourse) have the advatange of building a global discourse structure linking all the information in a text. Discursive annotations proposed in PDTB (Penn Discourse Tree Bank) have the advatange of identifying the "source" of each information – thereby answering to questions such as who says or thinks what?

In collaboration with Owen Rambow (Columbia University), we have proposed [26], [28] a unified approach for discursive annotations combining the strengths of these two streams of research. This unified approach relies crucially on factivity information, as encoded in the English corpus FactBank. We intend to pursue this avenue of research by initiating in 2012 the development of a French FactBank.

6.13. Modelling and extracting temporal structures

Participants: Pascal Denis, Philippe Muller.

\(^5\) http://www.idiap.ch/project/comtis
Temporal information has been the focus of recent attention in information extraction. An elegant approach to learning temporal orderings from texts is to formulate this problem as a constraint optimization problem, which can be then given an exact solution using Integer Linear Programming. This works well for cases where the number of possible relations between temporal entities is restricted to the mere precedence relation, but becomes impractical when considering all possible interval relations.

We have proposed this year two innovations [29], inspired from work on temporal reasoning, that control this combinatorial blow-up, therefore rendering an exact ILP inference viable in the general case. First, we propose to translate the network of constraints from temporal intervals to their end-points, to handle a drastically smaller set of constraints, while preserving the same temporal information. Second, we have show that additional efficiency is gained by enforcing coherence on particular subsets of the entire temporal graphs. We evaluate these innovations through various experiments on TimeBank 1.2 using standard evaluation metrics, and compare our ILP formulations with various baselines and oracle systems.

The evaluation of temporal information extraction, i.e., the comparison of two annotations of a given text, is also a scientific challenge. This is because relations between events in a story are intrinsically interdependent and cannot be evaluated separately. A proper evaluation measure is also crucial in the context of a machine learning approach to the problem. Finding a common comparison referent at the text level is not obvious, and we have argued, in collaboration with Xavier Tannier (LIMSI), in favor of a shift from event based measures to measures on a unique textual object, a minimal underlying temporal graph, or more formally the transitive reduction of the graph of relations between event boundaries [15].


**Participants:** Emmanuel Lassalle, Pascal Denis.

Bridging descriptions are a special kind of anaphora whose interpretation requires not only identifying an antecedent, but also inferring a specific relation linking it to the anaphor. The resolution of bridging anaphora represents a very challenging task in discourse processing. It is considerably much harder than standard coreferential anaphora resolution for which shallow predictors (like distance, string matching, or morphosyntactic agreement) have been shown to be rather effective. Part of the challenge is due to an important information bottleneck. Lexical resources like WordNet are still too poor and uneven in coverage to provide a realistic solution. In turn, more recent approaches to bridging resolution have turned to web-based extraction methods. To date, the most complete and best-performing approach combines focus and lexical distance predictors using machine learning techniques [105].

We have focused on mereological bridging anaphora (that is, cases wherein the inferred relation is a part-whole relation). Moreover, we have worked on French, a language for which current lexical resources have a very low coverage. The system, presented in [32] is similar to a system developed for English [105], but it was enriched to integrate meronymic information extracted automatically from both web queries and raw text using syntactic patterns. Through various experiments on the DEDE corpus [78], we show that although still mediocre the performance of our system compare favorably to those obtained for English by the above-mentioned system. In addition, our evaluation indicates that the different meronym extraction methods have a cumulative effect, but that the text pattern-based extraction method is more robust and leads to higher accuracy than the Web-based approach.

### 6.15. Statistical models of word order in French

**Participants:** Juliette Thuilier, Benoît Crabbé.

We study the problem of choice in the ordering of French words using statistical models along the lines of [66] and [67]. This work aims at describing and model preferences in syntax, bringing additional elements to Bresnan’s thesis, according to which the syntactic competence of human beings can be largely simulated by probabilistic models. We previously investigated the relative position of attributive adjectives with respect to the noun.

---

6 An illustrative English example is is the following discourse: *The car will not move. The engine is broken.*
This year, we mainly studied the problem of the relative ordering of postverbal complements. The focus of this investigation is the relative order of direct object and indirect object of French ditransitive verbs. The first part of this work is based on corpora data that we extracted from two journalistic corpora (French Tree Bank and Est-Républicain) and a radio corpus (ESTER). These data were manually annotated and validated for semantic categories (animacy and semantic class of the ditransitive verb). Based on these data, we built statistical models showing that the relative length of complements and verbal lemmas are the most important factors, and that, differently from English or German, categories as animacy or definiteness seem to play no role in the relative ordering.

In collaboration with Anne Abeillé (Laboratoire de Linguistique Formelle, Université Paris 7), we extended our corpora study with psycholinguistic questionnaires, in order to show that statistical models are reflecting some linguistic knowledge of French speakers. The preliminary results confirm that animacy is not a relevant factor in ordering French complements.

As regards to corpus work, we are extending the database with spontaneous speech corpora (CORAL-ROM and CORPAIX) and a wider variety of verbal lemmas, in order to enhance sample representativeness and statistical modelling. In a crosslinguistic perspective, we plan to strengthen the comparison with the constraints observed in other languages such as English or German.

As can be seen from the outline above, this line of research brings us closer to cognitive sciences. We hope in the very long run that these investigations will bring new insights on the design of probabilistic parsers or generators. In NLP the framework that is closest to implementing construction grammar is Data Oriented Parsing.

6.16. Assessing the Amazon Mechanical Turk plateform

Participant: Benoît Sagot.

In collaboration with Gilles Adda and Joseph Mariani from LIMSI and with Karên Fort from INIST, we have assessed some crowdsourced microworking systems and especially Amazon Mechanical Turk, the use of which has been steadily growing in language processing in the past few years [41], [17]. According to the mainstream opinion expressed in the articles of the domain, this type of on-line working platforms allows to develop very quickly all sorts of quality language resources, for a very low price, by people doing that as a hobby or wanting some extra cash. We have demonstrated that the situation is far from being that ideal, be it from the point of view of quality, price, workers’ status or ethics and bring back to mind already existing or proposed alternatives. Our goal was threefold:

- to inform researchers, so that they can make their own choices with all the elements of the reflection in mind,
- to ask for help from funding agencies and scientific associations, and develop alternatives,
- to propose practical and organizational solutions in order to improve new language resources development, while limiting the risks of ethical and legal issues without letting go price or quality.

6.17. Finite state formalisms for Egyptian Hieroglyphic transliteration

Participant: François Barthélémy.

The task of transliterating an Egyptian Hieroglyphic text into the latin alphabet was studied [20], as a model problem to compare two finite-state formalisms: the first one is a cascade of binary transducers; the second one is a class of multitape transducers expressing simultaneous constraints, implemented using the Karamel language [62]. The two systems were compared regarding their expressivity and readability.

The first system tends to produce smaller machines, but is more tricky. On the other hand, the Karamel language provides a more abstract description of the forms, using an explicit tree structure and separating the different pieces of information on different tapes, according to semantic criteria. But the Karamel machine is much larger. Karamel is a high-level declarative formalism whereas non contextual rewrite rules are an efficient low-level language.
6. New Results

6.1. Audio and speech content processing

6.1.1. Audio motif discovery

Participants: Frédéric Bimbot, Laurence Catanese, Armando Muscariello.

This work was performed in close collaboration with Guillaume Gravier from the Texmex project-team.

As an alternative to supervised approaches for multimedia content analysis, where predefined concepts are searched for in the data, we investigate content discovery approaches where knowledge emerge from the data. Following this general philosophy, we pursued work on motif discovery in audio contents.

Audio motif discovery is the task of finding out, without any prior knowledge, all pieces of signals that repeat, eventually allowing variability. In 2011, we extended our recent work on seeded discovery to near duplicate detection and spoken document retrieval from examples. First, we proposed algorithmic speed ups for the discovery of near duplicate motifs (low variability) in large (several days long) audio streams, exploiting subsampling strategies [muscariello-cbmi-11]. Second, we investigated the use of previously proposed efficient pattern matching techniques to deal with motif variability in speech data [muscariello-icassp-11] in a different setting, that of spoken document retrieval from an audio example. We demonstrated the potential of model-free approaches for efficient spoken document retrieval on a variety of data sets, in particular in the framework of the Spoken Web Search task of the MediaEval 2011 international evaluation [muscariello-is-11, muscariello-mediaeval-11].

This work is carried out in the context of the Quaero Project.

6.1.2. Landmark-driven speech recognition

Participant: Stefan Ziegler.

This work is supervised by Guillaume Gravier and Bogdan Ludusan from the Texmex project-team.

Speech recognition is a key issue to access multimedia spoken contents. In this context, speech recognition faces several challenges among which robustness to acoustic and linguistic variability.

In 2011, we initiated research on landmark-driven speech recognition to increase robustness. The idea of this approach consists in accurately detecting in the signal landmarks corresponding to broad phonetic classes (vowels, nasals, etc.). These landmarks, which represent almost certain knowledge about the phonetic content of the signal, are then used to bias the search space in Viterbi decoding towards solutions consistent with the landmarks. We proposed a landmark detection system, which employs numerous attributes extracted from a segment based representation of speech. We use a decision tree for BPC classification, since this allows the evaluation of each BPC on its most informative attributes, selected from a large variety of attributes. Then, each segment is converted into a landmark and a probability estimate for each BPC is provided. Second, we extend a previously proposed landmark-driven decoding strategy by a more flexible implementation, which reinforces paths at the detected landmarks according to the obtained BPC probabilities. Results obtained on French broadcast news data show a relative improvement in word error rate of about 2% with respect to the baseline.

6.2. Recent results on sparse representations

The team has had a substantial activity ranging from theoretical results to algorithmic design and software contributions in the field of sparse representations, which is at the core of the FET-Open European project (FP7) SMALL (Sparse Models, Algorithms and Learning for Large-Scale Data, see Section 7.2.1) and the ANR project ECHANGE (ECHantillonnage Acoustique Nouvelle GEnération, see, Section 6.3.1).
6.2.1. A new framework for sparse representations: analysis sparse models

Participants: Rémi Gribonval, Sangnam Nam.

Main collaboration: Mike Davies (Univ. Edinburgh), Michael Elad (The Technion), Hadi Zayyani (Sharif University)

In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of few columns from a given matrix (the dictionary). An alternative analysis-based model can be envisioned, where an analysis operator multiplies the signal, leading to a cosparse outcome. Within the SMALL project, we initiated a research programme dedicated to this analysis model, in the context of a generic missing data problem (e.g., compressed sensing, inpainting, source separation, etc.). We obtained a uniqueness result for the solution of this problem, based on properties of the analysis operator and the measurement matrix. We also considered a number of pursuit algorithms for solving the missing data problem, including an L1-based and a new greedy method called GAP (Greedy Analysis Pursuit). Our simulations demonstrated the appeal of the analysis model, and the success of the pursuit techniques presented. These results have been published in international conferences [64], [63], and a journal paper is in preparation.

Our simulations demonstrated the appeal of the analysis model, and the success of the pursuit techniques presented. These results have been published in conferences [64], [91], [92] and a journal paper submitted to Applied and Computational Harmonic Analysis is under revision [103]. Other algorithms based on iterative cosparse projections [57] as well as extensions of GAP to deal with noise and structure in the cosparse representation have been developed, with applications to toy MRI reconstruction problems and acoustic source localization and reconstruction from few measurements (submitted to ICASSP 2012).

6.2.2. Theoretical results on sparse representations and dictionary learning

Participants: Rémi Gribonval, Sangnam Nam, Nancy Bertin.

Main collaboration: Karin Schnass (EPFL), Mike Davies (University of Edinburgh), Volkan Cevher (EPFL), Simon Foucart (Université Paris 5, Laboratoire Jacques-Louis Lions), Charles Soussen (Centre de recherche en automatique de Nancy (CRAN)) Jérôme Idier (Institut de Recherche en Communications et en Cybernétique de Nantes (IRCCyN)), Cédric Herzet (Equipe-projet FLUMINANCE (INRIA - CEMAGREF, Rennes)) Morten Nielsen (Department of Mathematical Sciences [Aalborg]), Gilles Puy, Pierre Vandergheynst, Yves Wiaux (EPFL) Mehrdad Yaghoobi, Rodolphe Jenatton, Francis Bach (Equipe-projet SIERRA (INRIA, Paris)) Boaz Ophir, Michael Elad (Technion) Mark D. Plumbley (Queen Mary, University of London)

Sparse recovery conditions for Orthogonal Least Squares : We pursued our investigation of conditions on an overcomplete dictionary which guarantee that certain ideal sparse decompositions can be recovered by some specific optimization principles / algorithms. This year, we extended Tropp’s analysis of Orthogonal Matching Pursuit (OMP) using the Exact Recovery Condition (ERC) to a first exact recovery analysis of Orthogonal Least Squares (OLS). We showed that when ERC is met, OLS is guaranteed to exactly recover the unknown support. Moreover, we provided a closer look at the analysis of both OMP and OLS when ERC is not fulfilled. We showed that there exist dictionaries for which some subsets are never recovered with OMP. This phenomenon, which also appears with ℓ1 minimization, does not occur for OLS. Finally, numerical experiments based on our theoretical analysis showed that none of the considered algorithms is uniformly better than the other. This work has been submitted for publication in a journal [108].

New links between the Restricted Isometry Property and nonlinear approximations : It is now well known that sparse or compressible vectors can be stably recovered from their low-dimensional projection, provided the projection matrix satisfies a Restricted Isometry Property (RIP). We establish new implications of the RIP with respect to nonlinear approximation in a Hilbert space with a redundant frame. The main ingredients of our approach are: a) Jackson and Bernstein inequalities, associated to the characterization of certain approximation spaces with interpolation spaces; b) a new proof that for overcomplete frames which satisfy a Bernstein inequality, these interpolation spaces are nothing but the collection of vectors admitting a representation in the dictionary with compressible coefficients; c) the proof that the RIP implies Bernstein
inequalities. As a result, we obtain that in most overcomplete random Gaussian dictionaries with fixed aspect ratio, just as in any orthonormal basis, the error of best \( m \)-term approximation of a vector decays at a certain rate if, and only if, the vector admits a compressible expansion in the dictionary. Yet, for mildly overcomplete dictionaries with a one-dimensional kernel, we give examples where the Bernstein inequality holds, but the same inequality fails for even the smallest perturbation of the dictionary. This work has been submitted for publication in a journal \[ 102 \].

**Performance guarantees for compressed sensing with spread spectrum techniques:** We advocate a compressed sensing strategy that consists of multiplying the signal of interest by a wide bandwidth modulation before projection onto randomly selected vectors of an orthonormal basis. Firstly, in a digital setting with random modulation, considering a whole class of sensing bases including the Fourier basis, we prove that the technique is universal in the sense that the required number of measurements for accurate recovery is optimal and independent of the sparsity basis. This universality stems from a drastic decrease of coherence between the sparsity and the sensing bases, which for a Fourier sensing basis relates to a spread of the original signal spectrum by the modulation (hence the name "spread spectrum"). The approach is also efficient as sensing matrices with fast matrix multiplication algorithms can be used, in particular in the case of Fourier measurements. Secondly, these results are confirmed by a numerical analysis of the phase transition of the \( l_1 \)-minimization problem. Finally, we show that the spread spectrum technique remains effective in an analog setting with chirp modulation for application to realistic Fourier imaging. We illustrate these findings in the context of radio interferometry and magnetic resonance imaging. This work has been presented at a conference \[ 93 \] and accepted for publication in a journal \[ 105 \].

**Dictionary learning:** An important practical problem in sparse modeling is to choose the adequate dictionary to model a class of signals or images of interest. While diverse heuristic techniques have been proposed in the literature to learn a dictionary from a collection of training samples, there are little existing results which provide an adequate mathematical understanding of the behaviour of these techniques and their ability to recover an ideal dictionary from which the training samples may have been generated.

In 2008, we initiated a pioneering work on this topic, concentrating in particular on the fundamental theoretical question of the identifiability of the learned dictionary. Within the framework of the Ph.D. of Karin Schnass, we developed an analytic approach which was published at the conference ISCCSP 2008 \[ 13 \] and allowed us to describe "geometric" conditions which guarantee that a (non overcomplete) dictionary is "locally identifiable" by \( l_1 \) minimization.

In a second step, we focused on estimating the number of sparse training samples which is typically sufficient to guarantee the identifiability (by \( l_1 \) minimization), and obtained the following result, which is somewhat surprising considering that previous studies seemed to require a combinatorial number of training samples to guarantee the identifiability: the local identifiability condition is typically satisfied as soon as the number of training samples is roughly proportional to the ambient signal dimension. The outline of the second result was published in conferences \[ 12 \], \[ 25 \]. These results have been published in the journal paper \[ 15 \].

This year we have worked on extending the results to noisy training samples with outliers. A journal paper is in preparation, and the results will be presented at a workshop at NIPS 2011.

**Analysis Operator Learning for Overcomplete Cosparse Representations:** Besides standard dictionary learning, we also considered learning in the context of the cosparse model. We consider the problem of learning a low-dimensional signal model from a collection of training samples. The mainstream approach would be to learn an overcomplete dictionary to provide good approximations of the training samples using sparse synthesis coefficients. This famous sparse model has a less well known counterpart, in analysis form, called the cosparse analysis model. In this new model, signals are characterized by their parsimony in a transformed domain using an overcomplete analysis operator. We proposed two approaches to learn an analysis operator from a training corpus, both published in the conference EUSIPCO 2011 \[ 79 \], \[ 67 \].

The first one uses a constrained optimization program based on \( L_1 \) optimization. We derive a practical learning algorithm, based on projected subgradients, and demonstrate its ability to robustly recover a ground truth analysis operator, provided the training set is of sufficient size. A local optimality condition is derived,
providing preliminary theoretical support for the well-posedness of the learning problem under appropriate conditions. Extensions to deal with noisy training samples are currently investigated, and a journal paper is in preparation.

In the second approach, analysis "atoms" are learned sequentially by identifying directions that are orthogonal to a subset of the training data. We demonstrate the effectiveness of the algorithm in three experiments, treating synthetic data and real images, showing a successful and meaningful recovery of the analysis operator.

Connections between sparse approximation and Bayesian estimation: Penalized least squares regression is often used for signal denoising and inverse problems, and is commonly interpreted in a Bayesian framework as a Maximum A Posteriori (MAP) estimator, the penalty function being the negative logarithm of the prior. For example, the widely used quadratic program (with an $\ell^2$ penalty) associated to the LASSO / Basis Pursuit Denoising is very often considered as MAP estimation under a Laplacian prior in the context of additive white Gaussian noise (AWGN) reduction.

A first result, which has been published in IEEE Transactions on Signal Processing [35], highlights the fact that, while this is one possible Bayesian interpretation, there can be other equally acceptable Bayesian interpretations. Therefore, solving a penalized least squares regression problem with penalty $\phi(x)$ need not be interpreted as assuming a prior $C \cdot \exp(-\phi(x))$ and using the MAP estimator. In particular, we showed that for any prior $P_X$, the minimum mean square error (MMSE) estimator is the solution of a penalized least square problem with some penalty $\phi(x)$, which can be interpreted as the MAP estimator with the prior $C \cdot \exp(-\phi(x))$. Vice-versa, for certain penalties $\phi(x)$, the solution of the penalized least squares problem is indeed the MMSE estimator, with a certain prior $P_X$. In general $dP_X(x) \neq C \cdot \exp(-\phi(x))dx$.

A second result, obtained in collaboration with Prof. Mike Davies and Prof. Volkan Cevher (a paper is under revision) characterizes the "compressibility" of various probability distributions with applications to underdetermined linear regression (ULR) problems and sparse modeling. We identified simple characteristics of probability distributions whose independent and identically distributed (iid) realizations are (resp. are not) compressible, i.e., that can be approximated as sparse. We prove that many priors which MAP Bayesian interpretation is sparsity inducing (such as the Laplacian distribution or Generalized Gaussian distributions with exponent $p \leq 1$), are in a way inconsistent and do not generate compressible realizations. To show this, we identify non-trivial undersampling regions in ULR settings where the simple least squares solution outperform oracle sparse estimation in data error with high probability when the data is generated from a sparsity inducing prior, such as the Laplacian distribution.

6.2.3. Wavelets on graphs

**Participant:** Rémi Gribonval.

**Main collaboration:** Pierre Vandergheynst, David Hammond (EPFL)

Within the framework of the SMALL project 7.2.1, we investigated the possibility of developing sparse representations of functions defined on graphs, by defining an extension to the traditional wavelet transform which is valid for data defined on a graph.

There are many problems where data is collected through a graph structure: scattered or non-uniform sampling, sensor networks, data on sampled manifolds or even social networks or databases. Motivated by the wealth of new potential applications of sparse representations to these problems, the partners set out a program to generalize wavelets on graphs. More precisely, we have introduced a new notion of wavelet transform for data defined on the vertices of an undirected graph. Our construction uses the spectral theory of the graph laplacian as a generalization of the classical Fourier transform. The basic ingredient of wavelets, multi-resolution, is defined in the spectral domain via operator-valued functions that can be naturally dilated. These in turn define wavelets by acting on impulses localized at any vertex. We have analyzed the localization of these wavelets in the vertex domain and showed that our multi-resolution produces functions that are indeed concentrated at will around a specified vertex. Our theory allowed us to construct an equivalent of the continuous wavelet transform but also discrete wavelet frames.
Computing the spectral decomposition can however be numerically expensive for large graphs. We have shown that, by approximating the spectrum of the wavelet generating operator with polynomial expansions, applying the forward wavelet transform and its transpose can be approximated through iterated applications of the graph Laplacian. Since in many cases the graph Laplacian is sparse, this results in a very fast algorithm. Our implementation also uses recurrence relations for computing polynomial expansions, which results in even faster algorithms. Finally, we have proved how numerical errors are precisely controlled by the properties of the desired spectral graph wavelets. Our algorithms have been implemented in a Matlab toolbox that has been released in parallel to the main theoretical article [16]. We also plan to include this toolbox in the SMALL project numerical platform.

We now foresee many applications. On one hand we will use non-local graph wavelets constructed from the set of patches in an image (or even an audio signal) to perform de-noising or in general restoration. An interesting aspect in this case, would be to understand how wavelets estimated from corrupted signals deviate from clean wavelets. In a totally different direction, we will also explore the applications of spectral graph wavelets constructed from brain connectivity graphs obtained from whole brain tractography. Our preliminary results show that graph wavelets yield a representation that is very well adapted to how the information flows in the brain along neuronal structures.

6.2.4. Algorithmic breakthrough in sparse approximation: LoCOMP

Participants: Rémi Gribonval, Frédéric Bimbot, Ronan Le Boulch.

Main collaborations: Pierre Vandergheynst (EPFL), Boris Mailhé (former team member, now with Queen Mary University, London)

Our team had already made a substantial breakthrough in 2005 when first releasing the Matching Pursuit ToolKit (MPTK, see Section 5.3) which allowed for the first time the application of the Matching Pursuit algorithm to large scale data such as hours of CD-quality audio signals. In 2008, we designed a variant of Matching Pursuit called LoCOMP (ubiquitously for LOw Complexity Orthogonal Matching Pursuit or Local Orthogonal Matching Pursuit) specifically designed for shift-invariant dictionaries. LoCOMP has been shown to achieve an approximation quality very close to that of a full Orthonormal Matching Pursuit while retaining a much lower computational complexity of the order of that of Matching Pursuit. The complexity reduction is substantial, from one day of computation to 15 minutes for a typical audio signal [20], [19]. The main effort this year has been to integrate this algorithm into MPTK to ensure its dissemination and exploitation, and a journal paper has been published [22].

6.3. Emerging activities on compressive sensing and inverse problems

6.3.1. Nearfield acoustic holography (ECHANGE ANR project)

Participants: Rémi Gribonval, Nancy Bertin.

Main collaborations: Albert Cohen (Laboratoire Jacques-Louis Lions, Université Paris 6), Laurent Daudet, Gilles Chardon, François Ollivier, Antoine Peillot (Institut Jean Le Rond d’Alembert, Université Paris 6)

Compressed sensing is a rapidly emerging field which proposes a new approach to sample data far below the Nyquist rate when the sampled data admits a sparse approximation in some appropriate dictionary. The approach is supported by many theoretical results on the identification of sparse representations in overcomplete dictionaries, but many challenges remain open to determine its range of effective applicability. METISS has chosen to focus more specifically on the exploration of Compressed Sensing of Acoustic Wavefields, and we have set up the ANR collaborative project ECHANGE (ECHantillonnage Acoustique Nouvelle Génération) which began in January 2009. Rémi Gribonval is the coordinator of the project.

In 2010, the activity on ECHANGE has concentrated on Nearfield acoustic holography (NAH), a technique aiming at reconstructing the operational deflection shapes of a vibrating structure, from the near sound field it generates. In this application scenario, the objective is either to improve the quality of the reconstruction (for a given number of sensors), or reduce the number of sensors, or both, by exploiting a sparsity hypothesis which helps regularizing the inverse problem involved.
Contributions of the team in this task spans: notations and model definitions, experimental setting design and implementation, choice of an adapted dictionary in which the sparsity hypothesis holds, improved acquisition strategies through pseudo-random sensor arrays and/or spatial multiplexing of the inputs, experimental study of robustness issues, and theoretical study of potential success guarantees based on the restricted isometry property (which revealed being not verified in our case, despite improved experimental performance).

A paper about robustness issues and spatial multiplexing (an alternative to building antennas with random sensor position) was published in GRETSI [88]. A journal paper is under revision.

6.3.2. Sparse reconstruction for underwater acoustics (ECHANGE ANR project)

Participants: Rémi Gribonval, Valentin Emiya, Nikos Stefanakis, Nancy Bertin.

Main collaborations: Jacques Marchal, Pierre Cervenka (UPMC Univ Paris 06)

Underwater acoustic imaging is traditionally performed with beamforming: beams are formed at emission to insonify limited angular regions; beams are (synthetically) formed at reception to form the image. We proposed to exploit a natural sparsity prior to perform 3D underwater imaging using a newly built flexible-configuration sonar device. The computational challenges raised by the high-dimensionality of the problem were highlighted, and we described a strategy to overcome them. As a proof of concept, the proposed approach was used on real data acquired with the new sonar to obtain an image of an underwater target. We discussed the merits of the obtained image in comparison with standard beamforming, as well as the main challenges lying ahead, and the bottlenecks that will need to be solved before sparse methods can be fully exploited in the context of underwater compressed 3D sonar imaging. This work has been submitted to ICASSP 2012 and a journal paper is in preparation.

6.3.3. Audio inpainting (SMALL FET-Open project)

Participants: Rémi Gribonval, Valentin Emiya.

Main collaborations: Amir Adler, Michael Elad (Computer Science Department, The Technion, Israel); Maria G. Jafari, Mark D. Plumbley (Centre for Digital Music, Department of Electronic Engineering, Queen Mary University of London, U.K.).

Inpainting is a particular kind of inverse problems that has been extensively addressed in the recent years in the field of image processing. It consists in reconstructing a set of missing pixels in an image based on the observation of the remaining pixels. Sparse representations have proved to be particularly appropriate to address this problem. However, inpainting audio data has never been defined as such so far.

METISS has initiated a series of works about audio inpainting, from its definition to methods to address it. This research has begun in the framework of the EU Framework 7 FET-Open project FP7-ICT-225913-SMALL (Sparse Models, Algorithms and Learning for Large-Scale data) which began in January 2009. Rémi Gribonval is the coordinator of the project. The research on audio inpainting has been conducted by Valentin Emiya in 2010 and 2011.

The contributions consist of:

- defining audio inpainting as a general scheme where missing audio data must be estimated: it covers a number of existing audio processing tasks that have been addressed separately so far – click removal, declipping, packet loss concealment, unmasking in time-frequency;
- proposing algorithms based on sparse representations for audio inpainting (based on Matching Pursuit and on $l_1$ minimization);
- addressing the case of audio declipping (i.e. desaturation): thanks to the flexibility of our inpainting algorithms, they can be constrained so as to include the structure of signals due to clipping in the objective to optimize. The resulting performance are significantly improved. This work has been reported in [47] and it will appear as a journal paper [96].
Current and future works deal with developing advanced sparse decomposition for audio inpainting, including several forms of structured sparsity (e.g., temporal and multichannel joint-sparsity) and several applicative scenarios (declipping, time-frequency inpainting).

6.4. Music Content Processing and Music Information Retrieval

6.4.1. Acoustic music modeling

Participants: Nancy Bertin, Emmanuel Vincent.

Main collaborations: R. Badeau (Télécom ParisTech), J. Wu (University of Tokyo)

Music involves several levels of information, from the acoustic signal up to cognitive quantities such as composer style or key, through mid-level quantities such as a musical score or a sequence of chords. The dependencies between mid-level and lower- or higher-level information can be represented through acoustic models and language models, respectively.

Our acoustic models are based on nonnegative matrix factorization (NMF) and variants thereof. NMF models an input short-term magnitude spectrum as a linear combination of magnitude spectra, which are adapted to the input under suitable constraints such as harmonicity and temporal smoothness. While our previous work considered harmonic spectra only, we proposed the use of wideband spectra to represent attack transients and showed that this resulted in improved pitch transcription accuracy [77]. Our past work on the convergence properties of NMF was also disseminated [50].

We used the resulting model parameters to identify the musical instrument associated with each note, by means of a Support Vector Machine (SVM) classifier trained on solo data, and obtained improved instrument classification accuracy compared to state-of-the-art Mel-Frequency Cepstral Coefficient (MFCC) features [42], [78].

6.4.2. Music language modeling

Participants: Frédéric Bimbot, Emmanuel Vincent.

Main collaboration: S.A. Raczynski (University of Tokyo, JP)

We pursued our pioneering work on music language modeling, with a particular focus on the joint modeling of "horizontal" (sequential) and "vertical" (simultaneous) dependencies between notes by log-linear interpolation of the corresponding conditional distributions. We identified the normalization of the resulting distribution as a crucial problem for the performance of the model and proposed an exact solution to this problem.

We also applied the log-linear interpolation paradigm to the joint modeling of melody, key, chords and meter, which evolve according to different timelines. In order to synchronize these feature sequences, we explored the use of beat-long templates consisting of several notes as opposed to short time frames containing a fragment of a single note.

Both of these studies are ongoing.

6.4.3. Music structuring

Participants: Frédéric Bimbot, Gabriel Sargent, Emmanuel Vincent.

External collaboration: Emmanuel Deruty (as an independent consultant)

The structure of a music piece is a concept which is often referred to in various areas of music sciences and technologies, but for which there is no commonly agreed definition. This raises a methodological issue in MIR, when designing and evaluating automatic structure inference algorithms. It also strongly limits the possibility to produce consistent large-scale annotation datasets in a cooperative manner.

We have pursued our investigations on autonomous and comparable blocks, based on principles inspired from structuralism and generativism for producing music structure annotation. This work has allowed consolidating the methodology and producing additional annotations (over 400 pieces) [53].
We have also developed an algorithm aiming at the automatic inference of autonomous and comparable blocks using the timbral and harmonic content of music pieces, in combination with a regularity constraint [72]. Tested within the QUAERO project and during the MIREX 2011 campaign [94], the algorithm ranked among state-of-the-art methods.

6.5. Source separation

6.5.1. A general framework for audio source separation

Participants: Alexis Benichoux, Frédéric Bimbot, Charles Blandin, Ngoc Duong, Rémi Gribonval, Nobutaka Ito, Alexey Ozerov, Emmanuel Vincent.

Main collaborations: H. Tachibana (University of Tokyo, JP), N. Ono (National Institute of Informatics, JP)

Source separation is the task of retrieving the source signals underlying a multichannel mixture signal. The state-of-the-art approach, which we presented in a survey chapter [95], consists of representing the signals in the time-frequency domain and estimating the source coefficients by sparse decomposition in that basis. This approach relies on spatial cues, which are often not sufficient to discriminate the sources unambiguously. Last year, we proposed a general probabilistic framework for the joint exploitation of spatial and spectral cues [39] that was disseminated in several invited talks [43], [44]. This framework relies in particular on the thesis of Ngoc Duong, which was defended this year [30]. It makes it possible to quickly design a new model adapted to the data at hand and estimate its parameters via the EM algorithm. As such, it is expected to become the basis for a number of works in the field, including our own.

Since the EM algorithm is sensitive to initialization, we devoted a major part of our work to reducing this sensitivity. One approach is to set probabilistic priors over the model parameters, including spatial position priors [56] or temporal continuity priors [55]. A complementary approach is to initialize the parameters in a suitable way using source localization techniques specifically designed for environments involving multiple sources and possibly background noise [33], [54], [83]. In a longer-term perspective, we also investigated the design and exploitation of sparsity priors over time-domain acoustic transfer functions [52], [82].

6.5.2. Exploiting filter sparsity for source localization and/or separation

Participants: Alexis Benichoux, Prasad Sudhakar, Emmanuel Vincent, Rémi Gribonval, Frédéric Bimbot.

Main collaboration: Simon Arberet (EPFL)

Estimating the filters associated to room impulse responses between a source and a microphone is a recurrent problem with applications such as source separation, localization and remixing.

We considered the estimation of multiple room impulse responses from the simultaneous recording of several known sources. Existing techniques were restricted to the case where the number of sources is at most equal to the number of sensors. We relaxed this assumption in the case where the sources are known. To this aim, we proposed statistical models of the filters associated with convex log-likelihoods, and we proposed a convex optimization algorithm to solve the inverse problem with the resulting penalties. We provided a comparison between penalties via a set of experiments which shows that our method allows to speed up the recording process with a controlled quality tradeoff. This work was presented at two conferences [52], [82] and a journal paper including extensive experiments with real data is in preparation.

We also investigated the filter estimation problem in a blind setting, where the source signals are unknown. We proposed an approach for the estimation of sparse filters from a convolutive mixture of sources, exploiting the time-domain sparsity of the mixing filters and the sparsity of the sources in the time-frequency (TF) domain. The proposed approach is based on a wideband formulation of the cross-relation (CR) in the TF domain and on a framework including two steps: (a) a clustering step, to determine the TF points where the CR is valid; (b) a filter estimation step, to recover the set of filters associated with each source. We proposed for the first time a method to blindly perform the clustering step (a) and we showed that the proposed approach based on the wideband CR outperforms the narrowband approach and the GCC-PHAT approach by between 5 dB and 20 dB. This work has been published at ICASSP 2011 [49] and submitted for publication as a journal paper.
On a more theoretical side, we studied the frequency permutation ambiguity traditionally incurred by blind convolutive source separation methods. We focussed on the filter permutation problem in the absence of scaling, investigating the possible use of the temporal sparsity of the filters as a property enabling permutation correction. The obtained theoretical and experimental results highlight the potential as well as the limits of sparsity as an hypothesis to obtain a well-posed permutation problem. This work has been submitted for publication as a journal paper [99].

6.5.3. Towards real-world separation and remixing applications

Participants: Valentin Emiya, Alexey Ozerov, Laurent Simon, Emmanuel Vincent.

Shoko Araki (NTT Communication Science Laboratories, JP), Cédric Févotte (Télécom ParisTech, FR), Antoine Liutkus (Télécom ParisTech, FR), Volker Hohmann (University of Oldenburg, DE)

Following our founding role in the organization of a regular source separation evaluation campaign (SiSEC), we wrote an invited paper summarizing the outcomes of the three latest campaigns [41]. While some challenges remain, this paper highlighted that progress has been made and that audio source separation is closer than ever to successful industrial applications. This is also exemplified by the i3DMusic project and the contract recently signed with MAIA Studio.

In order to exploit our know-how for these real-world applications, we investigated issues such as how to implement our algorithms in real time and how best to exploit human input or metadata [68], [70]. In addition, while the state-of-the-art quality metrics previously developed by METISS remain widely used in the community, we proposed a new set of perceptually motivated metrics which greatly increase correlation with subjective assessments [34].

6.5.4. Source separation for multisource content indexing

Participants: Kamil Adiloglu, Alexey Ozerov, Emmanuel Vincent.

Main collaborations: J. Barker (University of Sheffield, UK), M. Lagrange (IRCAM, FR)

Another promising real-world application of source separation concerns information retrieval from multi-source data. Source separation may then be used as a pre-processing stage, such that the characteristics of each source can be separately estimated. The main difficulty is not to amplify errors from the source separation stage through subsequent feature extraction and classification stages. To this aim, we proposed a principled Bayesian approach to the estimation of the uncertainty about the separated source signals [45] and propagated this uncertainty to the features. We then exploited it in the training of the classifier itself, thereby greatly increasing classification accuracy [69].

While our work in this direction was initially motivated by music applications (e.g. artist recognition by separating the vocals from the accompaniment), we eventually applied it to noise-robust speech recognition, which is a better defined task [71]. In order to motivate further work by the community, we created a new international evaluation campaign on that topic (CHiME) [86].
6. New Results

6.1. Speech Analysis and Synthesis

Participants: Anne Bonneau, Vincent Colotte, Dominique Fohr, Yves Laprie, Joseph di Martino, Slim Ouni, Asterios Toutios, Sébastien Demange, Fadoua Bahja, Agnès Piquard-Kipffer, Utpala Musti.

6.1.1. Acoustic-to-articulatory inversion

6.1.1.1. Building new articulatory models

The possibility of generating the same sounds as those uttered by the speaker (or at least vocal tract transfer functions not too far from those observed) via the articulatory model and the acoustic simulation constitutes the underlying hypothesis of an analysis by synthesis method of acoustic-to-articulatory inversion. The articulatory model, and consequently its construction, thus plays a crucial role in inversion. An geometrical adaptation procedure has been developed in order to account for new speakers [28], [29]. It uses two scaling factors, one for the mouth cavity and the second for the pharyngeal cavity. In addition the model can be rotated globally and a second rotation controls the relative position of the pharynx with respect to the mouth cavity. In order to ensure a smooth transition from the mouth cavity to the pharynx cavity the angle of the rotation is a function of the distance with respect to the mouth axis.

The adaptation and model have been tested by using the X-ray data used by Maeda to construct his model. It should be noted that there are very few X-ray data with articulatory contour information available. These data correspond to a female speaker. The RMS reconstruction error reached by the adapted articulatory model is 0.550 mm what is very good for this particular speaker. Other data will be used in the future to validate the model and the adaptation procedure as soon as the contours will be delineated. An anatomical adpation procedure will also be developed in the future.

6.1.1.2. Determination of the vocal tract centerline

The connection of the articulatory model with the acoustic simulation requires the area function to be decomposed into elementary uniform tubes. The decomposition should respect the plane wave propagation. For that purpose the central line of the vocal tract has to be determined. The quality of the centerline strongly influences the closeness between natural and artificial formant frequencies.

We designed two complementary algorithms. The first exploits a dynamic programming approach to select points on interior and exterior walls of the vocal tract which minimize a global criterion combining the length of the centerline and the angle between the normal to the segments linking the points selected on both walls and the centerline [29]. It turned out that this first algorithm provides an insufficient smoothness of the centerline. A second algorithm has been designed by using an active curve which maximizes the smoothness of the centerline and the distance from any point of the centerline with exterior and interior walls. This second algorithm provides very good results.

6.1.1.3. Adaptation of cepstral coefficients for inversion

The inversion of speech requires spectra of natural speech to be compared with spectra synthesized via the articulatory synthesizer. This comparison cannot be carried out directly because the source is not taken into account in the synthetic spectra. Last year we thus investigated an affine adaptation of all the cepstral coefficients. This adaptation brings the spectral peaks of natural and synthetic spectra closer but at the same time tends to flatten the spectra. Moreover, it also appears that adaptation of only the very first cepstral coefficients (the first two except $C_0$ which represents energy) were sufficient to capture the spectral tilt. Since it is important to keep clear spectral peaks to explore the articulatory space, we used the bilinear transform in order to bring the two spectra closer [15]. The results are now better and the bilinear transform will be used to recover inverse solutions.
6.1.1.4. Acoustic-to-articulatory inversion using a generative episodic memory

We have developed an episodic based inversion method. Episodic modeling is interesting for two reasons. First, it does not rely on any assumption about the mapping relationship between acoustic and articulatory, but rather it relies on real synchronized acoustic and articulatory data streams. Second, the memory structurally embeds the naturalness of the articulatory dynamics as speech segments (called episodes) instead of single observations as for the codebook based methods. Estimating the unknown articulatory trajectories from a particular acoustic signal, with an episodic memory, consists in finding the sequence of episodes, which acoustically best explains the input acoustic signal. We refer to such a memory as a concatenative memory (C-Mem) as the result is always expressed as a concatenation of episodes. Actually a C-Mem lacks from generalization capabilities as it contains only several examples of a given phoneme and fails to invert an acoustic signal, which is not similar to the ones it contains. However, if we look within each episode we can find local similarities between them. We proposed to take advantage of these local similarities to build a generative episodic memory (G-Mem) by creating inter-episodes transitions. The proposed G-Mem allows switching between episodes during the inversion according to their local similarities. Care is taken when building the G-Mem and specifically when defining the inter-episodes transitions in order to preserve the naturalness of the generated trajectories. Thus, contrary to a C-Mem the G-Mem is able to produce totally unseen trajectories according to the input acoustic signal and thus offers generalization capabilities. The method was implemented and evaluated on the MOCHA corpus, and on a corpus that we recorded using an AG500 articulograph. The results showed the effectiveness of the proposed G-Mem which significantly outperformed standard codebook and C-Mem based approaches. Moreover similar performances to those reported in the literature with recently proposed methods (mainly parametric) were reached. [18]

The paradigm of episodic memories was also used for speech recognition. We do not extend the acoustic feature with any explicit articulatory measurements but instead we used the articulatory-acoustic generative episodic memories (G-mem). The proposed recognizer is made of different memories each specialized for a particular articulator. As all the articulators do not contribute equally to the realization of a particular phoneme, the specialized memories do not perform equally regarding each phoneme. We showed, through phone string recognition experiments that combining the recognition hypotheses resulting from the different articulatory specialized memories leads to significant recognition improvements. [19].

6.1.2. Using Articulography for Speech production

Since we have an articulograph (AG500, Carstens Medizinelektronik) available, we can easily acquire articulatory data required to study speech production. The articulograph is used to record the movement of the tongue (this technique is called electromagnetography - EMA). The AG500 has a very good time resolution (200Hz), which allows capturing all articulatory dynamics. The articulograph was used in a study about inversion (see the previous section) and to investigate pharyngealization.

Pharyngealized phonemes are commonly described as having the same place of articulation (dental) as their non-pharyngealized counterparts, but differ by the presence of a secondary articulation involving mainly the back of the tongue.

To study pharyngealized phonemes in Arabic from an articulatory point of view, our articulograph was used to record the movement of the tongue. Although EMA is not known as an optimal technique to cover the back of the tongue, good placement of the sensors and good interpretation of their positions can help to define pharyngealization relevantly. In fact, it is important to set one sensor as far as possible on the tongue (in our case, at 7cm from the tongue tip).

A corpus of several CVCVCVs was recorded using this articulograph, then phonetically labeled, and analyzed. The main finding of this work is that the coarticulation effect of the pharyngealized phonemes extends the immediate surrounding phonemes to influence the phonemes up to four-phoneme distance from the pharyngealized phoneme. The pharyngealization affects indifferently the previous and the following vowels and consonants.
We also investigated the effect of pharyngealization in Modern Standard Arabic (MSA) and Dialectal Arabic (DA). The acoustic material was more important than EMA. Although, we studied one speaker for EMA, the obtained results are encouraging to record more arabic speakers. [42]

6.1.3. Labial coarticulation

Results show that protrusion is a fragile cue to the rounding feature. Although we observe for each speaker a clear (but not large) separation between vowels /i/ and /y/ produced in isolation, many realizations of /i/ and /y/ come very close together and even overlap in few cases for vowels in contexts. The efficiency of the parameter depends on speakers and contexts. The distance between the corners is probably the most fragile cue to vowel roundedness. Many overlapping areas are observed for vowels in context. This is not good news for speech specialists since this parameter is easy to measure (with cameras and markers painted on the speaker’s face) and its evaluation can be fully automatic. Each of the three lip opening parameters constitutes a very efficient cue to the rounding feature. For vertical opening, the opposition between /i/ and /y/ in initial position appears to be endangered in bilabial context, due to the anticipation of lip closing during /i/. Nevertheless, the temporal variations of lip opening during the initial /i/ are very important, and more analyses, taking into account these variations, will be necessary to analyse /i/ vs. /y/ phonetic distinction more thoroughly.

6.1.4. Speech synthesis

Visual data acquisition was performed simultaneously with acoustic data recording, using an improved version of a low-cost 3D facial data acquisition infrastructure. The system uses two fast monochrome cameras, a PC, and painted markers, and provides a sufficiently fast acquisition rate to enable an efficient temporal tracking of 3D points. The recorded corpus consisted of the 3D positions of 252 markers covering the whole face. The lower part of the face was covered by 70% of all the markers (178 markers), where 52 markers were covering only the lips so as to enable a fine lip modeling. The corpus was made of 319 medium-sized French sentences uttered by a native male speaker and corresponding to about 25 minutes of speech.

We designed a first version of the text to acoustic-visual speech synthesis based on this corpus. The system uses bimodal diphones (an acoustic component and a visual one) and unit selection techniques (see 3.2.4). We have introduced visual features in the selection step of the TTS process. The result of the selection is the path in the lattice of candidates found in the Viterbi algorithm, which minimizes a weighted linear combination of three costs: the target cost, the acoustic joined cost, and the visual joined cost.

Finding the best set of weights is a difficult problem by itself mainly because of their highly different nature (linguistic, acoustic, and visual considerations). This year, we added the first derivative of the visual trajectories in the visual join cost and we developed a method to determine automatically the weights applied to each cost, using a series of metrics that assess quantitatively the performance of synthesis [37].

This year, more progress have been made regarding the definition of the target cost. Now, The target cost includes both acoustic target cost and visual target cost.

The visual target cost includes visual and articulatory information. We implemented and evaluated two techniques [32]: (1) Phonetic category modification, where the purpose was to change the current characteristics of some phonemes which were based on phonetic knowledge. The changes modified the target and candidate description for the target cost to better take into account their main characteristics as observed in the audio-visual corpus. The expectation was that their synthesized visual speech component would be more similar to the real visual speech after the changes. (2) Continuous visual target cost, where the visual target cost component is now considered as real value, and thus continuous, based on the articulatory feature statistics.

6.1.5. Phonemic discrimination evaluation in language acquisition and in dyslexia and dysphasia

6.1.5.1. Phonemic segmentation in reading and reading-related skills acquisition in dyslexic children and adolescents

Our computerized tool EVALEC was published [67] after the study of reading level and reading related skills of 400 hundred children from grade 1 to grade 4 (from age 6 to age 10) [69]. This research was supported by a grant from the French Ministry of Health (Contrat 17-02-001, 2002-2005). This first computerized battery
of tests in French language assessing reading and related skills (phonemic segmentation, phonological short term memory) comparing results both to chronological age controls and reading level age control in order to diagnostic Dyslexia. Both processing speed and accuracy scores are taken into account. This battery of tests is used by speech and language therapists. We keep on examining the reliability (group study) and the prevalence (multiple case study) of 15 dyslexics’ phonological deficits in reading and reading related skills in comparison with a hundred reading level children [68], and by the mean of longitudinal studies of children from age 5 to age 17 [66]. This year, we started the development of a project which examined multimodal speech both with SLI, dyslexics and control children (30 children). Our goal is to examine visual contribution to speech perception across different experiments with a natural face (syllables with several conditions). Our goal is to search what can improve intelligibility in children who have severe language acquisition difficulties.

6.1.5.2. Language acquisition and language disabilities (deaf children, dysphasic children)

Providing help for improving French language acquisition for hard of hearing (HOH) children or for children with language disabilities was one of our goal: ADT (Action of Technological Development) Handicom [piquardkipffer:2010:inria-00545856:2]. The originality of this project was to combine psycholinguistic and speech analyses researches. New ways to learn to speak/read were developed. A collection of three digital books has been written by Agnès Piquard-Kipffer for both 2-6, 5-9, 8-12 year old children (kindergarten, 1-4th grade) to train speaking and reading acquisition regarding their relationship with speech perception and audio-visual speech perception. A web interface has been created (using Symfony and AJAX technologies) in order to create others books for language impaired children. A workflow which transforms a text and an audio source in a video of digital head has been developed. This workflow includes an automatic speech alignment, a phonetic transcription, a speech synthesizer, a French cued speech coding and speaking digital head. A series of studies (simple cases studies, 5 deaf children and 5 SLI children and group studies with 2 kindergarten classes) were proposed to investigate the linguistic, audio-visual processing. ... presumed to contribute to language acquisition in deaf children. Publication are submitted.

6.1.6. Enhancement of esophageal voice

6.1.6.1. Detection of F0 in real-time for audio: application to pathological voices

The work first rested on the CATE algorithm developed by Joseph Di Martino and Yves Laprie, in Nancy, 1999. The CATE (Circular Autocorrelation of the Temporal Excitation) algorithm is based on the computation of the autocorrelation of the temporal excitation signal which is extracted from the speech log-spectrum. We tested the performance of the parameters using the Bagshaw database, which is constituted of fifty sentences, pronounced by a male and a female speaker. The reference signal is recorded simultaneously with a microphone and a laryngograph in an acoustically isolated room. These data are used for the calculation of the contour of the pitch reference. When the new optimal parameters from the CATE algorithm were calculated, we carried out statistical tests with the C functions provided by Paul BAGSHAW. The results obtained were very satisfactory and a first publication relative to this work was accepted and presented at the ISIVC 2010 conference [46]. At the same time, we improved the voiced / unvoiced decision by using a clever majority vote algorithm electing the actual F0 index candidate. A second publication describing this new result was published at the ISCT 2010 conference [45].

6.1.6.2. Voice conversion techniques applied to pathological voice repair

Voice conversion is a technique that modifies a source speaker’s speech to be perceived as if a target speaker had spoken it. One of the most commonly used techniques is the conversion by GMM (Gaussian Mixture Model). This model, proposed by Stylianou, allows for efficient statistical modeling of the acoustic space of a speaker. Let “x” be a sequence of vectors characterizing a spectral sentence pronounced by the source speaker and “y” be a sequence of vectors describing the same sentence pronounced by the target speaker. The goal is to estimate a function F that can transform each source vector as nearest as possible of the corresponding target vector. In the literature, two methods using GMM models have been developed: In the first method (Stylianou), the GMM parameters are determined by minimizing a mean squared distance between the transformed vectors and target vectors. In the second method (Kain), source and target vectors are combined in a single vector “z”. Then, the joint distribution parameters of source and target speakers is estimated using
the EM optimization technique. Contrary to these two well known techniques, the transform function $F$, in our laboratory, is statistically computed directly from the data: no needs of EM or LSM techniques are necessary. On the other hand, $F$ is refined by an iterative process. The consequence of this strategy is that the estimation of $F$ is robust and is obtained in a reasonable lapse of time. This interesting result was published and presented at the ISIVC 2010 conference [70].

6.1.7. Perception and production of prosodic contours in L1 and L2

6.1.7.1. Language learning (feedback on prosody)

Feedback on L2 prosody based upon visual displays, speech modifications and automatic diagnosis has been elaborated and a pilot experiment undertaken to test its immediate impact on listeners. Results show that the various kinds of feedback provided by the system enable French learners with a low production level to improve their realisations of English lexical accents more than (simple) auditory feedback. These results should be confirmed with a large number of speakers but based upon the important differences between results obtained for speakers in test and control conditions, we are confident in the interest of the system presented here [41]. In particular, the system analyses learners’ realisations and provide indications on what they should correct, a guidance which is considered as necessary by specialists in the oral aspects of language learning.

6.1.7.2. Production of prosody contour

We report here relevant observations for the study continuation in French. These observations were obtained in an ongoing project about non-conclusive prosodic patterns in French and English (“Intonale” project 7.2.2). We specifically discuss slope variations, estimated in semitones, concerning two kinds of non-conclusive configurations, which are inside a clause, or at the end of a clause, respectively: (i) the final segment of a subject NP in an assertive sentence, followed or not by another syntagm ended by a continuation contour (ii) the final segment of a A clause, in a two clause utterance AB, where A and B are assertive clauses connected by an discourse relation, marked or not with a conjunction.

Intonation slopes are computed as regression slopes using F0 values in semitones estimated every 10 ms. Slopes are calculated on the two last syllables of the target segments of every sentence. Results show that slopes for segments which are not at the end of a clause, and segments at the end of a clause followed by a conjunction are typically rising, and not significantly different the ones from the others. On the contrary, slopes for ends of clauses not followed by a conjunction are significantly different from the previous ones. More than 50

6.1.8. Pitch detection

Over the last two years, we have proposed two new real time pitch detection algorithms (PDAs) based on the circular autocorrelation of the glottal excitation, weighted by temporal functions, derived from the CATE [64] original algorithm (Circular Autocorrelation of the Temporal Excitation), proposed initially by J. Di Martino and Y. Laprie. In fact, this latter algorithm is not constructively real time because it uses a post-processing technique for the Voiced/Unvoiced (V/UV) decision. The first algorithm we developed is the eCATE algorithm (enhanced CATE) that uses a simple V/UV decision less robust than the one proposed later in the eCATE++ algorithm.

We propose a recent modified version called the eCATE++ algorithm which focuses especially on the detection of the F0, the tracking of the pitch and the voicing decision in real time. The objective of the eCATE++ algorithm consists in providing low classification errors in order to obtain a perfect alignment with the pitch contours extracted from the Bagshaw database by using robust voicing decision methods. The main improvement obtained in this study concerns the voicing decision, and we show that we reach good results for the two corpora of the Bagshaw database.

6.2. Automatic Speech Recognition

Participants: Christophe Cerisara, Sébastien Demange, Dominique Fohr, Christian Gillot, Jean-Paul Haton, Irina Illina, Denis Jouvet, Odile Mella, Luiza Orosanu, Othman Lachhab, Larbi Mesbah.
6.2.1. Core recognition

6.2.1.1. Broadcast News Transcription

In the framework of the Technolangue project ESTER, we have developed a complete system, named ANTS, for French broadcast news transcription (see section 5.4).

Extensions of the ANTS system have been studied, including the possibility to use the Sphinx recognizers. Training scripts for building acoustic models for the Sphinx recognizers are now available and take benefit of the computer cluster for a rapid optimization of the model parameters. The Sphinx models are also used for speech/text alignment on both French and English speech data. A new speech decoding program has been developed for efficient decoding on the computer cluster, and easy modification of the decoding steps (speaker segmentation and clustering, data classification, speech decoding in one or several passes, ...). It handles both the Julius and Sphinx (versions 3 and 4) decoders.

This year, we have proposed an approach to grapheme-to-phoneme conversion based on a probabilistic method: Conditional Random Fields (CRF). CRF gives a long term prediction, and assume a relaxed state independence condition. Moreover, we proposed an algorithm to the one-to-one letter to phoneme alignment needed for CRF training. This alignment is based on discrete HMMs. The proposed system was validated on two pronunciation dictionaries. Different set of input features were studied: POS-tag, context size, unigram versus bigram. Our approach compared favorably with the performance of the state-of-the-art Joint-Multigram Models (JMM) for the quality of the pronunciations, but provided better recall and precision measures for multiple pronunciation variants generation [22] [21].

As the pronunciation lexicon is one the key-points of a speech recognition system, we have investigated to which extent wiktionary data can be used to build such a lexicon. Collecting the pronunciations available for many entries of the wiktionary make possible the creation of an initial pronunciation lexicon. Such initial lexicon is then used for training grapheme-to-phoneme conversion systems (either CRF-based of JMM-based), in order to obtain pronunciation variants for words that are not in the initial pronunciation lexicon extracted from the web wiktionary data. Combining the pronunciation variants generated by the 2 grapheme-to-phoneme systems provides the best results. Although the achieved results are not as good as those obtained with a hand-made pronunciation lexicon, this automatic approach makes possible an easy creation of a pronunciation lexicon for a new language [26].

Confidence measures aim at estimating the confidence of a hypothesis result provided by the speech recognition engine. Two word confidence measures were proposed, which can be computed without waiting for the end of the audio stream; one frame-synchronous and one local. Our local measures achieved performance very close to a state-of-the-art measure which requires the recognition of the whole sentence. A preliminary experiment to assess the contribution of our confidence measure in improving the comprehension of automatic transcription results by hearing impaired was also conducted [10].

6.2.1.2. Speech recognition for interaction in virtual worlds

Automatic speech recognition is investigated for vocal interaction in virtual worlds, in the context of serious games in the EMOSPEECH project. This year, a wizard-of-oz experiment was carried out to collect speech data corresponding to the dialogs from 5 players interacting with a serious game. The players were invited to speak freely to any character of the game with whom it is possible to interact, while the wizard of Oz (a game expert localized in the same room) answered them. Hence, the recorded interactions between the player and the characters of the game are natural dialogs. The audio sessions have been manually transcribed. Each session comprises roughly 30 speech turns (one player’s sentence plus one wizard’s sentence).

For training the language models, the text dialogs recorded by the TALARIS team (Midiki corpus) on the same serious game (but in a text-based interaction), have been used on addition of available broadcast news corpus. For this purpose we have also manually corrected the Midiki sentences, in order to handle the numerous typos and misspellings as well as chat specific "words" such as smileys ("mdr" or "lol"), emphasized punctuations ("!!!!!") or over-segmentations such as "é-lec-tro-nique". This normalization step is a strong requirement for speech recognition models. Different language models have then been created using different vocabulary sizes.
The acoustic models are adapted from the radio broadcast news models, using state-of-the-art Maximum A Posteriori adaptation algorithm. This reduces the mismatch in recording conditions between the game devices and the original models trained on radio streams. We are currently investigating solutions to integrate this adaptation within the speech recognition component and perform it online. At runtime, the targeted strategy is to ask the player to utter some few predefined sentences and to use these sentences to adapt the generic acoustic models to the player’s voice.

6.2.2. Speech recognition modeling

Robustness of speech recognition to multiple sources of speech variability is one of the most difficult challenge that limits the development of speech recognition technologies. We are actively contributing to this area via the development of the following advanced modeling approaches.

6.2.2.1. Detection of Out-Of-Vocabulary words

One of the key problems for large vocabulary continuous speech recognition is the occurrence of speech segments that are not modeled by the knowledge sources of the system. An important type of such segments are so-called Out-Of-Vocabulary (OOV) words (words are not included in the lexicon of the recognizer). Mostly OOV words yield more than one error in the transcription result because the error can propagate due to the language model.

We have investigated, with Frederik Stouten (postdoctoral), to what extent OOV words can be detected. For this we used a classifier that makes a decision about each speech frame whether it belongs to an OOV word or not. Acoustic features for this classifier are derived from three recognition systems. On top of the acoustic features we also used four language model features: the ngram probability, the order of the gram that was used to calculate the language model probability, the unigram probability for the current word and a binary indicator that takes the value one if the word is preceded by a first name.

We propose to exploit the fact that 38% of the OOV word observations in the broadcast news data are pronounced more than one time in a time period of less than 1 minute. To improve the detection of repeated OOV words, we design a clustering module working on the detected OOV word segments. This algorithm is based on the estimation of the entropy. The proposed incremental clustering algorithm has been evaluated on the broadcast news corpus ESTER and gave better performance than a classical baseline incremental clustering algorithm based on a distance threshold [36].

6.2.2.2. Detailed modeling exploiting uncertainty

Modeling pronunciation variation is an important topic for automatic speech recognition. It has been widely observed that speech recognition performance degrades notably on spontaneous speech, and more precisely, that the word error rate increases when the degree of spontaneity increases. The rate of speech is also an important variability source which impacts notably on the acoustic realization of the sounds as well as on the pronunciation of the words, and consequently affects recognition performance. Large increases in word error rates are observed when speaking rate increases. And, it should be noted that rate of speech and spontaneous speech are not completely independent as the rate of speech is an important cue for detecting spontaneous speech.

This year, we have investigated further the detailed modeling of the probabilities of pronunciation variants for large vocabulary continuous speech recognition, and evaluated it on broadcast news transcriptions. In particular we have refined the modeling of the probabilities of the pronunciation variants dependent on the speaking rate. This was achieved by taking into account the uncertainty in the estimation of the speaking rate that results from the word and phoneme boundary uncertainty (speech signal - phoneme alignment errors). Such uncertainty was handled both in the training process and in the decoding step, leading to speech recognition performance improvements [25].

Detailed acoustic modeling was also investigated using automatic classification of speaker data. With such an approach it is possible to go beyond the traditional four class models (male vs female, studio quality vs telephone quality). However, as the amount of training data for each class gets smaller when the number of classes increases, this limits the amount of classes that can efficiently be trained. Hence, this year we have
investigated introducing a classification margin in the classification process. With such a margin, which handle boundary classification uncertainty, speech data at the class-boundary may belong to several classes. This increases the amount of training data in each class, which makes the class acoustic model parameters more reliable, and finally improved the overall recognition performance.

6.2.2.3. Speech recognition using distant recording

Speech recognition of distant recording of speech commands was investigated. A set of domotic commands were recorded from a few speakers using a far talking microphone. Acoustic models were adapted to this context using some training data played with a loud speaker, and recorded using a distant microphone. Among other results, preliminary experiments showed the benefit of adapting the models, as well as using a noise robust acoustic analysis when dealing with noisy data.

6.2.2.4. Training HMM acoustic models

At the beginning of his second internship at INRIA Nancy research laboratory, Othman Lachhab focused on the finalization of a speech recognition system based on context-independent HMMs models, using bigram probabilities for the phonotactic constraints and a model of duration following a normal distribution $N(\mu, \sigma^2)$ incorporated directly in the Viterbi search process. Currently, he built a reference system for speaker-independent continuous phone recognition using Context-Independent Continuous Density HMM (CI-CDHMM) modeled by Gaussian Mixture Models (GMMs). In this system he developed his own training technique, based on a statistical algorithm estimating the classical optimal parameters. This new training process compares favorably with already published HMM technology on the same test corpus (TIMIT).

6.2.3. Speech/text alignment

6.2.3.1. Alignment with native speech

Speech to text alignment is a research objective that is derived from speech recognition. While it seems easier to solve at first sight, expectations are also higher and new problems appear, such as how to handle very large audio documents, or how to handle out-of-vocabulary words. Another important challenge that motivated our work in this area concerns how to improve our results and meet the user expectation by exploiting as much as possible the interactions and feedback loop between the end-user and the system. This year, we kept on improving the open-source JTrans software platform for this task as described in Section (see section 5.6 ). We further submitted an ANR Corpus proposal in collaboration with University Paris 3. We also sent a new version of the software to the "Timecode" company to help them investigating the usefulness of this approach in the application context of foreign film dubbing (see section 7.4.1 ).

6.2.3.2. Alignment with non-native speech

Non-native speech alignment with text is one critical step in computer assisted foreign language learning. The alignment is necessary to analyze the learner’s utterance, in view of providing some prosody feedback (as for example bad duration of some syllables - too short or too long -). However, non-native speech alignment with text is much more complicated than native speech alignment. This is due to the pronunciation deviations observed on non-native speech, as for example the replacement of some target language phonemes by phonemes of the mother tongue, as well as errors in the pronunciations. Moreover, these pronunciation deviations are strongly speaker dependent (i.e. they depend on the mother tongue of the speaker, and on its fluency in the target foreign language) which makes their prediction difficult.

In this application context, the precision of phoneme boundaries is critical. Hence, speech-text alignment was investigated on non-native speech. A large non-native speech corpus has been manually segmented for building a reference corpus. Then automatic phonetic segmentation (resulting from the speech-text alignment) has been analyzed. The results shows that rather reliable boundaries are obtained for some phonetic classes [31] and that better results are obtained when only frequent pronunciation deviations are kept as variants in the pronunciation lexicon [27]. Further work is on-going to determine automatically a confidence value on the proposed alignments.
6.2.4. Computing and merging linguistic information on speech transcripts

The raw output of speech recognition is difficult to read for humans, and difficult to exploit for further automatic processing. We thus investigated solutions to enrich speech recognition outputs with non-lexical information, such as dialog acts, punctuation marks and syntactic dependencies. Computing such a linguistic information requires a corpus to train stochastic models, and we also worked out new semi-supervised training algorithms for building a French corpus dedicated to syntactic parsing of oral speech. The creation of this corpus is realized in collaboration with the TALARIS team. Finally, we designed a new solution to improve our core language models by integrating into them lexical semantic distances.

An important information for post-processing speech transcripts concerns dialog acts and punctuation marks. We initiated some work in this area several years ago with the Ph.D. thesis of Pavel Kral. Since then, we continued our collaboration in this domain by successively investigating specific challenges, such as finding the most relevant features, models and testing the adaptation of our approaches in two languages, Czech and French [59]. We further proposed this year an approach to improve commas generation with the help of syntactic features [17].

Inferring syntactic dependencies is an extremely important step towards structuring the text and an absolute prerequisite for working with relations between words and next interpreting the utterance. Yet, no state-of-the-art solutions designed for parsing written texts can be reliably adapted to parsing speech, and even less transcribed speech. The lack of such methods and resources is especially blatant in French. We started, in collaboration with the TALARIS team, to address this issue by building a new French treebank dedicated to speech parsing [52], as well as a software platform dedicated to working with this corpus (see section 5.5). We exploited this year this corpus to study specific syntactic structures, such as negations (Master internship in 2011) and left dislocations in French [13].

While a large part of our work is dedicated to enriching the output of our speech recognition system, we also tried integrating within the speech decoding process itself new information coming from the higher-levels. We thus extended the new approach proposed in 2010 about language model smoothing with a new probabilistic smoothing that takes into account much longer words history thanks to a Levenshtein-based clustering of the training sentences [20].

6.3. Speech-to-Speech Translation and Langage Modeling

Participants: Kamel Smaïli, David Langlois, Sylvain Raybaud.

Our work on Confidence Measures is now published in Machine Translation [9]. Now we are working on Speech-to-Text translation. We have proposed a method to segment audio input stream for machine translation. First, audio stream is split into overlapping segments useful for speech recognition; then, these segments are transcribed and regrouped; last, the obtained text flow is segmented into machine translation-friendly segments and translated. We incorporated this work into our speech-to-text machine translation system, and we evaluated our system for French-English broadcast news translation [34]. The following step consists in integrating Confidence Measures into the system in order to improve the integration between the both recognition-translation processes.

Moreover, we pursued our collaboration with Chiraz Latiri from the URPAH Team, University of Tunis. Running on our previous works (based on word-based machine translation system) we compared our respective methods in the scope of phrase-based machine translation [30].
6. New Results

6.1. Coordination Parsing

**Participants:** Bruno Guillaume, Guy Perrier.

In the development of the French grammar, FRIGRAM, Joseph Le Roux and Guy Perrier have tackled the difficult problem of modelling and parsing coordination [39]. They have enriched FRIGRAM with a module expressing different syntactic constructions with coordination. An important drawback of this approach is the number of elementary constructions that have to be introduced to obtain a reasonable coverage of the phenomenon.

In the continuation of his Master thesis, Valmi Dufour-Lussier with Bruno Guillaume and Guy Perrier worked on a different approach. They propose to process coordination at the parsing level as a linguistic performance issue, outside the grammar, rather than as a matter of competence [15]. They apply a specific algorithm to combine coordinated syntactic structures that were partially parsed using a coordination-less grammar, resulting in a directed acyclic parse graph in which constituent sharing appears sharply. They have experimented the algorithm within the framework of Tree-Adjoining Grammars (although it can be adapted to other formalisms) on a small subset of the Penn Treebank 2. They have shown that it is able to handle many types of coordinative constructions, including left and right node raising, argument clusters, and verb gapping.

6.2. Graph Rewriting

**Participants:** Bruno Guillaume, Mathieu Morey, Guy Perrier.

Guillaume Bonfante (from CARTE team), Bruno Guillaume, Mathieu Morey and Guy Perrier have improved their graph rewriting calculus, experimenting it in two directions. Taking an asynchronous perspective on the syntax-semantics interface, they have designed a modular graph rewriting system to produce underspecified semantic representations from a syntactic dependency graph [14]. They experimentally validated this approach on a set of sentences extracted from the French Treebank annotated with syntactic dependencies [27]. The results open the way for the production of underspecified semantic dependency structures from corpora annotated with syntactic dependencies and, more generally, for a broader use of modular rewriting systems for computational linguistics.

In a second application, they show how to enrich a syntactic dependency annotation of the French Treebank, using graph rewriting, in order to compute its semantic representation [13]. The rewriting system is composed of grammatical and lexical rules structured in modules. The lexical rules use a control information extracted from Dicovalence, a lexicon of French verbs 3.

6.3. ACG Type System

**Participants:** Philippe de Groote, Sylvain Pogodalla, Florent Pompigne.

In order to extend the flexibility and the expressiveness of the the ACG framework, [57] proposed a type-system extension. However, the formal properties of the system have to be proved. In his PhD work, Florent Pompigne is proposing alternate \(\eta\)-rules and commutative conversions in order to get the desirable properties. This work, currently in progress, relates to former proposals for a linear calculus with dependent types [28] and a calculus for extensionality with variants [40].

---

2 [http://www.cis.upenn.edu/~treebank/](http://www.cis.upenn.edu/~treebank/)

6.4. Logic and Grammars

**Participant:** Maxime Amblard.

Maxime Amblard has presented an extension of Minimalist Categorial Grammars (MCG) to encode Chomsky’s phases in [11]. These grammars are based on Partially Commutative Logic (PCL) and encode properties of Minimalist Grammars (MG) of Stabler. The first implementation of MCG were using both non-commutative properties (to respect the linear word order in an utterance) and commutative ones (to model features of different constituents). Here, we propose to adding Chomsky’s phases with the non-commutative tensor product of the logic.

6.5. Discourse dynamics

**Participants:** Maxime Amblard, Sai Qian.

Sai Qian and Maxime Amblard has presented a framework which constructs an event-style discourse semantics, [17]. The discourse dynamics are encoded in continuation semantics [54] and various rhetorical relations are embedded in the resulting interpretation of the framework. They assume that discourse and sentence are distinct semantic objects, that play different roles in meaning evaluation. Moreover, two sets of composition functions, for handling different discourse relations, are introduced.

6.6. Modeling pathological discourse

**Participant:** Maxime Amblard.

Maxime Amblard starts a conjoint work with a psychologist Michel Musiol (IntePsy) and a philosopher Manuel Rebuschi (Archives Poincaré) about developing a formal analysis of pathological conversations involving schizophrenic speakers [18]. Such conversations give rise to manifest incongruities or ruptures that can be seen as mere contradictions by any “normal” speaker. Our construal relies both on semantic and pragmatic features of conversation. After an overview on the making of the corpus, we propose a SDRT-inspired account of pathological conversations, and we apply it to two relevant excerpts. We conclude with a short discussion about the localization of incoherencies by schizophrenics, either in semantics or in pragmatics, and its importance for our understanding of thought disorders.
6. New Results

6.1. MLIF

TALARIS contributes to ISO TC 37 committee “Terminologies and other Language Resources”, and more specifically to the activities of its SC3 “Computer Applications in Terminology”, and SC4 “Linguistic Resources Management”. Within TC37/SC4, TALARIS is currently contributing, as project leader, to the definition and specification of the Multi Lingual Information Framework (MLIF) [ISO FDIS 24616]. MLIF is being designed with the objective of providing a common abstract model being able to generate several formats used in the framework of translation and localization. MLIF has been released as FDIS (Final Draft International Standard) and it should finally be published as an official ISO standard soon. MLIF has been extensively used within the ITEA2 META VERSE1 project. [42], [43], [12].

6.2. TEXT CLASSIFICATION

Neural clustering algorithms show high performance in the general context of the analysis of homogeneous textual datasets. We have recently proposed a new incremental growing neural gas algorithm using the cluster label maximization (IGNGF) [44] [34]. In this strategy the use of a standard distance measure for determining a winner is completely suppressed by considering the label maximization approach as the main winner selection process. One if its important advantage is that it provides the method with an efficient incremental character as it becomes independent of parameters. Although it performs better than the standard clustering methods on textual data, we have shown this year than the obtained results are not as efficient as expected whenever an analysis of very complex heterogeneous textual datasets is performed [33]. We have thus explored several variations of IGNG-F approach based on combination of distance based criteria and cluster label maximization. Our new results on all kinds of datasets, especially on the most complex heterogeneous textual datasets, clearly reflect the advantages of our new algorithm as compared to other existing algorithms and to our former adaptations [29]. Cluster quality evaluation represents a key process for all kinds of data analysis tasks, and more especially for textual data. We have recently presented different variations of unsupervised Recall/Precision and F-measures measures that cope with the defects of classical indexes, like inertia-based indexes. Our new indexes directly exploit the maximized features of the data associated to each cluster after the clustering process without prior consideration of clusters profiles. As compared to classical indexes, their main advantage is thus to be independent of the clustering methods and of their operating mode. They thus altogether permit the objective comparison of clustering methods and represent a sound technique for efficient cluster labeling. We have more especially worked this year on the large scale validation our indexes using reference labeled textual datasets [35].

We are also currently investigating to set up a platform for efficiently assisting the patents experts in the process of patents validation. Reaching such a goal has implied to develop new semi-supervised classification methods or propose in-deep adaptation of existing ones in order to establish relevant relationships between hierarchical patents classification and bibliographical references describing research covering the fields related to the different patents classes. In this context, we have successfully explored this year new classification techniques based on taboo search [14].

To cope with the current defects of existing incremental clustering methods, an alternative approach for analyzing information evolving over time consists in performing diachronic analysis. We have thus explored this year different an original technique based on this approach on texts by the use of the combination of cluster labeling with unsupervised Bayesian reasoning between cluster labels extracted from clustering model issued from different time periods. Based on a reference dataset issued from the IST-PROMTECH project, we have clearly shown that these new techniques, whilst providing a new framework for automatizing such kind of analysis, outperformed existing ones [32] [31] [30].
6.3. DIALOG
Within the Emospeech project, we developed the Emospeech Dialogue Toolkit (cf. Software section); used the Wizard of Oz infrastructure it includes to collect dialog data; and trained an interpreter and a dialog manager. The collected data comprises 591 dialogues in French collected within the context of the Mission Plastechnology serious game, 4874 utterances, 77854 words and 1321 player utterances containing 12901 word tokens and 1427 word types. We collected in average 50 conversations for each sub-dialogue in the game. Dialog length varies from 78 to 142 with an average length of 106 utterances per dialog.

6.4. GIVE
For the Generation of Instruction in Virtual Environment challenge edition 2.5 (GIVE), we developed two systems. The first system is the successor of the system that we presented to the GIVE 2 challenge (2010). We solved two known problems of this system, namely the indefinite presupposition problem and the ambiguity arising from underspecified referring expressions [26]. The GIVE 2.5 challenge proved that these improvements were efficient, and showed an increase of 21% in terms of task success (47% in GIVE 2, 68% in GIVE 2.5). The second system, developed in collaboration with the University of Cordoba is the first to our knowledge that uses a human-human corpus to provide whole utterances thanks to plan matching techniques [22], [21]. The system ranked fifth in terms of task success (58%), but second in terms of referring targets identification. The naturalness of instructions and the simplified system development makes it an interesting research track to follow.

6.5. Verb Classification
To help computer systems in the task of understanding and representing the full meaning of a text, verb classifications have been proposed which group together verbs with similar syntactic and semantic behaviour. For English verbs, VerbNet provides such a large scale classification but there are no similar French resource available. We investigated different ways both of automatically constructing such a resource; and of evaluating it.

Using Formal Concept Analysis (FCA), we developed a method for classifying verbs based on their (syntactic) subcategorisation information extracted from existing French lexical resources; and by translating the English Verbnet, we showed how to associate the obtained classes with semantic information represented by Verbnet’s thematic role sets ([27]). As a result, a VerbNet like classification for French verbs can be constructed fully automatically.

The FCA approach we pursued, first builds a classification based on verbs and verb features and second filters this classification using various metrics (e.g., concept probability, concept stability). We are currently comparing this approach with a clustering approach which makes use of detailed evaluation metrics [44] and uses probabilistic information to guide classification. First results are promising and outperform the state of the art methods in this domain [63].

One important difference between the clustering and the FCA approach we experimented with is that only the second, allows a verb to belong to several classes. Since verbs are highly ambiguous, this is an important difference. To evaluate the impact of this difference on the usability of the classifications built by each of the methods, we are currently conducting a task-based, extrinsic evaluation of both classifications by analysing their impact when used in a Semantic Role Labeling task on a French corpus.

6.6. I-FLEG
Within the Allegro project, we developed the I-FLEG game [16], [15], a virtual world in which the learner exercises French by clicking on objects and answering the questions raised by the system. The language learning exercises produced by I-FLEG are automatically generated using the GenI sentence generator from a knowledge base describing the virtual world. A preliminary evaluation of I-FLEG with school children [17] suggests that the “game” aspect increases learner motivation and that spoken output is essential in maintaining learner interest.
6. New Results

6.1. Geometry Processing

**Participants:** Laurent Alonso, Alejandro Galindo, Phuong Ho, Samuel Hornus, Bruno Lévy, David Lopez, Kun Liu, Vincent Nivoliers, Nicolas Ray, Dmitry Sokolov, Rhaleb Zayer.

- **Sampling:** In the frame of the Goodshape project, we continued developing new techniques for sampling shapes optimally, based on the notion of Centroidal Voronoi Tessellations. We developed a new technique for computing clipped 3D Voronoi diagrams, suitable to volumetric meshing [23], and an accelerated GPU-based centroidal Voronoi tesselation [20]. We developed an optimization technique to suppress obtuse triangles [21]. We also studied periodic boundary conditions [29], suitable to some numerical simulations.

- **Geometric modeling and computational geometry:** Dobrina Boltcheva joined the team, with her expertise on simplicial homology [10]. Vincent Nivoliers published a L-system based knot insertion rules [19] that he developed during his master in Gipsa lab. We also studied furthest polygon Voronoi diagrams [15].

- We gave an invited course on geometry processing at SIGGRAPH Asia [24].

6.2. Computer Graphics

**Participants:** Laurent Alonso, Anass Lasram, Samuel Hornus, Bruno Jobard, Anass Lasram, Sylvain Lefebvre, Bruno Lévy, Vincent Nivoliers, Nicolas Ray.

- **Foundations of Computer Graphics:** We developed new algorithms and data structures for spatial caching and hashing on the GPU [16]. We continued our work on noise generation based on Gabor convolution and proposed several improvements [17] (see Figure 5). We also studied the fundamental problem of interpolating functions and developed an algorithm based on optimal transport theory [11].

- **Applications:** We developed an algorithm to change the lighting in images that contain trees [12] (with REVES, see Figure 6).
Figure 6. Relighting photos of canopies.

Figure 7. Unified visualization of atomic bounds and molecular surfaces.
Figure 8. Left: flow lines between injectors (black squares) and well (center square). Right: Adapted Voronoi grid.

Figure 9. Fitting a spline surface (bottom) to a triangle mesh (top) starting from an initial template (center). Data courtesy of Distene.
6.3. Scientific Visualization and scientific computing

Participants: Samuel Hornus, Bruno Jobard, Bruno Lévy, Romain Merland, Vincent Nivoliers, Jeanne Pellerin, Nicolas Ray, Dmitry Sokolov, Rhaleb Zayer.

- **Molecular visualization:** We continued the development of our Micromegas software, and developed several improvements [13], [14] (see Figure 7).

- **Geo-Modeling and geo-visualization:** In the frame of our partnership with the Gocad consortium, we developed an evaluation of multi-valued data depiction techniques [22]. We also developed several meshing tools dedicated to the numerical simulation of oil exploitation [28], [26], [25] (see Figure 8).

- **Reverse engineering:** We developed methods to convert mesh surfaces and point sets into parametric splines, using either a global parameterization technique [18] or an approximation of surface-to-surface distance based on Voronoi diagrams [27] (see Figure 9).
6. New Results

6.1. Lighting and Rendering

Participants: Mahdi Bagher, Laurent Belcour, Georges-Pierre Bonneau, Eric Bruneton, Cyril Crassin, Jean-Dominique Gascuel, Olivier Hoel, Nicolas Holzschuch, Fabrice Neyret, Cyril Soler, Fabrice Neyret, Charles de Rousiers, Cyril Soler.

6.1.1. Non-linear Pre-filtering Methods for Efficient and Accurate Surface Shading

Participants: Eric Bruneton, Fabrice Neyret.

Rendering a complex surface accurately and without aliasing requires the evaluation of an integral for each pixel, namely a weighted average of the outgoing radiance over the pixel footprint on the surface. The outgoing radiance is itself given by a local illumination equation as a function of the incident radiance and of the surface properties. Computing all this numerically during rendering can be extremely costly. For efficiency, especially for real-time rendering, it is necessary to use precomputations. When the fine scale surface geometry, reflectance and illumination properties are specified with maps on a coarse mesh (such as color maps, normal maps, horizon maps or shadow maps), a frequently used simple idea is to pre-filter each map linearly and separately. The averaged outgoing radiance, i.e., the average of the values given by the local illumination equation is then estimated by applying this equation to the averaged surface parameters. But this is really not accurate because this equation is non-linear, due to self-occlusions, self-shadowing, non-linear reflectance functions, etc. Some methods use more complex pre-filtering algorithms to cope with these non-linear effects. In [14] we presented a survey of these methods. We have started with a general presentation of the problem of pre-filtering complex surfaces. We then present and classify the existing methods according to the approximations they make to tackle this difficult problem. Finally, an analysis of these methods allows us to highlight some generic tools to pre-filter maps used in non-linear functions, and to identify open issues to address the general problem.

6.1.2. Frequency-Based Kernel Estimation for Progressive Photon Mapping

Participants: Laurent Belcour, Cyril Soler.

We have developed an extension to Hachisuka et al.’s Progressive Photon Mapping (or PPM) algorithm [32] in which we estimate the radius of the density estimation kernels using frequency analysis of light transport [29]. We predict the local radiance frequency at the surface of objects using a Gaussian approximation, and use it to drive the size of the density estimation kernels, in order to accelerate convergence (see Figure 3). The key is to add frequency information to a small proportion of photons: frequency photons. In addition to contributing to the density estimation, they will provide frequency information. This work has been published in [20].

6.1.3. Efficiently Visualizing Massive Tetrahedral Meshes with Topology Preservation

Participant: Georges-Pierre Bonneau.
Figure 3. In this figure we compare against progressive photon mapping with our algorithm for the convergence of an indirectly lit part of the scene. In the closeup, we show that our algorithm produces a lower varying estimate at an earlier stage of its execution. The images were produced using 100,000 photons per pass and 25% of frequency photons to make timing comparable.

Figure 4. Left: Variable resolution visualization of a volume mesh with multiple linear features. The topology of the substructures is guaranteed to be preserved. Right: Snapshot of the multiresolution visualization tool to explore simulation data with embedded structures on a desktop PC.
This work is the result of a collaboration with S. Hahmann from the EVASION team-project and Prof. Hans Hagen partly done during a sabbatical of G.-P. Bonneau in the University of Kaiserslautern, Germany. Interdisciplinary efforts in modeling and simulating phenomena have led to complex multi-physics models involving different physical properties and materials in the same system. Within a 3d domain, substructures of lower dimensions appear at the interface between different materials. Correspondingly, an unstructured tetrahedral mesh used for such a simulation includes 2d and 1d substructures embedded in the vertices, edges and faces of the mesh. The simplification of such tetrahedral meshes must preserve (1) the geometry and the topology of the 3d domain, (2) the simulated data and (3) the geometry and topology of the embedded substructures. This work focuses on the preservation of the topology of 1d and 2d substructures embedded in an unstructured tetrahedral mesh, during edge collapse simplification. We derive a robust algorithm, based on combinatorial topology results, in order to determine if an edge can be collapsed without changing the topology of both the mesh and all embedded substructures. Based on this algorithm we have developed a system for simplifying scientific datasets defined on irregular tetrahedral meshes with substructures, illustrated in Figure 4. We presented and demonstrated the power of our system with real world scientific datasets from electromagnetism simulations in the Springer book chapter [27].

6.1.4. Real-Time Rough Refraction

Participants: Nicolas Holzschuch, Charles de Rousiers.

We have developed an algorithm to render objects of transparent materials with rough surfaces in real-time, under distant illumination. Rough surfaces cause wide scattering as light enters and exits objects, which significantly complicates the rendering of such materials. We present two contributions to approximate the successive scattering events at interfaces, due to rough refraction: First, an approximation of the Bidirectional Transmittance Distribution Function (BTDF), using spherical Gaussians, suitable for real-time estimation of environment lighting using pre-convolution; second, a combination of cone tracing and macro-geometry filtering to efficiently integrate the scattered rays at the exiting interface of the object. We demonstrate in I3D paper [24] the quality of our approximation by comparison against stochastic raytracing. This work is illustrated in Figure 5.
6.1.5. **Interactive Indirect Illumination Using Voxel Cone Tracing**  
**Participants:** Cyril Crassin, Fabrice Neyret.

Indirect illumination is an important element for realistic image synthesis, but its computation is expensive and highly dependent on the complexity of the scene and of the BRDF of the involved surfaces. While offline computation and pre-baking can be acceptable for some cases, many applications (games, simulators, etc.) require real-time or interactive approaches to evaluate indirect illumination. We present in the Pacific Graphics paper [16] a novel algorithm to compute indirect lighting in real-time that avoids costly precomputation steps and is not restricted to low-frequency illumination. An illustration is given in Figure 6. It is based on a hierarchical voxel octree representation generated and updated on the fly from a regular scene mesh coupled with an approximate voxel cone tracing that allows for a fast estimation of the visibility and incoming energy. Our approach can manage two light bounces for both Lambertian and glossy materials at interactive framerates (25-70FPS). It exhibits an almost scene-independent performance and can handle complex scenes with dynamic content thanks to an interactive octree-voxelization scheme. In addition, we demonstrate that our voxel cone tracing can be used to efficiently estimate Ambient Occlusion. A primer of this work has been published as a poster (Best Poster Award [22]). Insights of the method were given in the Siggraph Talk 2011 [23].

The publication [22] has received the Best Poster Award at I3D’2011.

6.1.6. **Fast multi-resolution shading of acquired reflectance using bandwidth prediction**  
**Participants:** Mahdi Bagher, Laurent Belcour, Nicolas Holzschuch, Cyril Soler.

Shading complex materials such as acquired reflectances in multi-light environments is computationally expensive. Estimating the shading integral involves stochastic sampling of the incident illumination independently at several pixels. The number of samples required for this integration varies across the image, depending on an intricate combination of several factors. Ignoring visibility, adaptively distributing computational budget across the pixels for shading is already a challenging problem. In the paper [28] we present a systematic approach to accelerate shading, by rapidly predicting the approximate spatial and angular variation in the local light field arriving at each pixel. Our estimation of variation is in the form of local bandwidth, and accounts for combinations of a variety of factors: the reflectance at the pixel, the nature of the illumination, the local geometry and the camera position relative to the geometry and lighting. An illustration is given in Figure 7. The speed-up, using our method, is from a combination of two factors. First, rather than shade every pixel, we use this predicted variation to direct computational effort towards regions of the image with high local variation. Second, we use the predicted variance of the shading integrals, to cleverly distribute a fixed total budget of shading samples across the pixels. For example, reflection off specular objects is estimated using fewer samples than off diffuse objects.
Figure 7. The technique developed by Mahdi Bagher allows to predict in real time the local bandwidth of the image obtained by shading a measured material with all frequency distant illumination (See inset colored top-right image). This information allows a drastic economy of samples in the computation of the integrals that are needed to produce an accurate image. In particular this allows to perform multi-sampling anti-aliasing in a deferred shading pipeline with much less image-space samples than the brute-force solution.
6.2. Expressive Rendering and Visualization

Participants: Pierre Bénard, Georges-Pierre Bonneau, Alexandre Coninx, Joëlle Thollot.

6.2.1. Temporal Coherence for Stylized Animation

Participants: Pierre Bénard, Joëlle Thollot.

Figure 8. In our state-of-the-art report we review and carefully compare Temporal Coherence techniques for stylized animations.

Non-photorealistic rendering (NPR) algorithms allow the creation of images in a variety of styles, ranging from line drawing and pen-and-ink to oil painting and watercolor. These algorithms provide greater flexibility, control and automation over traditional drawing and painting. Despite significant progress over the past 15 years, the application of NPR to the generation of stylized animations remains an active area of research. The main challenge of computer generated stylized animations is to reproduce the look of traditional drawings and paintings while minimizing distracting flickering and sliding artifacts present in hand-drawn animations. These goals are inherently conflicting and any attempt to address the temporal coherence of stylized animations is a trade-off. We have published the state-of-the-art report [15] motivated by the growing number of methods proposed in recent years and the need for a comprehensive analysis of the trade-offs they propose. We formalize the problem of temporal coherence in terms of goals and compare existing methods accordingly. We propose an analysis for both line and region stylization methods and discuss initial steps toward their perceptual evaluation. The goal of our report is to help uninformed readers to choose the method that best suits their needs, as well as motivate further research to address the limitations of existing methods.

6.2.2. Visualization of data with uncertainty using perceptually guided procedural noise

Participants: Alexandre Coninx, Georges-Pierre Bonneau.

This work is the result of a collaboration with EdF R&D and Jacques Droulez, Director of Research at CNRS in Collège de France. In his PhD work, Alexandre Coninx has introduced a new method to visualize uncertain scalar data fields by combining color scale visualization techniques with animated, perceptually adapted Perlin noise. The parameters of the Perlin noise are controlled by the uncertainty information to produce animated patterns showing local data value and quality, as illustrated in Figure 9. In order to precisely control the perception of the noise patterns, we perform a psychophysical evaluation of contrast sensitivity thresholds for a set of Perlin noise stimuli. We validate and extend this evaluation using an existing computational model. This allows us to predict the perception of the uncertainty noise patterns for arbitrary choices of parameters. We demonstrate and discuss the efficiency and the benefits of our method with various settings, color maps and data sets. This work has been published at APGV’2011 [21].
Interaction and Visualization - New Results - Project-Team ARTIS

Figure 9. Left: classical colormap visualization of scalar data without uncertainty. Right: in our technique, we perturb the input of the colormap using a perceptually guided procedural noise, scaled by the uncertainty of the data. The data and its uncertainty can be visualized in the same image.

6.3. Modeling and Animation

Participants: Georges-Pierre Bonneau, Alexandre Derouet-Jourdan, Nicolas Holzschuch, Nassim Jibai, Cyril Soler, Joëlle Thollot.

6.3.1. Multiscale Feature-Preserving Smoothing of Tomographic Data

Participants: Nassim Jibai, Nicolas Holzschuch, Cyril Soler.

Figure 10. Left: A contour surface extracted from noisy tomographic data contains surface noise and several topological artifacts such as small handles and holes. Right: The surface extracted from our smoothed volume is clean, and yet small features, such as the thread in the screw, and sharp edges have been preserved.

Computer tomography (CT) has wide application in medical imaging and reverse engineering. Due to the limited number of projections used in reconstructing the volume, the resulting 3D data is typically noisy. Contouring such data, for surface extraction, yields surfaces with localised artifacts of complex topology. To avoid such artifacts, we propose a method for feature-preserving smoothing of CT data, illustrated in Figure 10. The smoothing is based on anisotropic diffusion, with a diffusion tensor designed to smooth
noise up to a given scale, while preserving features. We compute these diffusion kernels from the directional histograms of gradients around each voxel, using a fast GPU implementation. This work has been published as a Siggraph’2011 Poster [26].

6.3.2. 3D Inverse Dynamic Modeling of Strands

Participants: Alexandre Derouet-Jourdan, Joëlle Thollot.

![Figure 11. Our model can physically animate the tail. The tail will retrieve its initial shape at the end of slight (possibly strong) motions.](image)

In this work, we propose a new method to automatically and consistently convert 3D splines into dynamic rods at rest under gravity, bridging the gap between the modeling of 3D strands (such as hair, plants) and their physics-based animation. An illustration is given in Figure 11. This work is done in collaboration with F. Bertails from the BIPOP team-project. It has been published in a Siggraph’2011 poster [25].

6.3.3. Lagrangian Texture Advection: Preserving both Spectrum and Velocity Field

Participants: Eric Bruneton, Nicolas Holzschuch, Fabrice Neyret.

Texturing an animated fluid is a useful way to augment the visual complexity of pictures without increasing the simulation time. But texturing flowing fluids is a complex issue, as it creates conflicting requirements: we want to keep the key texture properties (features, spectrum) while advecting the texture with the underlying flow — which distorts it. In this context we present a new, Lagrangian, method for advecting textures: the advected texture is computed only locally and follows the velocity field at each pixel (see Figure 12). The texture retains its local properties, including its Fourier spectrum, even though it is accurately advected. Due to its Lagrangian nature, our algorithm can perform on very large, potentially infinite scenes in real time. Our experiments show that it is well suited for a wide range of input textures, including, but not limited to, noise textures. This work has been published in the IEEE Transactions on Visualization and Computer Graphics (TVCG) [18].

6.3.4. Feature-Based Vector Simulation of Water Waves

Participant: Fabrice Neyret.

We have developed a method for simulating local water waves caused by obstacles in water streams for real-time graphics applications. Given a low-resolution water surface and velocity field, our method is able to decorate the input water surface with high resolution detail for the animated waves around obstacles. We construct and animate a vector representation of the waves. It is then converted to feature-aligned meshes for capturing the surfaces of the waves (see Figure 13). Results demonstrate that our method has the benefits of real-time performance and easy controllability. The method also fits well into a state-of-the-art river animation system. This work has been published in the Journal of Computer Animation and Virtual Worlds [19].
Figure 12. Left: Our method advects open-domain textures preserving both the spectrum and the motion field, in real-time. Right: Various applications in 2D and 3D, with procedural, image, bump, displacement textures.

Figure 13. Our method permits the real-time rendering of highly detailed animated features on large river scenes.
6.3.5. Volume-preserving FFD for Programmable Graphics Hardware

Participant: Georges-Pierre Bonneau.

Figure 14. Left: FFD deformation of the armadillo mesh, without volume preservation. Right: our technique: GPU-based volume preservation of the FFD deformation.

This work is the result of a collaboration with S. Hahmann from the EVASION team-project, Prof. Gershon Elber from Technion and Prof. Hans Hagen from the University of Kaiserslautern.

Free Form Deformation (FFD) is a well established technique for deforming arbitrary object shapes in space. Although more recent deformation techniques have been introduced, amongst them skeleton-based deformation and cage based deformation, the simple and versatile nature of FFD is a strong advantage, and justifies its presence in nowadays leading commercial geometric modeling and animation software systems. Since its introduction in the late 80’s, many improvements have been proposed to the FFD paradigm, including control lattices of arbitrary topology, direct shape manipulation and GPU implementation. Several authors have addressed the problem of volume preserving FFD. These previous approaches either make use of expensive non-linear optimization techniques, or resort to first order approximation suitable only for small-scale deformations. In this work we take advantage from the multi-linear nature of the volume constraint in order to derive a simple, exact and explicit solution to the problem of volume preserving FFD (see Figure 14). Two variants of the algorithm are given, without and with direct shape manipulation. Moreover, the linearity of our solution enables to implement it efficiently on GPU. The results have been published in a Visual Computer journal paper [17].
6. New Results

6.1. Visual Analytics of EA Data

Participants: Jean-Daniel Fekete, Évelyne Lutton [correspondant].

An experimental analysis of Evolutionary Algorithms (EAs) usually generates a huge amount of multidimensional data, including numeric and symbolic data. It is difficult to efficiently navigate in such a set of data, for instance to be able to tune the parameters or evaluate the efficiency of some operators. Usual features of existing EA visualisation systems consist in visualising time- or generation-dependent curves (fitness, diversity, or other statistics). When dealing with genomic information, the task becomes even more difficult, as a convenient visualisation strongly depends on the considered fitness landscape. In this latter case the raw data are usually sets of successive populations of points of a complex multidimensional space.

The purpose of this study was to evaluate GraphDice on complex sets of EA data (for artificial and real test-cases), and to sketch future developments of this tool, in order to better adapt it to the needs of EA experimental analysis (Fig. 6). An output of this study is the acceptance of the EASEA-Cloud ANR-Emergence project, in which developments will aim at adding tools in GraphDice specific for:

- visualisation of Evolutionary Algorithms written in the EASEA language,
- monitoring the execution of these algorithms on a cloud of computers (CPU + GPU).

Figure 6. Visual Analytics of EA Data: GraphDice showing fitness versus generation view. Red points correspond to the first 10 generations, yellow points, to the 10 last ones, and green points, to the best fitness areas.
6.2. Interactive Evolutionary Algorithms for Visual decision making

**Participants:** Nadia Boukhelifa, Waldo Cancino, Jean-Daniel Fekete, Evelyne Lutton [correspondant].

When dealing with very large datasets with many dimensions, it is often difficult to efficiently navigate and find interesting viewpoints, significative compound variables, unexpected behaviour, and other remarkable characteristics.

Our aim within the System@tic CSDL project (Complex Systems Design Lab, 2009–2012) is to use interactive evolutionary algorithm to assist the user in its exploration task. Finding an interesting, non obvious, viewpoint on a complex dataset can be formulated as an interactive optimisation problem. Population-based evolutionary search mechanisms can then efficiently be exploited for suggesting new viewpoints on data, that progressively adapt to the needs of an user.

In September 2011 (arrival of Nadia Boukhelifa and Waldo Cancino) we started to build a prototype based on GraphDice, that proposes new dimensions in the scatterplot matrix. These secondary set of dimensions are compositions of the dimensions of the initial dataset. Starting from an initial set of suggested dimensions (PCA analysis of the dataset), an evolutionary algorithm progressively refines the compound dimensions according to a measurement of the activity of the user on the corresponding views.

6.3. Optimisation of Food Models

**Participant:** Évelyne Lutton [correspondant].

*In collaboration with Alberto Tonda and Romain Reuillon, ISC-PIF*

The European project DREAM (http://dream.csregistry.org/) managed by INRA-CEPIA, aims at building decision support tools for better managing product quality and, by the way, manufacturing processes in the domain of agrifood industry.

Our contribution to this project is focused on the evolutionary optimisation of Bayesian Networks models, on the development of efficient cooperative-co-evolution schemes to solve some food modeling problems (milk gel, cheese ripening), and on the efficient visualisation of output data of these algorithms.

6.4. A Study on Dual-Scale Data Charts

We presented the results of a user study that compares different ways of representing dual-scale data charts (see Fig. 7). Dual-scale charts incorporate two different data resolutions into one chart in order to emphasize data in regions of interest or to enable the comparison of data from distant regions. While some design guidelines exist for these types of charts, there is currently little empirical evidence on which to base their design. We filled this gap by discussing the design space of dual-scale cartesian-coordinate charts and by experimentally comparing the performance of different chart types with respect to elementary graphical perception tasks such as comparing lengths and distances. Our study suggests that cut-out charts which include collocated full context and focus are the best alternative, and that superimposed charts in which focus and context overlap on top of each other should be avoided.

6.5. Information Visualization Evaluation

Petra Isenberg has contributed to three articles on evaluation methodologies: The first article “Collaborative Visualization: Definition, Challenges, and Research Agenda” [12] deals with challenges of collaborative visual analytics and includes a discussion on the challenges of evaluating tools during multi-person use. The second article, “Information Visualization Evaluation in Large Companies: Challenges, Experiences and Recommendations” [14], discusses challenges of evaluating and deploying visual analytics tools in a large company setting. It lists several challenges and provides concrete guidance to others who seek to evaluate tool within domain experts in their work environment. Finally, “Seven Guiding Scenarios for Information Visualization Evaluation” [33] is a pre-print of a journal article (in press) which provides a new viewpoint on evaluation in information visualization. Instead of giving an overview of methods, it cites evaluation goals and questions and can, thus, provide clear considerations for practitioners and researchers in the area.
6. New Results

6.1. High level model for shapes

6.1.1. Constructive implicit modeling

**Participants:** Adrien Bernhardt, Marie-Paule Cani, Maxime Quiblier, Cédric Zanni.

Implicit surfaces are an appealing representation for free-form, volumetric shapes. In addition to being able to represent shapes of arbitrary topological genus, they have the ability to be constructed by successively blending different components, which eases interactive modeling.

Within Cédric Zanni’s PhD, we are collaborating with a researcher in formal computation, Evelyne Hubert, to improve and extend the analytical methods for computing closed form solutions for convolution surfaces. We introduced a warping method for enabling the modeling of complex helical shapes from a single implicit primitive (fig. 3), which greatly enhances efficiency [14]. We also proposed a method based on anisotropic, surface Gabor noise, for generating procedural details on skeleton-based implicit surfaces. The surfaces enhanced with details can still be smoothly blended, with a natural transition between the details they carry. A paper has been submitted for publication. We are currently developing normalized convolution surfaces, invariant through homothetic transformations, and which will provide an intuitive blending sharpness parameter, usable with simple additive blending.

![Figure 3. Top: Helical primitives. Bottom: Implicit modeling of a squid. Each tentacle is made of two helical primitives.](image)

Lastly, we contributed to a new blending operator, gradient blending, which enables us to blend implicit shapes not only in function of the field values but also of their gradients. This solves a number long standing problems in implicit modeling: we can generate bulge-free blending, ensure that the topological genius of the blended shape remains the one of the union of the input one, and avoid the blur of small details. A paper is currently submitted for publication.
6.1.2. Ontology-based mesh segmentation

Participant: Olivier Palombi.

The smart use of data by automated systems is now a problem having implications as various as the optimization of the functioning of web search engine or medico-surgical simulation. The program MyCF is an attractive innovation in this field, gathering an organization of the biomedical knowledge in the form of ontologies, 3D acquisition of anatomical structures and also the possibility to export these 3D structures to biomechanics simulation programs, like SOFA. Our work consisted in creating, thanks to the program Protégé, an ontology of the functions of the human body (which didn’t existed then) and to couple it with the anatomical ontology FMA (ontology we significantly reworked). The objective was to build connections introducing a link between anatomical structures and functions they bear. Once this goal reached, it became possible for a computer to link the elimination of an anatomical structure and the loss of a function, this real-time. Thus we established the foundations of an ontology which currently gather 84484 entities (4330 of which are functional entities) and 4159 relations. So, our contributions are threefold: first of all, the creation of an ontology of functions of the human body which is an original one; then the redrafting of the FMA ontology which allowed us to complete some lacks and above all to make of it a tool more oriented towards the practical applications we waited about it, and finally, the contribution which seems to us the more significant is, without any doubt, the institution of a link between these two ontologies. This work is at is moment unfinished and should be pursued in order to approach always more the reality in the field of medico-surgical simulation.

6.1.3. French translation of the Foundational Model of Anatomy (FMA) ontology

Participant: Olivier Palombi.

The goal of this study, performed in collaboration with the LITIS was to facilitate the translation of FMA vocabulary into French. We compare two types of approaches to translate the FMA terms into French. The first one is UMLS-based on the conceptual information of the UMLS metathesaurus. The second method is lexically-based on several Natural Language Processing (NLP) tools. The two approaches permitted us to semi-automatically translate 3,776 FMA terms from English into French, this was to add to the existing 10,844 French FMA terms in the HMTP (4,436 FMA French terms and 6,408 FMA terms manually translated) [9].

6.1.4. Homology computation

Participants: Dobrina Boltcheva, Jean-Claude Léon.

This work is a part of the BQR project IDEAL (see Section 8.1.2) which is performed in collaboration with Leila de Floriani from the University of Genova in Italy. The main goal of this project is to study non-manifold geometrical models and to find out features allowing to classify these models and criteria for determining their shape. We are interested in non-manifold models such as idealized industrial CAD models, since they are still ill-understood even if they are frequently used in computer graphics and many engineering applications. This year, we have worked on the computation of topological invariants on non-manifold simplicial complexes, such as the homology groups, since they play a crucial role in the field of shape description and analysis. The goal was also to acquire a better understanding of the behaviour of the homological groups on non-manifold models. A first step towards this goal has been achieved this year and we have developed an efficient method for computing the homology of a large simplicial complex from the homologies of its sub-complexes. This work has been already published in a research report in 2010 and a journal paper in the context of the international conference on Solid and Physical Modeling [[2]].

6.1.5. Creased paper modeling

Participants: Marie-Paule Cani, Stefanie Hahmann, Damien Rohmer.

Although very common in real life, 3D creased paper models are rarely seen in virtual scene due to the lack of available modeling tools.

---

1Laboratoire d’Informatique, de Traitement de l’Information et des Systèmes (LITIS) Institut National des Sciences Appliquées de Rouen – Université du Havre – Université de Rouen
We developed a new approach to efficiently generate a 3D model of creased paper from a boundary curve (see fig. 4). The generated surface lie on the given boundary curve while preserving the lengths with respect to the original pattern. Contrary to other approaches, this method can seamlessly handles sharp creases while automatically generating the optimal mesh to this shape. The generation is fast enough to be used interactively, and the physical properties such as developability are approximatively preserved. This generation of static surface generation has been published in EG short paper [22].

6.1.6. Spline surface models for arbitrary topologies

Participant: Stefanie Hahmann.

In geometric modeling quad meshes have always been popular, in the sense that NURBS surfaces which are composed of a tensor product network of quadrilateral patches, are the inevitable standard for describing free-form shapes. They are defined on a chess board like assembly of quadrilateral parameter domains. But when modeling shapes of arbitrary topological type with tensor product NURBS, it is necessary to overcome the restriction of the tensor product configuration where always 4 patches meet at a common corner by using singular parameterizations.

We are developing new smooth parametric surface models defined on irregular quad meshes are in fact a powerful alternative to singularly parameterized tensor product surfaces since they combine the advantages of both, the arbitrary topology of quad meshes and the smoothness of the tensor product patches. Herein, tensor product polynomial patches are assembled with tangent plane continuity, in one-to-one correspondence to the mesh faces. They are thus capable to represent manifold shapes of arbitrary topological type since no restriction on the number of patches assembled around a mesh vertex exists.

While subdivision surfaces can also produce a smooth shape with only a few subdivision steps from a coarse mesh, our parametric surfaces have the advantage to provide an explicit parameterization. Moreover, all classical modeling operations such as trimming, intersection, blending and boolean operations can be preformed with parametric patches. Tensor product patches can furthermore make profit from the powerful tools of existing modeling systems for purposes of evaluation, display, interrogation and all operations cited above.

In collaboration with G.-P. Bonneau (Artis team) several parametric triangular surface models for arbitrary topologies have been published in the past (CAGD, IEEE TVCG and ACM ToG). A new tensor product spline surface model has been developed this year. It solves the problem of defining a $G^1$-continuous surface interpolating the vertices of an irregular quad mesh with low degree polynomial tensor product patches. It further aims to produce shapes of very high visual quality while reducing the number of control points. A comparison with existing methods and a journal paper are on-going work.
6.1.7. Point sampled surfaces

**Participant:** Stefanie Hahmann.

Point sampled geometry from scanned data exhibits very characteristic shapes, due to the presence of sharp features in most manufactured and designed objects. Therefore, reconstruction of surfaces from unorganized point sets using MLS fitting requires additional attention. In fact, it is an inherent property of MLS fitting to produce smooth surfaces, thus all sharp features in the point cloud may also be smoothed out. Instead of searching for appropriate new fitting functions our approach was to introduce a new method for selecting an appropriate local point neighborhood for the projection operator so that a standard MLS fitting automatically reproduces sharp features.

This work was part of Christopher Weber’s Ph.D. thesis, which has been co-advised by S. Hahmann and H. Hagen from TU Kaiserslautern, Germany. First part of the work on Gauss map clustering for feature point detection has been published in the SMI 2010 proceedings. The second part has been submitted to a Computer Graphics journal. The thesis has been defended in August 2011.

6.1.8. Volume preserving Free-Form Deformations

**Participant:** Stefanie Hahmann.

![Figure 5. Comparison between standard FFD deformation (top) and our method preserving the volume (bottom).](image)

Free Form Deformation (FFD) is a well established technique for deforming arbitrary object shapes in space. Although more recent deformation techniques have been introduced, amongst them skeleton-based deformation and cage based deformation, the simple and versatile nature of FFD is a strong advantage, and justifies its presence in nowadays leading commercial geometric modeling and animation software systems. Several authors have addressed the problem of volume preserving FFD. These previous approaches however make either use of expensive non-linear optimization techniques, or resort to first order approximation suitable only for small-scale deformations. Our approach was to take advantage from the multi-linear nature of the volume constraint in order to derive a simple, exact and explicit solution to the problem of volume preserving FFD. Two variants of the algorithm have been developed, without and with direct shape manipulation. Moreover, we showed that the linearity of our solution enables to implement it efficiently on GPU. This work has been done in collaboration with G. Elber from TECHNION, H. Hagen from TU Kaiserslautern, G.-P. Bonneau and S. Barbier from Artis INRIA. It has been accepted for publication in the journal The Visual Computer [6].

6.2. High level models for animation

6.2.1. Geometrical methods for skinning character animations

**Participants:** Marie-Paule Cani, Stefanie Hahmann, Damien Rohmer.
Skeletal animation is an efficient and widely used technique in video games or movie industry due to its flexibility and simplicity. Still, the skinning method do not take into account informations about the physical properties of the underlying material. Therefore effects such as muscle bulging or fat tissue compression cannot be modeled without the addition of a tedious manual correction. Within Damien Rohmers PhD [1], an active geometry framework was proposed in order to enhance geometry information with a priori knowledge about how the underlying material can deform. For instance, bending the belly of an animal will be constraint to generate bulges that will preserve locally the volume (see fig. 6). The process can be either purely automatically generated, or it can be artistically controled.

6.2.2. Action representation, segmentation and recognition

Participant: Remi Ronfard.

Following Daniel Weinland’s PhD thesis, we published a survey of modern methods for representing, segmenting and recognizing full-body actions in video [12]. A taxonomy of methods is elaborated in that paper, where actions can be represented with local, structured or global features both in time and in space. The potential for future work in grammar-based action recognition is emphasized, with possible applications in corpus-based procedural modeling of actions.

6.2.3. Frame-based simulation of deformable solids

Participants: Guillaume Bousquet, François Faure.

We present a new type of deformable model which combines the realism of physically based continuum mechanics models and the usability of frame-based skinning methods [4]. The degrees of freedom are coordinate frames (see Figure 15). In contrast with traditional skinning, frame positions are not scripted but move in reaction to internal body forces. The displacement field is smoothly interpolated using blending techniques such as dual quaternions. The deformation gradient and its derivatives are computed at each sample point of a deformed object and used in the equations of Lagrangian mechanics to achieve physical realism. This allows easy and very intuitive definition of the degrees of freedom of the deformable object. The meshless discretization allows on-the-fly insertion of frames to create local deformations where needed. We formulate the dynamics of these models in detail and describe some pre-computations that can be used for speed. We show that our method is effective for behaviors ranging from simple unimodal deformations to complex realistic deformations comparable with Finite Element simulations.
We extend the approach (see fig. 7) to the simulation of complex, intricated material distributions using material-aware shape functions [3]. Given a volumetric map of the material properties of an object and a number of control nodes, a distribution of the nodes is computed automatically, as well as the associated shape functions. Reference frames are attached to the nodes, and deformations are applied to the object using linear blend skinning. A continuum mechanics formulation is derived from the displacements and the material properties. We introduce novel material-aware shape functions in place of the traditional radial basis functions used in meshless frameworks. These allow coarse deformation functions to very efficiently resolve non-uniform stiffnesses. Complex models can thus be simulated at high frame rates using a small number of control nodes.

To encourage its use, the software is freely available in the simulation platform SOFA.

6.3. Towards interactive digital creation media

6.3.1. Sketch-based modeling and shape editing

Participants: Adrien Bernhardt, Rémi Brouet, Marie-Paule Cani, Jean-Claude Léon, Olivier Palombi.
3D modeling from a sketch is a fast and intuitive way of creating digital content.

We developed a method based on convolution surfaces for inferring free-form shapes in 3D from arbitrary progressive sketches, without any a priori knowledge on the objects being represented (see the section describing the Aestem Studio software). We recently investigated whether 2D deformation could be a better approach than sketching for defining the 2D sketch given as input [18]. Results are very promising: We are planning both to allow such intuitive deformations, combined with sketching, within the Aestem software, and to extend them to the editing of the 3D deformed shape. This will be done in the context of Rémi Brouet’s PhD thesis, co-advised by Renaud Blanch from the IIHM/LIG team.

We also develop methods for interpreting complex sketches (contours with T-junctions) based on some a priori knowledge. Our first work on this topic used the conventions of anatomical drawing to infer the 3D geometry of vascular systems, with branching and occlusions, from a single sketch [10]. We are also investigating the design of realistic terrains from a single sketch, within the PhD thesis of Adrien Bernhardt (see fig. 8). Our first advances include a new representation and new methods for generating a high-field from user-sketched constraint curves [30], [17].

### 6.3.2. Free-form sculpture
**Participants:** Marie-Paule Cani, Lucian Stanculescu.

In the context of Lucian Stanculescu’s PhD thesis, co-advised by Raphaëlle Chaine from LIRIS (Lyon), we developed an interactive sculpting system enabling both arbitrary deformation and topological changes of a free-form shape [11]. Our method is based on a semi-regular mesh which adaptively refines and changes its topology according to the need. See Figure 9. We are currently extending the method for handling the sculpting of composite objects made of many different components.

![Figure 9. Sculpting with topological changes made within the framework from [11].](image)

### 6.3.3. Hand Navigator
**Participants:** Jean-Rémy Chardonnet, Jean-Claude Léon.

The different deformation models we developed in the past few years open the problem of providing intuitive interaction tools for specifying the desired deformations in real-time. Therefore, our recent work focused on developing new devices for interacting with the model to deform. For the past two years, we focused on developing a peripheral device similar to a mouse, called the HandNavigator, enabling to control simultaneously ten or more degrees of freedom of a virtual hand. This device consists in a 3D mouse for the position and orientation of the hand in 3D space, enhanced with many sensors for moving and monitoring the virtual fingers. Thanks to a pre-industrialization project funded by the incubator GRAVIS, the first prototype, patented by INRIA, has been extended with the incorporation of new sensors and new shapes to improve the device efficiency. An ongoing extension of the patent and a partnership with HAPTION company are new
step toward the industrialization of this device. Dissemination to general public has been performed at the “Fête de la Science” and another exhibition. Publications will take place after setting up the patent extension. The ongoing BQR INTOUACTIVE funded by Grenoble-INP will lead to further scientific topics regarding interactions during grasping as well as with deformable bodies and a partnership has been set up with GIPSA-Lab to study the muscular activity during interactions.

6.3.4. Procedural modeling
Participants: Marie-Paule Cani, Arnaud Emilien.

We developed a method for procedurally generating villages with the appropriate roads and streets on arbitrary terrains, in collaboration with Eric Galin from LIRIS, Lyon. This work will be continued within Arnaud Emilien’s PhD thesis towards more general models for populating terrains with houses, vegetation, and animals. We will focus on the development of intuitive ways to edit procedural models, to overcome the main drawback of these approaches.

6.3.5. Computational model of film editing
Participant: Remi Ronfard.

Building on Remi Ronfard’s experience leading the virtual cinematography research team at Xtranormal Technology, Montreal, we designed a novel computational model for automatic editing of animated movies (see fig. 10). A prototype has been implemented in a collaboration with the Bunraku/mimetic team, and demonstrated in poster sessions at the Symposium on Computer Animation (SCA) [31] and International Conference on Interactive Digital Storytelling (ICIDS) [21]. This early work opens new directions that will be further explored by the IMAGINE team, including corpus-based learning of cinematography and editing styles.

Figure 10. Interface developed for automatic editing of animated movies [31], [21].
6. New Results

6.1. Interaction Techniques

**Participants:** Caroline Appert, Michel Beaudouin-Lafon [correspondant], David Bonnet, Anastasia Bezerianos, Olivier Chapuis, Guillaume Faure, Emilien Ghomi, Stéphane Huot, Mathieu Nancel, Wendy Mackay, Cyprien Pindat, Emmanuel Pietriga, Theophanis Tsandilas, Julie Wagner.

Acquiring a target, such as pointing to an icon, a button or a landmark on a digital map, is the most common action in today’s graphical user interfaces. We have continued our work to better understand this seemingly simple action and make it faster and more reliable. This year we have conducted theoretical work on small target [12] and more practical work with TorusDesktop [4].

Targets of only a few pixels are notoriously difficult to acquire. Despite many attempts at facilitating pointing, the reasons for this difficulty are poorly understood. We confirm a strong departure from Fitts’ Law for small target acquisition using a mouse and investigate three potential sources of problems: motor accuracy, legibility, and quantization. We find that quantization is not a problem, but both motor and visual sizes are limiting factors. This suggests that small targets should be magnified in both motor and visual space to facilitate pointing. Since performance degrades exponentially as targets get very small, we further advocate the exploration of uniform, target-agnostic magnification strategies. We also confirm Welford’s 1969 proposal that motor inaccuracy can be modeled by subtracting a “tremor constant” from target size. We argue for the adoption of this model, rather than Fitts’ law, when reflecting on small target acquisition.

With TorusDesktop [4], we revisited a pointing technique that allows to wrap the mouse cursor around screen edges in conventional desktop environments. Allowing the cursor to jump from one edge of the screen to the opposite one (i.e., turning the desktop into a torus) was already explored, but never studied empirically nor designed for everyday desktop usage. We have introduced a dead zone and an off-screen cursor feedback that ease the use of this technique. We also conducted three controlled experiments to refine the design and evaluate its performance. Our results suggest clear benefits in several conditions, but also some potential limitations due to users’ over-estimation of cursor wrapping advantages. An implementation of TorusDesktop for the Mac OS X desktop can be downloaded for free at [http://insitu.lri.fr/TorusDesktop](http://insitu.lri.fr/TorusDesktop).

We continued our work on wall-sized displays, focusing on the study of high-level tasks such as pan-zoom navigation (Figure 10), that have received little attention. Indeed, while pointing on this type of display has been studied extensively, it remains unclear which techniques are best suited to perform multiscale navigation in these environments. Building upon empirical data gathered from studies of pan-and-zoom on desktop computers and studies of remote pointing, we identified three key factors for the design of mid-air pan-and-zoom techniques: uni- vs. bimanual interaction, linear vs. circular movements, and level of guidance to accomplish the gestures in mid-air. After an extensive phase of iterative design and pilot testing, we ran a controlled experiment aimed at better understanding the influence of these factors on task performance. This work received a best paper award at CHI 2011 [6].

On the opposite side, we have studied small displays such as the ones smartphones are equipped with. One major challenge with this type of device is to make the user able to interact in parallel with both the device and other artefacts in his environment (e.g., giving a phone call while holding a paper document). The Swiss Army Menu (SAM) [27] is a radial menu that enables a very large number of functions accessible via small thumb movements. The design of SAM relies on four different kinds of items, support for navigating in hierarchies of items and a control that requires only the thumb of the hand that holds the device. SAM can offer a set of functions so large that it would typically have required a number of widgets that could not have been displayed in a single viewport at the same time.
Interaction and Visualization - New Results - Project-Team IN-SITU

Figure 10. Panning and zooming in Spitzer’s 396 032 x 12 000 = 4.7 billion pixels images of the inner part of our galaxy.

The different interaction techniques presented above are aimed at improving the control within a given representation. As a complement, we have also worked on improving the user’s experience by providing him with two (or more) representations of the data he is interacting with and ways to transition between these representations.

Gliimpse [17] is a quick preview technique that smoothly transitions between document markup code (HTML, wiki markup or LaTeX) and its visual rendering (see Figure 11). It allows users to regularly check the code they are editing in-place, without leaving the text editor. Gliimpse can complement classical preview windows by offering rapid overviews of code-to-document mappings and leaving more screen real-estate. As the technique smoothly show the links between the code and the rendered result, it can also help to learn how complex markup code will result in the final document (e.g., HTML tables or LaTeX formulae).

In collaboration with University of Toronto and OCAD University, we designed a novel visualization technique called ChronoLenses [14], aimed at supporting users in time-series visual exploration tasks. ChronoLenses perform on-the-fly transformation of the data points in their focus area, tightly integrating visual analysis with user actions, and enabling the progressive construction of advanced visual analysis pipelines, supporting tasks that require visualizing derived values, identifying correlations, or discovering anomalies beyond obvious outliers.

We further explored user understanding of data presented in Dual-Scale data charts, charts that incorporate two different data resolutions into one chart in order to emphasize data in regions of interest (focus) or to enable the comparison of data from distant regions (context) [13]. In collaboration with researchers from inria AVIZ, we presented a unified description of different Dual-Scale data charts, and we compared them in terms of user understanding using elementary graphical perception tasks, such as comparing lengths and distances. Our study suggests that cut-out charts which include collocated full context and focus are the best alternative, and that superimposed charts in which focus and context overlap on top of each other should be avoided.

6.2. Research Methods
Participants: Caroline Appert, Michel Beaudouin-Lafon, Anastasia Bezerianos, Olivier Chapuis, Jérémie Garcia, Stéphane Huot, Ilaria Liccardi, Wendy Mackay [correspondant], Emmanuel Pietriga.

InkSplore [19]. We conducted three studies with contemporary music composers at IRCAM. We found that even highly computer-literate composers use an iterative process that begins with expressing musical ideas on paper, followed by active parallel exploration on paper and in software, prior to final execution of their ideas as an original score. We conducted a participatory design study that focused on the creative exploration phase, to design tools that help composers better integrate their paper-based and electronic activities. We then developed InkSplorer as a technology probe that connects users? hand-written gestures on paper to Max/MSP and OpenMusic. Composers appropriated InkSplorer according to their preferred composition styles, emphasizing its ability to help them quickly explore musical ideas on paper as they interact with the computer. We conclude with recommendations for designing interactive paper tools that support the creative process, letting users explore musical ideas both on paper and electronically.

Wikibook: [22]. With the Wikibook project [22] we investigate how Wikibooks authors collaborate to create high-quality books. We combined Information Retrieval and statistical techniques to examine the complete multi-year lifecycle of over 50 high-quality Wikibooks. We found that: 1. The presence of redundant material is negatively correlated with collaboration mechanisms; 2. For most books, over 50% of the content is written by a small core of authors; and 3. Use of collaborative tools (predicted pages and talk pages) is significantly correlated with patterns of redundancy. Non-redundant books are well-planned from the beginning and require fewer talk pages to reach high-quality status. Initially redundant books begin with high redundancy, which drops as soon as authors use coordination tools to restructure the content. Suddenly redundant books display sudden bursts of redundancy that must be resolved, requiring significantly more discussion to reach high-quality status. These findings suggest that providing core authors with effective tools for visualizing and removing redundant material may increase writing speed and improve the book’s ultimate quality.

6.3. Engineering of interactive systems

Participants: Caroline Appert, Michel Beaudouin-Lafon, Olivier Chapuis, James Eagan, Tony Gjerlufsen, Stéphane Huot, Wendy Mackay, Clemens Nylandsted Klokmose, Emmanuel Pietriga [correspondant], Clément Pillias, Romain Primet.

We started working on jBricks [24], a Java toolkit enabling the exploratory prototyping of interaction techniques and rapid development of post-WIMP applications running on cluster-driven interactive visualization platforms such as the WILD wall display (Section 7.1). Research on cluster-driven wall displays has mostly focused on techniques for parallel rendering of complex 3D models. There has been comparatively little research effort dedicated to other types of graphics and to the software engineering issues that arise when prototyping novel interaction techniques or developing full-featured applications for such displays. To fill this gap, jBricks integrates a high-quality 2D graphics rendering engine and a versatile input configuration module into a coherent framework, hiding low-level details from the developer. The goal of this framework is to ease the development, testing and debugging of interactive visualization applications for wall-sized displays. It also offers an environment for the rapid prototyping of novel interaction techniques and their evaluation through controlled experiments, such as the one we recently conducted about mid-air pan-and-zoom techniques for wall-sized displays (see Section 6.1).

We developed the Shared Substance framework for multisurface interaction [20]. It is based on Substance, which implements a novel programming model called data orientation that separates functionality from data. Shared Substance extends Substance to distributed environments. It makes distribution explicit so that the programmer can dynamically add, reconfigure and remove components at runtime. An application built with Shared Substance is a collection of processes called environments that run on different machines. Environments are discovered dynamically and can appear and disappear at any time. Each environment contains a hierarchical data structure that can be shared, in whole or in part, with other environments. Sharing can be done through replication or mounting, which entail different performance trade-offs. Shared Substance also includes the Instrumental Interaction Kit (IIK) to facilitate the development of instruments in a multisurface environment [34], [42]. We have used Shared Substance to develop several applications.
for our WILD multisurface environment: Substance Canvas manages a virtual canvas that can span multiple interactive surfaces managed by different computers, such as the tiled display, interactive table and users’ laptops of the WILD room; Content can be added to the canvas from various an extensible set of sources, including live applications using Scotty (see below). Substance Grise wraps an existing application for displaying 3D brain scans into a Shared Substance environment; This allows us to run 64 copies of the application, each showing a different brain scan, and synchronize the 3D orientation of the scans using a brain prop that the user turns in his hand.

We explored the notion of user interface programming at run-time to create more malleable software [18]. Rather than creating a new user interface toolkit or supporting the customization of an interactive application from outside, we explored how well-defined hooks and a few high-level constructs could allow a programmer to modify an application “from inside”, i.e. using code that is dynamically loaded by the application at run-time. Compared with existing approaches, this supports deep customization that involve the behavior of the application, not just the surface of its user interface. The Scotty prototype implements run-time interface programming in the Mac OS X environment for any application written with the native Cocoa framework. We have used Scotty to distribute the user interface of an application over multiple devices, e.g. to move the Print button of an application onto an iPhone so the user can safely print while physically close to the printer; to replace a tool palette in an application with a toolglass; to check for the presence of attachments in an email application before sending an email; to add subtitles to a video viewer that does not have this functionality. We have also used Scotty in connection with the work on Shared Substance (see above) to teleport a live vector-based representation of a running applications to the WILD wall display. The advantage of this approach over, e.g., VNC, is that the content is properly scaled, taking advantage of the full resolution of the wall.
6. New Results

6.1. Modeling

6.1.1. Reassembly

Participants: Nicolas Mellado, Patrick Reuter, Gaël Guennebaud, Pascal Barla, Christophe Schlick.

In the context of cultural heritage, 3D laser scanning and photogrammetric 3D acquisition of broken content is becoming increasingly popular, resulting in large collections of detailed fractured archaeological 3D objects that have to be reassembled virtually. We recently investigated a semi-automatic reassembly approach for pairwise matching of digital fragments, that makes it possible to take into account both the archeologist’s expertise, as well as the power of automatic geometry-driven matching algorithms. In order to increase matching efficiency and robustness, we currently focus on shape analysis with higher level representation to guide ICP-like algorithms.

6.2. 3D Data Rendering and Visualization

6.2.1. Soft shadows

Participant: Gaël Guennebaud.

Shadows are a fundamental visual effect which both increase the level of realism of a 3D scene, and help to identify spatial relationships between objects. This latter observation makes them particularly important in the context of interactive 3D applications. Generating high quality soft shadows in real-time is still an open challenge. In the continuity of our previous collaboration with the State Key Lab of CAD&CG of Zhejiang University (China) [39], we developed a perceptually based metric dedicated to the prediction of ideal shadow map resolutions [16]. This metric allows us to adaptively generate shadow map tiles. As a result, we managed to render wide and complex exterior scenes with high quality while maintaining high performance (see figure 2).

![Figure 2. Our soft shadow rendering system generates adaptive shadow map tiles and can therefore render softer shadows (right @ 25 fps) faster than their hard shadow counterpart (left @ 15 fps).](image)

6.2.2. Synthesis and control of breaking waves

Participants: Nicolas Maréchal, Pascal Barla, Gaël Guennebaud, Patrick Reuter.
Modeling complex breaking waves over arbitrary bathymetry is a tedious problem. Currently, most of the existing methods are based on physical simulations by solving the Navier-Stokes equation. Controlling the shape of breaking waves is almost impossible with such approaches, and the simulation does not run in real-time. In order to overcome these limitations, we propose a phenomenological approach based on a real-time simulation using Airy’s wave theory. Our system handles phenomena such as shoaling, refraction and grouping (see figure 3), and the rendering style can be adapted by the user.

Figure 3. Breaking waves generated using our system over an arbitrary bathymetry. 2D wave shape profiles are shown for better legibility.

6.2.3. Analysis and visualization of surface relief

Participants: Lucas Ammann, Pascal Barla, Gaël Guennebaud, Xavier Granier, Patrick Reuter.

Given a base surface with relief, we developed an analysis technique that leverages the complexity found in detailed 3D models for illustrative shading purposes. The key originality of our approach is to extract the relief features such as concavities, convexities and inflections at multiple scales and directions using local cubic-polynomial fitting. We use this information to guide a variety of shading techniques. Our approach is parametrization-free and meshless, allowing for a wide variety of applications ranging from scientific visualization to special effects for the movie industry.

6.3. Expressive Rendering

Figure 4. Conveying visual information through expressive rendering
6.3.1. Line-based Rendering

**Participants:** Pascal Barla, Jiazhou Chen, Xavier Granier, Christophe Schlick.

We have introduced [18] a new technique called Implicit Brushes to render animated 3D scenes with stylized lines in real-time with temporal coherence. An Implicit Brush is defined at a given pixel by the convolution of a brush footprint along a feature skeleton; the skeleton itself is obtained by locating surface features in the pixel neighborhood. Features are identified via image-space fitting techniques that not only extract their location, but also their profile, which permits to distinguish between sharp and smooth features. Profile parameters are then mapped to stylistic parameters such as brush orientation, size or opacity to give rise to a wide range of line-based styles. This work has won the 3rd best paper award at Eurographics annual conference.

6.3.2. Shape Depiction through Shading

**Participants:** Pascal Barla, Xavier Granier, Christophe Schlick.

Recently, a number of techniques have been proposed to exaggerate the depiction of shape through the shading of 3D objects. However, existing methods are limited to a single type of material, simple light sources, and they give a fake percept where 3D shape seems to be flattened or embossed, or produce temporal artifacts. We have recently shown that adjusting lighting amplitude for each direction (Radiance Scaling [38], selected as a best paper at I3D 2010 and extended as a TVCG journal paper [17]) may enhance the shape depiction. The technique has been ported to Meshlab (http://meshlab.sourceforge.net/).

6.3.3. Dynamic Expressive Shading Primitives

**Participant:** Pascal Barla.

Shading appearance in illustrations, comics and graphic novels is designed to convey illumination, material and surface shape characteristics at once. Moreover, shading may vary depending on different configurations of surface distance, lighting, character expressions, timing of the action, to articulate storytelling or draw attention to a part of an object. We have developed [31] a method that imitates such expressive stylized shading techniques in dynamic 3D scenes, and which offers a simple and flexible means for artists to design and tweak the shading appearance and its dynamic behavior. The key contribution of our approach is to seamlessly vary appearance by using a combination of shading primitives that take into account lighting direction, material characteristics and surface features.

6.3.4. Non-Uniform Compositing of Styles

**Participants:** Jiazhou Chen, Xavier Granier.

In order to investigate how the composition of different styles may help in directing user attention, we have developed [22] a non-uniform composition that integrates multiple rendering styles in a picture driven by an importance map. This map, either issued from salience estimation or designed by a user, is introduced both in the creation of the multiple styles and in the final composition. Our approach accommodates a variety of stylization techniques, such as color desaturation, line drawing, blurring, edge-preserving smoothing and enhancement.

6.4. Interaction

6.4.1. Toucheo: Multitouch + Stereo

**Participants:** Martin Hachet, Benoit Bossavit, Aurélie Cohé.
We propose a new system that efficiently combines direct multitouch interaction and 3D stereoscopic visualization (see Figure 5). In our approach, the users interact by means of simple 2D gestures on a monoscopic touchscreen, while visualizing occlusion-free 3D stereoscopic objects floating above the surface at an optically correct distance. By coinciding the 3D virtual space with the physical space, we produce a rich seamless workspace where both the advantages of direct and indirect interaction are jointly exploited. In addition to standard multitouch gestures and controls (e.g., pan, zoom, and standard 2D widgets) from which we take advantage, we have designed a dedicated multitouch 3D transformation widget. This widget allows the near-direct control of rotations, scaling, and translations of the manipulated objects. To illustrate the power of our setup, we have designed a demo scenario where participants reassemble 3D virtual fragments. This scenario, as many others, takes benefit of our proposal, where the strength of both multitouch interaction and stereoscopic visualization are unified in an innovative and relevant workspace [19][25]. See highlights.

Figure 5. Toucheo

6.4.2. Touch-based interaction

Participants: Jérémy Laviole, Aurélie Cohé, Martin Hachet.

We have continued exploring 3D User Interfaces for [multi-]touch screens. In particular, in [26], we conducted a user study to better understand the impact of directness on user performance for a RST docking task, for both 2D and 3D visualization conditions. We have also designed a new 3D transformation widget, called tBox, that can be operated easily and efficiently from simple gestures on touch-screens. In our approach, users apply rotations by means of physically plausible gestures, and we have extended successful 2D tactile principles to the context of 3D interaction [23].

6.4.3. Immersive environments

Participant: Martin Hachet.

We have continued working on immersive environments. In particular, with the "Digital Sound" group of LaBRI, we have studied how sound processes should be visualized in immersive setups [13]. Another collaboration is with the REVES Inria project-team, where we have explored how to design 3D pieces of architecture in a CAVE [21].

6.4.4. Brain-Computer Interaction

Participant: Fabien Lotte.

With Fabien Lotte joining the IPARLA team as a research scientist in January 2011, a new research topic related to interaction is being explored: Brain-Computer Interfaces (BCI). BCI are communication systems that enable its users to send commands to the computer by means of brain activity only, this activity being generally measured using ElectroEncephaloGraphy (EEG). This is therefore a new way to interact with computers and interactive 3D applications. In this area, we have explored new techniques to analyze and process EEG signals in order to identify the mental state of the user [14][20]. This has led to improved robustness and mental state recognition performances. Another challenge in BCI is that it requires the collection of several examples
of EEG signals from the user in order to calibrate the system. This makes the calibration step long and inconvenient. In order to alleviate this problem, we have proposed to generate artificial EEG signals from a few EEG signals already available. Our evaluations have shown that it can indeed significantly reduce the calibration time [28]. Together with the Inria VR4I-team, we have also explored the use of a new mental state to drive a BCI: attention and relaxation states. We have shown that it is indeed possible to identify relaxation and concentration in EEG signals, and that it can be used to drive a BCI [24]. Finally, we have critically analyzed the usefulness and potential of BCI for 3D video games [27].

6.4.5. Tangible user interfaces

Participant: Patrick Reuter.

Tangible user interfaces have proven to be useful for the manipulation of 3D objects, such as for selection and navigation tasks, and even for deformation tasks. Deforming 3D models realistically is a crucial task when it comes to study the physical behavior of 3D objects, for example in engineering, in sculpting applications, and in other domains. With recent progress in physical deformation models and the increasing computing power, physically-realistic deformation simulations can now be driven at interactive rates. Consequently, there is an increasing demand for efficient and user-friendly user interfaces for physically-realistic deformation in real-time. We designed a general concept for designing physically-realistic deformations of 3D models with a tangible user interface, and instantiated our concept with a concrete prototype using a passive tangible user interface that incarnates the 3D model and that runs in real-time [32].
6. New Results

6.1. Motion Sensing and analysis

6.1.1. Sensing human activity for detecting falling motions

Participants: Franck Multon [contact], Richard Kulpa, Anthony Sorel, Edouard Auvinet.

Sensing human activity is a very active field of research, with a wide range of applications ranging from entertainment and serious games to personal ambient living assistance. MimeTIC aims at proposing original methods to process raw motion capture data in order to compute relevant information according to the application.

In personal ambient living monitoring, we have collaborated with University of Montreal, Department of Computer Science and Operations Research (DIRO) which main activity is biomedical engineering. A co-supervised student is addressing two complementary problems: detecting people falling in everyday environment and providing easy-to-use clinical gait analysis systems for early detection of potential risks of falling. These two problems have been addressed by reconstructing the visual hull of a subject according to synchronized classical of depth cameras. As visual hull is based on videos it’s subject to occlusions which generally occur in natural environment, such as a room with furniture. We have adapted the classical visual hull algorithm in order to be less sensible to occlusions. We also have proposed an index based on 3D silhouette vertical distribution which enhance this property to tackle occlusion problems [1]. This index is based on a ratio: the volume above a given threshold divided by the total body volume. It has been successfully applied to dozens of falling scenarios involving natural occlusions with furniture. The second problem consists in extracting relevant indexes in gait that could enable clinicians to identify elderly people who have a risk of falling. Classical indexes are based on gait regularity and asymmetry in dual tasks protocols (such as walking while counting downward). 3D silhouettes intrinsically contain all the required information in a unique representation contrary to multi-point motion capture systems. However extracting the relevant information from 3D volumes is complex. We have proposed an original approach based on statistical analysis of the volumes in order to compute indexes for gait asymmetry while simply using 3 depth-cameras (Microsoft Kinects) [1] (see figure 4).

![Figure 4. 3D silhouettes reconstructed with three depth-cameras in order to analyze gait asymmetry.](image-url)
In entertainment and serious games, the problem is different as we need to accurately now the action performed by the user in order to react in a convenient manner. Collaboration with Artefacto Company enabled us to develop such motion recognition methods in serious games scenarios. Given motion capture data provided by an optical motion capture system lead to large state vectors in which the relevant information is hidden. Mixture of Gaussians is generally used as an input of Hidden Markov Models to recognize a motion according to this raw data. To simplify, features are generally introduced in order to capture the relevant geometrical property of the motion with either general information (such as joint angles or Cartesian positions) or application-specific information. The former type of information has the advantage to be generic but leads to recognizers that are very sensitive to style and morphology variations. We have proposed a new generic feature based on morphology-independent representation that enables to tackle this problem (submitted to Eurographics2012). The recognition rate is above 75% for very similar upper-limb motions (see figure 5) while classical methods fail to recognize the same motions (recognition rate below 50%).

![Image](image_url)

*Figure 5. The motion of a user equipped with reflective markers is automatically recognized with a HMM system based on morphology-independent features.*

### 6.1.2. The Joyman: a novel immersive locomotion device for virtual environments

**Participant:** Julien Pettré [contact].

We proposed a novel interface called Joyman, designed for immersive locomotion in virtual environments [18]. Whereas many previous interfaces preserve or stimulate the users proprioception, the Joyman aims at preserving equilibrioception in order to improve the feeling of immersion during virtual locomotion tasks. The proposed interface is based on the metaphor of a human-scale joystick. The device has a simple mechanical design that allows a user to indicate his virtual navigation intentions by leaning accordingly. We also propose a control law inspired by the biomechanics of the human locomotion to transform the measured leaning angle into a walking direction and speed - i.e., a virtual velocity vector. We aim at using this interface to enable natural interaction with virtual humans with low-cost devices. The Joyman is patented and was presented at the Emerging Technologies, Siggraph Asia, Hong-Kong [20].

These are joint results with the VR4i team (Anatole Lécuyer and Maud Marchal).

### 6.2. VR and Sports

**Participants:** Richard Kulpa, Benoit Bideau [contact], Brault Sébastien, Burns Anne-Marie.
In the past, we have worked on the interaction between two opponents in handball. We have designed a framework to animate virtual throwers in a reality center and to analyze the gestures of real goalkeepers whose objective was to intercept the corresponding virtual balls. This VR framework was then validated by showing that behaviors in real and virtual environment were similar. These works have been extended by using perception-action coupling and perception-only studies to evaluate the anticipation of opponents. In order to evaluate the importance of perceived parameters, the ball and/or the character animation was successively hidden to determine their importance and the same kind of study was done on the graphical level of details.

These works have been extended to the study of deceptive movements and gaps evaluation in rugby. Combining perceptual analysis based on the use of cutoffs with biomechanical analysis, we have extracted important kinematic information that could explain differences between experts and novices. Indeed, thanks to the cutoffs, it is possible to determine how early each of these two levels of practice can perceive the correct final direction of the opponent. Then this information is correlated to kinematical parameters of this player. Finally, we have embedding these knowledge on the evaluation of novices and expertsd to create models of rugby defenders. We are currently working on integrating these models in a VR experiment in which the real user is this time the attacker and our model the virtual defender.

Concurrently, studies are experimented to determine if VR can be used for training in sports [9]. The first step was to compare if trainees learned the same way in real situation, facing a video of the lesson or facing a virtual teacher that is animated from the motions of the real course. Based on evaluation of an expert, the results showed that the three groups evolved the same way and reached the same level of practice. The second step is then to have more experts to complete the evaluation but also to combine these results with objective analyses based on kinematics data.

This work is partially funded by the Biofeedback project (DGCIS "Serious Gaming" project) of the M2S laboratory, University Rennes 2. Its goal was to create a training tool that can be used and configured by coaches in order to train athletes to repetitive motions such as katas in karate. The evaluation is made by an automatic module that compares the kinematics of the trainee to a database of expert movements.

6.3. Crowds

Participants: Julien Pettré [contact], Richard Kulpa, Anne-Hélène Olivier, Samuel Lemercier, Yijiang Zhang, Jonathan Perrinet.

6.3.1. Perception of collision in crowds

We designed a level-of-detail (LOD) selection function to determine where collision avoidance constraints in crowd simulation can be relaxed without being perceived by spectators [4]. Collision avoidance is probably the most time consuming parts of crowd simulator. However, when only believable results are required, we argue that visually similar results can be obtained a low computational costs based on macroscopic crowd simulation. Based on a perception study, we determined the conditions for collision to be or not to be detected. We discovered that the camera tilt angle was playing a great effect on perception, whereas distance to camera (usually used in previous works) was only the third most important factor to be considered.

6.3.2. Mixed Reality Crowds

In the task of making virtual crowds and real people interact together, we explore a mixed reality solution [22]. The seamless integration of virtual characters into dynamic scenes captured by video is a challenging problem. In order to achieve consistent composite results, both the virtual and real characters must share the same geometrical space and their interactions must adhere to the physical coherence criteria. One essential question is how to detect the motion of real objects - such as real characters moving in the video - and how to steer virtual characters accordingly to avoid unrealistic collisions. We propose an online solution. First, by analysis of the input video, the motion states of the real objects are recovered into a common world 3D coordinate system. Meanwhile, a simplified accuracy measurement is defined to represent the confidence of the motion estimate. Then, under the constraints imposed by the real dynamic objects, the motion of virtual characters are accommodated by a uniform steering model. The final step is to merge virtual objects back the
real video scene by taking into account visibility and occlusion constraints between real foreground objects and virtual ones.

6.3.3. Experiments on crowds

Evaluating crowd simulation models is a difficult task. In the frame of the ANR PEDIGREE project, we put in a lot of effort to perform experiments on groups of walking people in order to dispose of a reference database on groups motion. In order to obtain high-quality data, we measure people locomotion by using optoelectronic motion capture systems. In 2011, we starting obtaining detailed analysis on such motion after large efforts on motion analysis and processing. We had to develop dedicated reconstruction techniques because of the challenging conditions in which we performed our motion capture [12]. We submitted two papers on following modeling and simulation stages (submitted to Eurographics 2012 and Physical Review E).

6.4. Interactive Virtual Cinematography

**Participants:** Marc Christie [contact], Christophe Lino.

The domain of Virtual Cinematography explores the operationalization of rules and conventions pertained to camera placement, light placement and staging in virtual environments. In 2011, two major challenges were tackled (i) the proposition of intelligent interactive assistants integrating users in the process of selecting viewpoints and editing a virtual movie, with the capability of adapting to the user choices, and (ii) the design and implementation of evaluation functions for precisely ranking the quality of viewpoints of a virtual 3D environment.

Our intelligent assistant is designed around (i) an intelligent cinematography engine that can compute, at the request of the filmmaker, a set of suitable camera placements (called suggestions) for starting a shot, representing semantically and cinematically distinct choices for visualizing the current narrative, considering established cinema conventions of continuity and composition along with the filmmaker’s previous selected suggestions, and also his or her manually crafted camera compositions, by a machine learning component that adapts shot editing preferences from user-created camera edits, (ii) a user interface, where the suggestions are presented as small movie frames, arranged in a grid whose rows and columns correspond to visual composition properties of the suggested cameras, and (iii) a motion-tracked camera system that makes the user able of modifying the low-level parameters of the camera in shots in the same way a real operator would.

The result of this work [16] is a novel workflow based on interactive collaboration of human creativity with automated intelligence that enables efficient exploration of a wide range of cinematographic possibilities, and rapid production of computer-generated animated movies. A full prototype has been built and demonstrated at ACM Multimedia conference [15] as well as ParisFX conference. A patent protecting this technology is currently under evaluation [25].

The second challenge is related to the design of efficient and precise metrics for measuring the quality of viewpoints. For efficiency, we have proposed parallel GPU-based evaluation techniques for the estimation of multiple viewpoints [8] coupled within a Genetic Algorithm (Swarm Particle Optimization) to rapidly explore the space of possible viewpoints. For preciseness, we have designed a large range of quality functions relative to screen composition and transition between shots, and employed these functions to either automatically generate movies from actions occuring in the virtual environment [13] or interactively generating movies by letter the users select best shots and best transitions between shots [14].

Finally we have been exploring the use of tactile devices to the interactive construction of narratives following Prop’s computational model of stories [10].

6.5. Biomechanics and Motion Analysis

6.5.1. Balance in highly dynamic situations

**Participants:** Franck Multon [contact], Ludovic Hoyet.
Balance is a key problem in humans as people stand on two feet which leads to a small base of support area compared to the overall body volume. This unstable state has been widely analyzed in static situation but is still difficult to understand when velocity and acceleration reach ineligible values. We thus have proposed an experimental protocol in order to evaluate if criteria published in the literature for specific motions could be generalized to any dynamic motions (see figure 7). To this end, each studied criterion was tested on various dynamic motions and the number of false falling alarms was reported in each case: the number of frames where the criterion detects loss of balance while the subject is actually balanced. The tested criteria where: the projection of the center of mass on the ground which should remain in the base of support, the Zero Moment Point widely used in robotics, the Zero Rate of Angular Momentum, the Foot Rotation Index and the extrapolated center of mass which was introduced in biomechanics recently. The results demonstrate that none of the criteria succeeded in correctly predicting loss of balance in highly dynamic motions [7]. It thus demonstrate the need to continue some fundamental work on this topic which is a key problem in many applications, including robotics, detection and prevention of falls in the Elderly, understanding performance in sports, improving realism in virtual human simulation...

6.5.2. Interaction strategies between two walkers to avoid collision
**Participants:** Armel Crétual, Julien Pettré [contact], Anne-Hélène Olivier, Jan Ondrej, Antoine Marin.

In the everyday life situation where two humans walk in the same nearness, each can be considered as a moving obstacle for the other one. They adapt their locomotion with respect to this external disturbance to avoid any collision. Collision avoidance between two humans has been largely neglected in the literature despite it lets us expect specific interactions. The main question we raised was to identify the conditions that induce avoidance manoeuvres in locomotor trajectories: what are the relations between the respective positions and velocities which induce motion adaptations? To answer this question, we proposed an original experiment: thirty participants were asked to walk two-by-two in a motion captured area. We assigned them locomotion tasks in order to provoke varied situations of potential future collisions (see figure 8 a). Following the hypothesis of a reciprocal interaction, we suggested a variable which is common to both of the walkers, the Minimum Predicted Distance (MPD), to predict the actual presence of physical interactions as well as to describe their properties. At each instant t, MPD was computed as the distance the walkers would meet if they did not perform motion adaptation after this instant t. Results showed that walkers adapted their motions only when required, i.e., when initial MPD was too low (<1m). We concluded that human walkers are able to accurately estimate future crossing distance and to mutually adapt it. The evolution of MPD enabled decomposing collision avoidance into 3 successive phases: observation, reaction, and regulation (see figure 8 b). Respectively, these phases corresponded to periods of time when, first, MPD was constant, second, increased to acceptable values by motion adaptation and, third, reached a plateau and even slightly decreased. This final phase demonstrates that collision avoidance is actually performed with anticipation. Future work is needed to inspect individual motion adaptations and to relate them with variation of MPD.

![Figure 8](image.png)

*Figure 8. a) Experimental set-up to study collision avoidance between 2 walkers. b) Mean MPD evolution for trials for which initial MPD is below 1m. Interaction follows 3 successive phases: observation, reaction and regulation phases. Normalized time was computed between the time when participants were able to see each other to the time they crossed.*

### 6.5.3. Quantification of pathological gait in adults

**Participants:** Armel Crétual [contact], Kristell Bervet.

Quantifying gait deviation is still a challenge in adults patient follow-up within a rehab process. This quantification can be done on several levels. Among them, the most useful for practitioners are surely kinematics and muscular activation. On the first one, Gillette Gait Index (GGI) has now become a common tool in rehabilitation centers to assess gait abnormalities. However, one limitation of this index is that it is based on some peak values and is thus sensitive to measurement noise. A new index, the Gait Deviation Index (GDI) which is based on joints angles patterns has been developed by the same team to avoid this problem. Nevertheless, both of them have only been validated in children with cerebral palsy. On the second level, no
satisfying global index has yet been developed. The first part of our study was to validate the GDI in adults. From a database of 74 healthy subjects and 48 patients we did demonstrate that GDI is a relevant index to quantify kinematic gait pattern in adults. Then, we developed a new index called KeR-EGI (for Kerpace-Rennes EMG-based Gait Index) which accounts for the muscular activation patterns of patients. There also, results were conclusive relying on a good correlation between GDI and KeR-EGI. Finally, our recommendation to practitioners was to use both of these two index as they account for complementary aspects of pathology. This allow to better understand if gait disorder is more due to neurological injury or on the opposite to mechanical constraints such as joint stiffness.

6.5.4. Modeling gesture in sports: tennis serve

Participants: Nicolas Bideau [contact], Guillaume Nicolas, Benoit Bideau, Caroline Martin, Richard Kulpa.

Most experimental studies on tennis focus on the segmental coordination in connection with the ball speed, but do not consider the resulting traumatisms. To this end, we currently develop an inverse dynamics modeling approach, based on musculoskeletal parameters. As a first step to this work, we calculated the joint constraints on the upper limb in the tennis serve, for professional and regional players. Eleven high level, professional players were compared to seven regional players during this specific motion. Each player was equipped with 42 reflective markers and tennis serve was analyzed using an optoelectronical system composed of 12 infrared cameras cadenced at 300Hz. For each player, values of force, power and internal work (in absolute value) were calculated for the three joints (shoulder, elbow, wrist) using a musculoskeletal model. The results showed that professional players produced higher power and internal work for each joint in comparison with the regional players. Results also showed a decrease in the values of internal work from the distal (wrist joint) to the proximal joint (shoulder joint). These results may explain shoulder pain in tennis, which is commonly depicted for high-level players in epidemiological studies. A first perspective to this work is to better take individual parameters (inertia, muscle parameters, pre-constraint, etc.) into account. Another perspective is to test various objective functions in order to predict which parameter is optimized during tennis serve.

6.5.5. Modeling gesture in sports: fin swimming

Participants: Nicolas Bideau [contact], Guillaume Nicolas, Benoit Bideau, Richard Kulpa.

In swimming, experimental approaches are commonly used to analyze performance. However, due to obvious limitations in experimental approaches (impossibility to standardize any situations etc.), it is difficult to characterize surrounding fluid. To overcome this limitation, we currently develop analysis, modeling and simulation of aquatic locomotion, using CFD computer simulation and new methods based on animation of virtual characters.

A first application of this topic enables to evaluate the influence of swim fin flexibility on efficiency during swimming based on a CFD structure interaction model. Finite elements simulations are carried out for various material properties and various prescribed kinematics. Besides the significant effect of flexibility on propulsive forces, the results indicate that the propulsive efficiency is greatly influenced by the stroke frequency and the initial angle of attack. For the selected material properties, the results show that efficiency increases from 3.6 percents to 11.9 percents when the stroke frequency is increased from 0 to 1.7 Hz. Moreover efficiency is clearly increased from 5.0 percents to 24.2 percents when increasing the angle of attack from 0 to 45 degrees. Therefore, an interesting prospect of the present work could be an enhancement of the design of better performing swim fins.

A second application of this topic related to aquatic propulsion deals with a new method to evaluate cross-sectional area based on computer animation of swimming. Indeed, reducing cross sectional area (CSA) during starts and turns is a key part of performance optimisation. Different methods have been used to obtain this parameter without any standard: total human body volume to the power 2/3, wetted area or frontal area based on planimetry technique (PT). These different methods can lead to discrepancies in drag values. Recently, we used two synchronized camcorders to evaluate drag parameters during the different phases of an undulatory stroke cycle.
However, such a technique needs accurate synchronization and calibration of the different camcorders views. The aim of this study is to provide a new method based on animation of virtual characters to obtain instantaneous cross-sectional area in an undulatory stroke cycle. Its main advantage is to obtain cross-sectional area as well as biomechanical analysis with a single camcorder in a sagittal plan and without space calibration. From this, we intend to better understand swimming hydrodynamics and the way CSA influences active drag. More generally, this approach has been designed to provide new practical insights into swimming analysis protocols.

6.6. Path planning and environment analysis

6.6.1. Space-Time planning in dynamic environments

Participants: Fabrice Lamarche [contact], Thomas Lopez.

When automatically populating 3D geometric databases with virtual humanoids, modeling the navigation behavior is essential since navigation is used in most exhibited behaviors. In many application fields, the need to manage navigation in dynamic environments arises (virtual worlds taking physics laws into account, numerical plants in which step stools can be moved,...). This study focuses on the following issue: how to manage the navigation of virtual entities in such dynamic environments where topology may change at any time i.e. where unpredictable accessibility changes can arise at runtime. In opposition to current algorithms, movable items are not only considered as obstacles in the environment but can also help virtual entities in their navigation.

The proposed algorithm [17] splits that problem into two complementary processes: a topology tracking algorithm and a path planning algorithm. The aim of the topology tracking algorithm is to continuously detect and update topological relations between moving objects i.e. accessibility or obstruction, while storing temporal information when recurring relations are observed. The path planning algorithm uses this information to plan a path inside the dynamic environment. The coupling of those algorithms endows a virtual character with the ability to immediately use inserted / moved object to reach previously unreachable locations. Moreover, this algorithm is able to find a path through moving platforms to reach a target located on a surface that is never directly accessible.

6.6.2. Automated environment analysis

Participants: Fabrice Lamarche [contact], Carl-Johan Jorgensen.

To populate a virtual environment, modeling the navigation behavior is crucial. This behavior relies on the ability of planning a path inside a complex environment, which itself relies on an adequate representation of the environment structure. Most often, virtual environments are represented has 3D databases that are analyzed to produce data structures that are suitable for path planning and navigation. However, without any user intervention, those data structures lack of information about the nature of identified navigable zones that are crucial for navigation credibility.

We proposed an environment analysis algorithm [11] that automatically extracts a meaningful spatial representation of 3D virtual environments, suitable for spatial reasoning. This algorithm automatically differentiates indoor, outdoor and covered parts of the environment. It separates buildings into floors linked by stairs and represent floors as rooms linked by doorsteps. On this basis, we propose a natural hierarchical representation of the environment. This representation is used for spatial reasoning including zone selection and multi-criterion path planning that enhances path credibility.
6. New Results

6.1. Improvement of the force-feedback in a 1-ddl device

**Participants:** Michel Amberg, Frédéric Giraud, Betty Lemaire-Semail.

Traveling Wave Ultrasonic Motor have many advantages compared to the classical electromagnetic motors: they are lightweight, they don’t need any speed reducer and they make no noise. In a 1-ddl force feedback device, they can help to reduce the bulk size of the mechanism by simplifying the kinematic chain. However, their control has to be very precise because the torque produced is not a straightforward function of the electrical parameters. Previously, we proposed several control algorithms and we obtained good results. But at low speed, problems still remains, like a stick-slip phenomena which makes the motor producing a cogging torque.

To cope with this problem, we first proposed an accurate modeling of the motor and its torque production \[ 24 \]. We introduced a friction torque \( T_f \) which holds the non linearity of the torque production. The evolution of this friction torque has been identified through an experimental study. Then we obtained by inversion a control scheme \[ 13 \]. The basic idea is to compensate the virtual friction torque. In order to achieve a more accurate control of the torque, we proposed to identify on-line the parameters of the equation of \( T_f \).

A one-degree-of-freedom force feedback lever was built to verify the control laws. The experiment involves use of the lever of the digitracker which is free to rotate about the horizontal axis, and is presented in figure 1. In the same figure, we plotted the output torque of the motor, compared to its reference. Both are consistent, showing a good accuracy of the torque controller. To achieve that, the estimator’s parameters are time-varying.

![Image of a force feedback lever](image)

(a) (b)

*Figure 1. (a) The 1-ddl haptic device; (b) experimental run of the torque controller with its parameter estimator; in H, results are plotted in the Torque-position plane, reference is in green and measurements is in blue while C, D and G show the estimator’s parameters*

6.2. Haptic Perception of Curvature through active touch

**Participants:** Michel Amberg, Frédéric Giraud, Betty Lemaire-Semail.
Haptic perception of curvature can be achieved by passive or active finger touch. In this study we proposed a new haptic device that could independently orient, elevate and translate a flat plate. User is free to move his finger on the plate; by controlling plate’s orientation and position accordingly to the position of the finger, we can render a curved shape. The device is composed of two 6-dof haptic devices (Novint Falcon) on which we attached the plate (Figure 2). A force sensor is used to compute the position of the finger on the plate.

Several modelings have been proposed to calculate the orientation and position of the plate. We then measured how accurate simulations of curved shapes are. To achieve that work, we simulated several curved surfaces with different curvature. As the perception performance of curvature is dependent on local surface orientation, the plate was always kept tangent to a virtual shape at the contact point. We then asked people to compare simulated curvature to the real ones. We found that users are able to find the real shape (among five) corresponding to the simulated one.

6.3. Tactile input with programmable friction

Participants: Michel Amberg, Géry Casiez, Frédéric Giraud, Betty Lemaire-Semail, Paolo Olivo, Nicolas Roussel.

Our work on programmable friction relies on a particular technology we have been developing for several years. The STIMTAC is a touchpad device that supports friction reduction by means of a squeeze film effect. It uses a controlled vibration at an ultrasonic frequency with a few micrometers amplitude to create an air bearing between a user’s finger and the device’s surface. As the frequency is outside skin mechanoreceptors’ bandwidth, one does not feel this vibration but its effect on tribological contact mechanisms: the touchpad feels more slippery as the amplitude is raised.

We have used this touchpad to create Surfpad, a pointing facilitation technique that operates in the tactile domain. Experiments comparing this technique to semantic pointing and constant control-display gain with and without distractor targets clearly show the limits of traditional target-aware control-display gain.
adaptation in the latter case, and the benefits of the tactile approach in both cases [17]. Surfpad leads to a performance improvement close to 9% compared to unassisted pointing at small targets with no distractor. It is also robust to high distractor densities, keeping an average performance improvement of nearly 10% while semantic pointing can degrade up to 100%. Our results also suggest the performance improvement is caused by tactile information feedback rather than mechanical causes, and that the feedback is more effective when friction is increased on targets using a simple step function.

This year’s work on the hardware aspects of the STIMTAC resulted in a compact and quiet prototype powered by the USB cable used for data communication and supporting precise and reliable finger tracking based on multiple force sensors (Figure 3, left). Within the context of the 3DTOUCH project, efforts have also been targeted at the adaptation of the STIMTAC operating principles to off-the-shelf transparent touch sensors. Our latest prototypes demonstrate the compatibility of our approach with resistive (Figure 3, right) and capacitive technologies. In order to facilitate the design and evaluation of novel interaction techniques taking advantage of these prototypes, we have started developing a specific library, tIO, that supports all of them in a unified way.

6.4. Methods and tools to characterize, replicate and compare pointing transfer functions

Participants: Géry Casiez, Damien Marchal, Nicolas Roussel.

Transfer functions are the only pointing facilitation technique actually used in modern graphical interfaces involving the indirect control of an on-screen cursor. But despite their general use, very little is known about them. We developed EchoMouse, a device we created to characterize the transfer functions of any system, and libpointing, a toolkit that we developed to replicate and compare the ones used by Windows, OS X and Xorg [16]. We described these functions and reported on an experiment that compared the default one of the three systems. Our results show that these default functions improve performance up to 24% compared to a unitless constant CD gain. We also found significant differences between them, with the one from OS X improving performance for small target widths but reducing its performance up to 9% for larger ones compared to Windows and Xorg. These results notably suggest replacing the constant CD gain function commonly used by HCI researchers by the default function of the considered systems.

6.5. Multimodal pen input for interactive multitouch surfaces

Participant: Géry Casiez.
Touch interaction is arguably more immediate and natural in many situations, but fingers are imprecise and
difficult to write with. Alternatively, using a pen (or stylus) makes writing more natural and pointing more
precise. Luckily, this does not need to be a unilateral choice; pen and touch can be used simultaneously.
However, without non-dominant hand coordination or graphical buttons, the pen itself supports few modes.
This makes single-handed mobile usage difficult and reduces the number of combined touch and pen modes.
When frequently switching between pen-oriented modes, such as drawing, handwriting, gestures, and lasso
selection, this can hurt performance. Inferring modes is difficult, and most users prefer explicit control.
Schemes for squeezing multiple explicit modes from a pen include adding barrel buttons and classifying
pressure, tilt, barrel rotation, or grip. But these can be error-prone and ambiguous. A simple way to add a
second mode is by adding an “eraser,” a second contact point. The pencil analogy lends intuition and users
have explicit control.
Conté is a small input device inspired by the way artists manipulate a real Conté crayon. By changing which
corner, edge, end, or side is contacting the display, the operator can switch interaction modes using a single
hand. Conté’s rectangular prism shape enables both precise pen-like input and tangible handle interaction.
Conté also has a natural compatibility with multi-touch input: it can be tucked in the palm to interleave
same-hand touch input, or used to expand the vocabulary of bimanual touch. Inspired by informal interviews
with artists, we catalogue Conté’s characteristics, and use these to outline a design space. We describe a
prototype device using common materials and simple electronics. With this device, we demonstrate interaction
techniques in a test-bed drawing application [19].

Figure 4. Illustration of different modes defined by Conté: (left) freehand drawing and annotating with the corner;
(middle) revealing attribute palettes using a thick side face; (right) contextual commands using the end contact.

6.6. Perceived difficulty of pen gestures

Participants: Géry Casiez, Laurent Grisoni.

There are three primary factors which contribute to a successful gesture-based interface: the acquisition
technology, the recognizer, and the design of the gesture set. Technologies to acquire gestures, and gesture
recognition algorithms, are now quite robust and widely available. However, developing techniques and criteria
to help designers create an intuitive and easy-to-perform gesture set remain an active area of research. The
challenge is that in order to successfully integrate into an application, a gesture has to satisfy multiple criteria:
it must be unambiguously recognized; fit well with its associated function; be easy to learn and recall; and be
efficient to perform.
Our empirical results show that users perceive the execution difficulty of single stroke gestures consistently, and execution difficulty is highly correlated with gesture production time. We use these results to design two simple rules for estimating execution difficulty: establishing the relative ranking of difficulty among multiple gestures; and classifying a single gesture into five levels of difficulty. We confirm that the CLC model does not provide an accurate prediction of production time magnitude, and instead show that a reasonably accurate estimate can be calculated using only a few gesture execution samples from a few people. Using this estimated production time, our rules, on average, rank gesture difficulty with 90% accuracy and rate gesture difficulty with 75% accuracy. Designers can use our results to choose application gestures, and researchers can build on our analysis in other gesture domains and for modeling gesture performance [18].
5. New Results

5.1. Plausible Image Rendering

5.1.1. Filtering Solid Gabor Noise

Participants: Ares Lagae, George Drettakis.

Solid noise is a fundamental tool in computer graphics. Surprisingly, no existing noise function supports both high-quality anti-aliasing and continuity across sharp edges. Existing noise functions either introduce discontinuities of the solid noise at sharp edges, which is the case for wavelet noise and Gabor noise, or result in detail loss when anti-aliased, which is the case for Perlin noise and wavelet noise. In this project, we therefore present a new noise function that preserves continuity over sharp edges and supports high-quality anti-aliasing. We show that a slicing approach is required to preserve continuity across sharp edges, and we present a new noise function that supports anisotropic filtering of sliced solid noise. This is made possible by individually filtering the slices of Gabor kernels, which requires the proper treatment of phase. This in turn leads to the introduction of the phase-augmented Gabor kernel and random-phase Gabor noise, our new noise function. We demonstrate that our new noise function supports both high-quality anti-aliasing and continuity across sharp edges, as well as anisotropy. Fig. 3 shows several solid procedural textures generated with our new random-phase solid Gabor noise.

This work was presented at ACM SIGGRAPH 2011 in Vancouver and published in ACM Transactions on Graphics [18] (SIGGRAPH paper).

5.1.2. Image-Guided Weathering for Flow Phenomena

Participants: Carles Bosch, Pierre-Yves Laffont, George Drettakis.

The simulation of weathered appearance is essential in the realistic modeling of urban environments. With digital photography and Internet image collections, visual examples of weathering effects are readily available. These images, however, mix the appearance of the weathering phenomena with the specific local context. In [12], we have introduced a new methodology to estimate parameters of a phenomenological weathering simulation from existing imagery, in a form that allows new target-specific weathering effects to be produced. In addition to driving the simulation from images, we complement the visual result with details and colors extracted from the images. This methodology has been illustrated using flow stains as a representative case, demonstrating how a rich collection of flow patterns can be generated from a small set of exemplars. In Fig. 4, we show the major components required for this approach.
Figure 4. From left to right: given a photograph of a stain effect, we extract its stain degree map, we estimate the information required for the simulation (color, simulation parameters, and high frequency details), and run simulations on new scenes using this information, resulting in a wide variety of similar-looking stains.

5.1.3. Relighting Photographs of Tree Canopies

Participants: Marcio Cabral, George Drettakis.

We present an image-based approach to relighting photographs of tree canopies. Our goal is to minimize capture overhead; thus the only input required is a set of photographs of the tree taken at a time of day, while allowing relighting at any other time. We first analyze lighting in a tree canopy, both theoretically and using simulations. From this analysis, we observe that tree canopy lighting is similar to volumetric illumination. We assume a single-scattering volumetric lighting model for tree canopies, and diffuse leaf reflectance. To validate our assumptions, we apply our method on several synthetic renderings of tree models, for which we are all able to compute all quantities involved.

Our method first creates a volumetric representation of the tree using 10-12 images taken at a single time of day and use a single-scattering participating media lighting model - these photos are taken from different viewpoints, around the tree. An analytical sun and sky illumination model, namely the Preetham model, provides consistent representation of lighting for the captured input and unknown target times.

We relight the input image by applying a ratio of the target and input time lighting representations. We compute this representation efficiently by simultaneously coding transmittance from the sky and to the eye in spherical harmonics. We validate our method by relighting images of synthetic trees and comparing to path-traced solutions. We also present results for photographs, validating with time-lapse ground truth sequences. An example is shown in Fig. 5. This work was published in the IEEE Transactions on Computer Graphics and Visualization [15], and was in collaboration with the past members of REVES, N. Bonneel and S. Lefebvre.

5.1.4. Silhouette-aware Warping for Image-based Rendering

Participants: Gaurav Chaurasia, George Drettakis.
Figure 5.

Mulberry tree Top row: input images and 4 target images with corresponding times of day. Middle row: 4 resulting relit images using our approach. Bottom row: $E_{in}$ and the four $E_{tor}$ images.
Figure 6. Top: Overview of our approach starting from input images to rendered novel views along with novel view generated using current state of the art IBR. Bottom: Novel views rendered using our approach.
We have presented a novel solution for image-based rendering (IBR) of urban scenes based on image warping. IBR techniques allow capture and display of 3D environments using photographs. Modern IBR pipelines reconstruct proxy geometry using multi-view stereo, reproject the photographs onto the proxy and blend them to create novel views. The success of these methods depends on accurate 3D proxies, which are difficult to obtain for complex objects such as trees and cars. Large number of input images do not improve reconstruction proportionally; surface extraction is challenging even from dense range scans for scenes containing such objects. Our approach does not depend on dense accurate geometric reconstruction; instead we compensate for sparse 3D information by variational warping. In particular, we formulate silhouette-aware warps that preserve salient depth discontinuities. This improves the rendering of difficult foreground objects, even when deviating from view interpolation. We use a semi-automatic step to identify depth discontinuities and extract a sparse set of depth constraints used to guide the warp. On the technical side, our formulation breaks new ground by demonstrating how to incorporate discontinuities in variational warps. Our framework is lightweight and results in good quality IBR for previously challenging environments as shown in figure. 6.

**Applications.** Robust image-based rendering can be used to generate photo-realistic visual content easily which can be very useful for virtual reality applications. Commercial products like Google StreetView and Microsoft PhotoSynth use rudimentary image-based rendering for large scale visualization of cities. Our work advances the state of the art by treating the hardest class of scenes.

This work is a collaboration with Prof. Dr. Olga Sorkine (ETH Zurich). The work was published in the special issue of the journal Computer Graphics Forum [16], presented at the Eurographics Symposium on Rendering 2011 in Prague, Czech Republic.

### 5.1.5. Real Time Rough Refractions

**Participant:** Adrien Bousseau.

![Figure 7](image.png)

*Figure 7. Compared to an expensive ray-traced reference (a), our method produces plausible results in real-time (b).*

We have proposed an algorithm to render objects made of transparent materials with rough surfaces in real-time, under environment lighting. Rough surfaces such as frosted and misted glass cause wide scattering as light enters and exits objects, which significantly complicates the rendering of such materials. We introduced two contributions to approximate the successive scattering events at interfaces, due to rough refraction: First, an approximation of the Bidirectional Transmittance Distribution Function (BTDF), using spherical Gaussians, suitable for real-time estimation of environment lighting using pre-convolution; second, a combination of cone tracing and macro-geometry filtering to efficiently integrate the scattered rays at the exiting interface of the object. Our method produces plausible results in real-time, as demonstrated by comparison against stochastic ray-tracing (Figure 7).
This work is a collaboration with Charles De Rousiers, Kartic Subr, Nicolas Holzschuch (ARTIS / INRIA Rhône-Alpes) and Ravi Ramamoorthi (UC Berkeley) as part of our CRISP associate team with UC Berkeley. The work was presented at the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games 2011 and won the best paper award.

5.1.6. Single view intrinsic images

**Participants:** Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

![Figure 8. Starting from multiple views of the scene (a), our method automatically decomposes photographs into three intrinsic components — the reflectance (b), the illumination due to sun (c) and the illumination due to sky (d). Each intrinsic component can then be manipulated independently for advanced image editing applications.](image)

We introduced a new algorithm for decomposing photographs of outdoor scenes into intrinsic images, i.e. independent layers for illumination and reflectance (material color).

Extracting intrinsic images from photographs is a hard problem, typically solved with image-driven methods using numerous user indications. Recent methods in computer vision allow easy acquisition of medium-quality geometric information about a scene using multiple photographs from different views. We developed a new algorithm which allows us to exploit this noisy and somewhat unreliable information to automate and improve image-driven propagation algorithms to deduce intrinsic images. In particular, we develop a new approach to estimate cast shadow regions in the image, by refining an initial estimate available from reconstructed geometric information. We use a voting algorithm in color space which robustly identifies reflectance values for sparse reconstructed 3D points, allowing us to accurately determine visibility to the sun at these points.
This information is propagated to the remaining pixels. In a final step we adapt appropriate image propagation algorithms, by replacing manual user indications by data inferred from the geometry-driven shadow and reflectance estimation. This allows us to automatically extract intrinsic images from multiple viewpoints, thus allowing many types of manipulation in images.

As illustrated on Figure 8 , our method can extract reflectance at each pixel of the input photographs. The decomposition also yields separate sun illumination and sky illumination components, enabling easier manipulation of shadows and illuminant colors in image editing software.

This work won the best paper award at “les journées de l’AFIG (Association Francophone d’Informatique Graphique) 2011”, and will appear in the French journal REFIG [ 21 ].

5.1.7. Single view intrinsic images with indirect illumination

**Participants:** Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

Building on top of our work published in [ 21 ], we extended our intrinsic image decomposition method to handle indirect illumination, in addition to sun and sky illumination. We introduced a new algorithm to reliably identify points in shadow based on a parameterization of the reflectance with respect to sun visibility, and proposed to separate illumination components using cascaded image-based propagation algorithm. We also demonstrated direct applications of our results for the end user in widespread image editing software. This rich intrinsic image decomposition method has been submitted for publication.

5.1.8. Multi-lighting multi-view intrinsic images

**Participants:** Pierre-Yves Laffont, Adrien Bousseau, George Drettakis.

We compute intrinsic image decompositions using several images of the same scene under different viewpoints and lighting conditions. Such images can be easily gathered for famous landmarks, using photo sharing websites such as Flickr. Our method leverages the heterogeneity of photo collections (multiple viewpoints and lighting conditions) to guide the intrinsic image separation process. With this automatic decomposition, we aim to facilitate many image editing tasks and improve the quality of image-based rendering from photo collections.

This work is in collaboration with Frédo Durand (associate professor at MIT) and Sylvain Paris (researcher at Adobe).

5.1.9. Warping superpixels

**Participants:** Gaurav Chaurasia, Sylvain Duchêne, George Drettakis.

We are working on developing a novel representation of multi-view scenes in the form of superpixels. Such a representation segments the scene into semantically meaningful regions called superpixels which admit variational warps [ 16 ], thereby leveraging the power of shape-preserving warps while providing automatic silhouette and occlusion handling. This entails several contributions namely incorporating geometric priors in superpixel extraction, generating consistent warping constraints and a novel rendering pipeline that assembles warped superpixels into the novel view maintaining spatio-temporal coherence.

This work is a collaboration with Prof. Dr. Olga Sorkine (ETH Zurich).

5.1.10. State-of-the-art Report on Temporal Coherence for Stylized Animations

**Participant:** Adrien Bousseau.
Non-photorealistic rendering (NPR) algorithms allow the creation of images in a variety of styles, ranging from line drawing and pen-and-ink to oil painting and watercolor. These algorithms provide greater flexibility, control and automation over traditional drawing and painting. The main challenge of computer generated stylized animations is to reproduce the look of traditional drawings and paintings while minimizing distracting flickering and sliding artifacts present in hand-drawn animations. These goals are inherently conflicting and any attempt to address the temporal coherence of stylized animations is a trade-off. This survey is motivated by the growing number of methods proposed in recent years and the need for a comprehensive analysis of the trade-offs they propose. We formalized the problem of temporal coherence in terms of goals and compared existing methods accordingly (Figure 9). The goal of this report is to help uninformed readers to choose the method that best suits their needs, as well as motivate further research to address the limitations of existing methods.

Figure 9. The temporal coherence problem involves three goals represented by the axes of these diagrams. Fully ignoring one of them produces the opposite artifacts.

This work is a collaboration with Pierre Bénard and Joëlle Thollot (ARTIS / INRIA Rhône-Alpes) and was published in the journal Computer Graphics Forum [14].

5.1.11. Improving Gabor Noise

Participant: Ares Lagae.

We have recently proposed a new procedural noise function, Gabor noise, which offers a combination of properties not found in existing noise functions. In this project, we present three significant improvements to Gabor noise: (1) an isotropic kernel for Gabor noise, which speeds up isotropic Gabor noise with a factor of roughly two, (2) an error analysis of Gabor noise, which relates the kernel truncation radius to the relative error of the noise, and (3) spatially varying Gabor noise, which enables spatial variation of all noise parameters. These improvements make Gabor noise an even more attractive alternative for existing noise functions. Fig. 3 shows a procedural textures generated with our new improved Gabor noise. This work was published in IEEE Transactions on Computer Graphics and Visualization [19].

5.2. Visual Perception and Audio Rendering
5.2.1. Perception of Visual Artifacts in Image-Based Rendering of Façades

Participants: Peter Vangorp, Gaurav Chaurasia, Pierre-Yves Laffont, George Drettakis.

Image-based rendering (IBR) techniques allow users to create interactive 3D visualizations of scenes by taking a few snapshots (Figure 11 (left)). However, despite substantial progress in the field, the main barrier to better quality and more efficient IBR visualizations are several types of common, visually objectionable artifacts. These occur when scene geometry is approximate or viewpoints differ from the original shots, leading to parallax distortions (Figure 11 (mid)), blurring, ghosting (Figure 11 (right)) and popping errors that detract from the appearance of the scene. We argue that a better understanding of the causes and perceptual impact of these artifacts is the key to improving IBR methods.

We present a series of psychophysical experiments in which we systematically map out the perception of artifacts in IBR visualizations of façades as a function of the most common causes. We separate artifacts into different classes and measure how they impact visual appearance as a function of the number of images available, the geometry of the scene and the viewpoint. The results reveal a number of counter-intuitive effects in the perception of artifacts.

We summarize our results in terms of the following practical guidelines for improving existing and future IBR techniques:
• When the total number of available images is small, e.g., because of storage limitations, it is preferable to use a sudden transition with its associated popping artifact rather than a gradual blending transition with its associated ghosting artifact.

• Interestingly, the depth range of the façade does not affect the perceived parallax distortions, even though it clearly does affect the objective parallax distortions. Only the intended output viewing angle should be taken into account when capturing images.

• For Google Street View™-like visualizations, a shorter cross-fading transition would improve the perceived quality.

This work is a collaboration with Roland W. Fleming (Justus-Liebig-University Giessen, Germany). The work was published in the special issue of the journal Computer Graphics Forum [20] and presented at the Eurographics Symposium on Rendering 2011.

### 5.2.2. Perception of Slanted, Textured Façades

**Participants:** Peter Vangorp, Adrien Bousseau, Gaurav Chaurasia, George Drettakis.

In large-scale urban visualizations, buildings are often geometrically represented by simple boxes textured with images of the façades. Any depth variations in the façade, such as balconies, are perceived to have distorted angles when the viewer is not at the capture camera position. The retinal hypothesis provides the most likely prediction of the magnitude of the perceived distortion. We conduct psychophysical experiments to measure the perceived distortion, thereby validating the retinal hypothesis, and to measure the threshold for detecting any distortion. The result is a prediction of the valid range of viewer motion for a given capture.

This work is a collaboration with Martin S. Banks (UC Berkeley).

### 5.2.3. Binocular and Dynamic Cues to Glossiness

**Participants:** Peter Vangorp, George Drettakis.

Recent advances in display technology have made it possible to present high quality stereoscopic imagery with accurate head tracking. This improves not only depth perception but also affects the perception of glossy materials. Previous work has shown that these conditions can increase the perceived gloss by a small amount. We conduct psychophysical experiments to measure this effect quantitatively.

This work is a collaboration with Roland W. Fleming (Justus-Liebig-University Giessen, Germany) and Martin S. Banks (UC Berkeley).

### 5.2.4. Sound Particles

**Participants:** Charles Verron, George Drettakis.

This research deals with a sound synthesizer dedicated to particle-based environmental effects, and intended to be used in interactive virtual environments. The synthesis engine is based on five physically-inspired basic elements (sound atoms) that can be parameterized and stochastically distributed in time and space. Physically-inspired controls simultaneously drive graphics particle models (e.g., distribution of particles, average particles velocity etc.) and sound parameters (e.g., distribution of sound atoms, spectral modifications etc.). The simultaneous audio/graphics controls result in an intricate interaction between the two modalities that enhances the naturalness of the scene. The approach is currently illustrated with three environmental phenomena: fire, wind, and rain.

### 5.2.5. Sound Synthesis for Crowds

**Participants:** Charles Verron, George Drettakis.

We are currently investigating new methods for synthesis of crowd sounds in virtual environments. Crowd sounds are constituted of many overlapping voices spatialized at different positions in the environment. A novel level of detail for crowd sounds is desirable: the cost of spatializing many individual voice sounds can be replaced by an efficient babble noise synthesis model. Furthermore, high-level control should be proposed to modify the crowd sounds by semantic parameters, related to the crowd emotional state (e.g., calm, angry...). This research should result in a new real-time crowd sound synthesizer with semantic controls for virtual environments.
5.3. Interaction and Design for Audiovisual Virtual Environments

5.3.1. Lighting Design for Material Depiction

Participants: Adrien Bousseau, Emmanuelle Chapoulie.

Shading, reflections and refractions are important visual features for understanding the shapes and materials in an image. While well designed lighting configurations can enhance these features and facilitate image perception (Figure 12 b), poor lighting design can lead to misinterpretation of image content (Figure 12 a).

![Our optimized lighting emphasizes materials](image1)

![Poor lighting de-emphasizes materials](image2)

*Figure 12. Our method (a) automatically optimizes the lighting to enhance material-specific visual features. The lighting reveals the thin and thick parts of the subsurface scattering wax candle, it accentuates the Fresnel reflections along the side of the porcelain vase and it adds strong specular highlights to emphasize the shiny chrome metal of the sculpture. Poorly designed lighting (b) diminishes these characteristic visual features of the materials. The candle appears more like solid plastic, the vase appears to be made of diffuse clay and the sculpture no longer looks like it is made of chrome.*

We have presented an automated system for optimizing and synthesizing environment maps that enhance the appearance of materials in a scene. We first identified a set of lighting design principles for material depiction. Each principle specifies the distinctive visual features of a material and describes how environment maps can emphasize those features. We then proposed a general optimization framework to solve for the environment map that best fulfill the design principles. Finally we described two techniques for transforming existing photographic environment maps to better emphasize materials. Our approach generates environment maps that enhance the depiction of a variety of materials including glass, metal, plastic, marble and velvet.

This work is a collaboration with Ravi Ramamoorthi and Maneesh Agrawala (UC Berkeley) as part of our CRISP associate team with UC Berkeley. The work was published in the special issue of the journal Computer Graphics Forum, presented at the Eurographics Symposium on Rendering 2011.

5.3.2. A Multimode Immersive Conceptual Design System for Architectural Modeling and Lighting

Participants: Marcio Cabral, Peter Vangorp, Gaurav Chaurasia, Emmanuelle Chapoulie, Martin Hachet, George Drettakis.

We developed a system which allows simple architectural design in immersive environments. The user is able to define the initial conceptual design of the model and can take into account the effects of daylight. Our system allows the manipulation of simple elements such as windows, doors and rooms while the overall model is automatically adjusted to the manipulation. The system runs on a four-sided stereoscopic, head-tracked immersive display. We also provide simple lighting design capabilities, with an abstract representation of sunlight and its effects when shining through a window. Our system provides three different modes
of interaction: a miniature-model table mode, a fullscale immersive mode and a combination of table and immersive which we call mixed mode (see Figure 13). Our goal is to study direct manipulation for basic 3D modeling in an immersive setting, in the context of conceptual or initial design for architecture.

We performed an initial pilot user test to evaluate the relative merits of each mode for a set of basic tasks such as resizing and moving windows or walls, and a basic light-matching task. The study indicates that users appreciated the immersive nature of the system, and found interaction to be natural and pleasant. In addition, the results indicate that the mean performance times seem quite similar in the different modes, opening up the possibility for their combined usage for effective immersive modeling systems for novice users.

Figure 13. (a) Table mode, (b) Immersive mode, (c) Mixed mode

5.3.3. Walking in a Cube: Novel Metaphors for Safely Navigating Large Virtual Environments in Restricted Real Workspaces

Participants: Peter Vangorp, Emmanuelle Chapoulie, George Drettakis.

Immersive spaces such as 4-sided displays with stereo viewing and high-quality tracking provide a very engaging and realistic virtual experience. However, walking is inherently limited by the restricted physical space, both due to the screens (limited translation) and the missing back screen (limited rotation). We propose three novel navigation techniques that have three concurrent goals: keep the user safe from reaching the translational and rotational boundaries; increase the amount of real walking; and finally, provide a more enjoyable and ecological interaction paradigm compared to traditional controller-based approaches. We notably introduce the “Virtual Companion”, which uses a small bird to guide the user through VEs larger than the physical space. We evaluate the three new techniques through a user study with pointing and path following tasks. The study provides insight into the relative strengths of each new technique for the three aforementioned goals.

This work is a collaboration with Gabriel Cirio, Maud Marchal, and Anatole Lécuyer (VR4I / INRIA Rennes) in the context of ARC NIEVE (Section 6.2.1) and has been accepted for publication [17].

5.3.4. Inferring Normals Over Design Sketches

Participant: Adrien Bousseau.

We are currently working on a sketch-based tool to infer normals over a 2D drawing. Our tool should allow users to apply realistic and non-photorealistic shading over the drawing, with applications in product design. This work is a collaboration with Alla Sheffer (University of British Columbia), Cloud Shao and Karan Singh (University of Toronto).
5.3.5. Using natural gestures into virtual reality immersive space

Participants: Emmanuelle Chapoulie, George Drettakis.

We are studying the use of gestures which are as natural as possible in a context of virtual reality environments. We define a scenario which is a sequence of tasks (hiding, finding, pushing, pulling, grabbing, picking up, putting down objects) that the users will perform with hands and with wands, in order to evaluate the usability of our interaction approach. Each task will be used to evaluate a specific criterion.
6. New Results

6.1. Physical modelling and simulation

6.1.1. Modal analysis for haptic manipulation of deformable models

Participants: Zhaoguang Wang, Georges Dumont [contact].

Real-time interaction between designer and deformable mock-up in VR (Virtual Reality) environment is a natural and promising manner to evaluate designing feasibility. Using finite element method (FEM) for solving this issue leads to high fidelity simulation but to simulation rates that do not meet the requirements (1000Hz) of real time haptic applications. We have proposed a two-stage method based on linear modal analysis. In this method, different modal subspaces, related to use scenarios, are pre-computed offline. These data are then combined online with respect to a simulation division scheme to obtain real time deformations of the parts with respect to the modal response. Two main features are developed in the method. First, we apply an adapted meshing method during the pre-computation process. This method allows to automatically switch between different modal subspaces depending on the interaction region. Second, we divide the real time deformation computation into two separate modules by extracting sub-matrixes from the pre-computed modal matrixes. This separates the haptic simulation loop from the whole deformation computation an thus preserves the haptic response. This work was presented in WINVR 2011 conference [31] is accepted for publication [8] and was the subject of the PhD Thesis of Zhaoguang Wang, that was defended in june 2011 [3].

6.1.2. Real-time mechanical simulation of brittle fracture

Participants: Loeiz Glondu, Georges Dumont [contact], Maud Marchal [contact].

Simulating brittle fracture of stiff bodies is now commonplace in computer graphics. However, simulating the deformations undergone by the bodies in a realistic way remains computationally expensive. Thus, physically-based simulation of brittle fracture in real-time is still challenging for interactive applications. We are currently working on a new physically-based approach for simulating realistic brittle fracture in real-time. Our method is composed of two main original parts: (1) a fracture initiation model based on modal analysis and a new contact force model and (2) a fracture propagation model based on a novel physically-based algorithm (Figure 2). First results of this method have been published in [32].

Adding physical properties to objects within a virtual world can not generally be handled in real-time during a simulation. For that reason, it is still difficult nowadays to physically simulate fragments of fractured objects or parts of teared/cut objects. We have proposed a method for handling the real-time physical simulations of arbitrary objects that are represented by their surface mesh. Our method is based on a pre-computed shape database in which physical data are stored for a wide variety of objects. When a query object needs to be physically simulated in the virtual world, a similarity search is performed inside the database and the associated physical data are extracted. Our approach proposes to compare three different similarity search methods that fit with our real-time needs. Our results show that our approach has a great potential for the physical simulation of arbitrary objects in interactive applications. These results have been published in the Eurographics International Workshop on Virtual Reality Interaction and Physical Simulation (Vriphys) [21].

6.1.3. Collision detection in large scale environments with High Performance Computing

Participants: Quentin Avril, Valérie Gouranton [contact], Bruno Arnaldi.
Virtual reality environments are becoming increasingly large and complex and real-time interaction level is becoming difficult to stably insure. Indeed, because of their complexity, detailed geometry and specific physical properties, these large scale environments create a critical computational bottleneck on physical algorithms. Our work focused on the first step of the physical process: the collision detection. These algorithms can sometimes have a quadratic complexity. Solving and simplifying the collision detection problem is integral to alleviating this bottleneck. Hardware architectures have undergone extensive changes in the last few years that have opened new ways to relieve this computational bottleneck. Multiple processor cores offer the ability to execute algorithms in parallel on one single processor. At the same time, graphics cards have gone from being a simple graphical display device to a supercomputer. These supercomputers now enjoy attention from a specialized community dealing solely with physical simulation. To perform large scale simulations and remain generic on the runtime architecture, we proposed unified and adaptive mapping models between collision detection algorithms and the runtime architecture using multi-core and multi-GPU architectures. We have developed innovative and effective solutions to significantly reduce the computation time in large scale environments while ensuring the stability and reproducibility of results (cf. Figure 3). We proposed a new pipeline of collision detection with a granularity of parallelism on multicore processors or multi-GPU platforms [11]. It enables simultaneous execution of different stages of the pipeline and a parallel internal to each of these steps. This was the subject of the PhD thesis of Quentin Avril [1].

6.1.4. Assessment of inverse dynamics method for muscle activity analysis

Participants: Georges Dumont [contact], Charles Pontonnier.

The use of virtual reality tools for ergonomics applications is a very important challenge. In order to improve the design of workstations, an estimation of the muscle forces involved in the work tasks has to be done. Several methods can lead to these muscle forces. In this study, we try to assess the level of confidence for results obtained with an inverse dynamics method from real captured work tasks. The chosen tasks are meat cutting tasks, well known to be highly correlated to musculoskeletal troubles appearance in the slaughter industry.

The experimental protocol consists in recording three main data during meat cutting tasks, and analyse their variation when some of the workstation design parameters are changing [26].

1. External (cutting)force data: for this purpose, a 3D instrumented knife has been designed in order
Figure 3. Simulation of moving objects with varying size. Our approach enables to perform the Broad phase step in interactive time using optimized spatial brute force algorithm. (Left: 2,000 Objects - Right: 50,000 Objects.

to record the force applied by the subject during the task;
2. Motion Capture data: for this purpose, we use a motion capture system with active markers (Visualeyez II, Phoenix Technologies, Canada);
3. EMG data: several muscle activities are recorded using electromyographic electrodes, in order to compare these activities to the ones obtained from the inverse dynamics method.

Then the motion is replayed in the AnyBody modeling system (AnyBody, Aalborg, Denmark) in order to obtain muscle forces generated during the motion. A trend comparison is then done [27], comparing recorded and computed muscle activations. Results show that most of the computed activations are qualitatively close from the recorded ones (similar shapes and peaks), but quantitative comparison leads to major differences between recorded and computed activations (the trend followed by the recorded activations in regard of a workstation design parameter, such as the table height, is not obtained with the computed activations). We currently explore those results to see if the fact that co-contraction of single joints muscles is badly estimated by classical inverse dynamics method can be a reason of this issue. We also work on the co-contraction simulation in order to improve the results [28].

This work has been done in collaboration with the Center for Sensory-motor Interaction (SMI, Aalborg University, Aalborg, Denmark), particularly Mark de Zee (Associate Professor) and Pascal Madeleine (Professor). Charles Pontonnier spent a 9 months post-doctoral fellowship at SMI from December 2010 to August 2011.

6.2. Multimodal immersive interaction

6.2.1. Brain-Computer Interaction based mental state

Participants: Anatole Lécuyer [contact], BrunoArnaldi, LaurentGeorge, YannRenard.

In [20], presented at IEEE EMBS conference, we have explored the use of electrical biosignals measured on scalp and corresponding to mental relaxation and concentration tasks in order to control an object in a video game as illustrated in Figure 4. To evaluate the requirements of such a system in terms of sensors and signal processing we compared two designs. The first one used only one scalp electroencephalographic (EEG) electrode and the power in the alpha frequency band. The second one used sixteen scalp EEG electrodes and machine learning methods. The role of muscular activity was also evaluated using five electrodes positioned on the face and the neck.
Results show that the first design enabled 70% of the participants to successfully control the game, whereas 100% of the participants managed to do it with the second design based on machine learning. Subjective questionnaires confirm these results: users globally felt to have control in both designs, with an increased feeling of control in the second one. Offline analysis of face and neck muscle activity shows that this activity could also be used to distinguish between relaxation and concentration tasks. Results suggest that the combination of muscular and brain activity could improve performance of this kind of system. They also suggest that muscular activity has probably been recorded by EEG electrodes.

In [19], presented in the 5th International Brain-Computer Interface Conference, we introduce the concept of Brain-Computer Interface (BCI) inhibitor, which is meant to standby the BCI until the user is ready, in order to improve the overall performance and usability of the system. BCI inhibitor can be defined as a system that monitors user’s state and inhibits BCI interaction until specific requirements (e.g. brain activity pattern, user attention level) are met.

We conducted a pilot study to evaluate a hybrid BCI composed of a classic synchronous BCI system based on motor imagery and a BCI inhibitor (Figure 5). The BCI inhibitor initiates the control period of the BCI when requirements in terms of brain activity are reached (i.e. stability in the beta band).

Preliminary results with four participants suggest that BCI inhibitor system can improve BCI performance.

### 6.2.2. Navigating in virtual worlds using a Brain-Computer Interface

**Participants:** Anatole Lécuyer [contact], Jozef Legény.

When a person looks at a light flickering at a constant frequency, we can observe a corresponding electrical signal in their EEG. This phenomenon, located in the occipital area of the brain is called Steady-State Visual-Evoked Potential (SSVEP).
In [7] we introduce a novel paradigm for a controller using SSVEP. Compared to the state-of-the-art implementations which use static flickering targets, we have used animated and moving objects. In our example applications we have used animated butterflies flying in front of the user as show in Figure 6. A study has revealed that, at the cost of decreased performance, this controller increases the personal feeling of presence.

These results show that integrating visual SSVEP stimulation into the environment is possible and that further study is necessary in order to improve the performance of the system.

6.2.3. Walking-in-place in virtual environments

**Participants:** Anatole Lécuyer [contact], Maud Marchal [contact], Léo Terziman, Bruno Arnaldi, Franck Multon.

The Walking-In-Place interaction technique was introduced to navigate infinitely in 3D virtual worlds by walking in place in the real world. The technique has been initially developed for users standing in immersive setups and was built upon sophisticated visual displays and tracking equipments. We have proposed to revisit the whole pipeline of the Walking-In-Place technique to match a larger set of configurations and apply it notably to the context of desktop Virtual Reality. With our novel “Shake-Your-Head” technique, the user has the possibility to sit down, and to use small screens and standard input devices for tracking. The locomotion simulation can compute various motions such as turning, jumping and crawling, using as sole input the head movements of the user (Figure 7).

In a second study [29] we analyzed and compared the trajectories made in a Virtual Environment with two different navigation techniques. The first is a standard joystick technique and the second is the Walking-In-Place (WIP) technique. We proposed a spatial and temporal analysis of the trajectories produced with both techniques during a virtual slalom task. We found that trajectories and users’ behaviors are very different across the two conditions. Our results notably showed that with the WIP technique the users turned more often and navigated more sequentially, i.e. waited to cross obstacles before changing their direction. However, the users were also able to modulate their speed more precisely with the WIP. These results could be used to optimize the design and future implementations of WIP techniques. Our analysis could also become the basis of a future framework to compare other navigation techniques.

6.2.4. Improved interactive stereoscopic rendering : SCVC

**Participants:** Jérôme Ardouin, Anatole Lécuyer [contact], Maud Marchal [contact], Eric Marchand.
Frame cancellation comes from the conflict between two depth cues: stereo disparity and occlusion with the screen border. When this conflict occurs, the user suffers from poor depth perception of the scene. It also leads to uncomfortable viewing and eyestrain due to problems in fusing left and right images.

In [10], presented at the IEEE 3DUI conference, we propose a novel method to avoid frame cancellation in real-time stereoscopic rendering. To solve the disparity/frame occlusion conflict, we propose rendering only the part of the viewing volume that is free of conflict by using clipping methods available in standard real-time 3D APIs. This volume is called the Stereo Compatible Volume (SCV) and the method is named Stereo Compatible Volume Clipping (SCVC).

Black Bands, a proven method initially designed for stereoscopic movies is also implemented to conduct an evaluation. Twenty two people were asked to answer open questions and to score criteria for SCVC, Black Bands and a Control method with no specific treatment. Results show that subjective preference and user’s depth perception near screen edge seem improved by SCVC, and that Black Bands did not achieve the performance we expected.

At a time when stereoscopic capable hardware is available from the mass consumer market, the disparity/frame occlusion conflict in stereoscopic rendering will become more noticeable. SCVC could be a solution to recommend. SCVC’s simplicity of implementation makes the method able to target a wide range of rendering software from VR application to game engine.

6.2.5. Six degrees-of-freedom haptic interaction

Participants: Anatole Lécuyer [contact], Maud Marchal [contact], Gabriel Cirio.

Haptic interaction with virtual objects is a major concern in the virtual reality field. There are many physically-based efficient models that enable the simulation of a specific type of media, e.g. fluid volumes, deformable and rigid bodies. However, combining these often heterogeneous algorithms in the same virtual world in order to simulate and interact with different types of media can be a complex task.

In [5], published at IEEE Transactions on visualization and Computer Graphics, we propose a novel approach that allows real-time 6 Degrees of Freedom (DoF) haptic interaction with fluids of variable viscosity. Our haptic rendering technique, based on a Smoothed-Particle Hydrodynamics (SPH) physical model, provides a realistic haptic feedback through physically-based forces. 6DoF haptic interaction with fluids is made possible thanks to a new coupling scheme and a unified particle model, allowing the use of arbitrary-shaped rigid bodies. Particularly, fluid containers can be created to hold fluid and hence transmit to the user force feedback coming from fluid stirring, pouring, shaking or scooping. We evaluate and illustrate the main features of our approach through different scenarios, highlighting the 6DoF haptic feedback and the use of containers.
The Virtual Crepe Factory [14] illustrates this approach for 6DoF haptic interaction with fluids. It showcases a 2-handed interactive haptic scenario: a recipe consisting in using different types of fluid in order to make a special pancake also known as "crepe". The scenario (Figure 8) guides the user through all the steps required to prepare a crepe: from the stirring and pouring of the dough to the spreading of different toppings, without forgetting the challenging flipping of the crepe. With the Virtual Crepe Factory, users can experience for the first time 6DoF haptic interactions with fluids of varying viscosity.

Figure 8. A complete use-case of our approach: a virtual crepe preparation simulator. The user manipulates a bowl (left hand, left haptic device) and a pan (right hand, right haptic device).

In [15], presented at the IEEE Virtual Reality Conference, we propose the first haptic rendering technique for the simulation and the interaction with multistate (Figure 9) media, namely fluids, deformable bodies and rigid bodies, in real-time and with 6DoF haptic feedback. The shared physical model (SPH) for all three types of media avoids the complexity of dealing with different algorithms and their coupling. We achieve high update rates while simulating a physically-based virtual world governed by fluid and elasticity theories, and show how to render interaction forces and torques through a 6DoF haptic device.

Figure 9. 6DoF haptic interaction in a medical scenario. Fluid blood pours from the deformable intestine when the user penetrates it with the rigid probe.

6.2.6. Joyman: a human-scale joystick for navigating in virtual worlds

Participants: Maud Marchal [contact], Anatole Lécuyer, Julien Pettré.
We have proposed a novel interface called Joyman (Figure 10), designed for immersive locomotion in virtual environments. Whereas many previous interfaces preserve or stimulate the users proprioception, the Joyman aims at preserving equilibrioception in order to improve the feeling of immersion during virtual locomotion tasks. The proposed interface is based on the metaphor of a human-scale joystick. The device has a simple mechanical design that allows a user to indicate his virtual navigation intentions by leaning accordingly. We have also proposed a control law inspired by the biomechanics of the human locomotion to transform the measured leaning angle into a walking direction and speed - i.e., a virtual velocity vector. A preliminary evaluation was conducted in order to evaluate the advantages and drawbacks of the proposed interface and to better draw the future expectations of such a device.

This principle of this new interface was published at international conference IEEE 3DUI [25] and a patent has been filed for the interface. A demonstration of this interface was proposed at ACM Siggraph Asia Emerging Technologies [33].

![Figure 10. Prototype of the "Joyman"](image)

### 6.2.7. Interactions within 3D virtual universes

**Participants:** Thierry Duval [contact], Valérie Gouranton [contact], Bruno Arnaldi, Laurent Aguerreche, Cédric Fleury, Thi Thuong Huyen Nguyen.

Our work focuses upon new formalisms for 3D interactions in virtual environments, to define what an interactive object is, what an interaction tool is, and how these two kinds of objects can communicate together. We also propose virtual reality patterns to combine navigation with interaction in immersive virtual environments.

We have worked upon generic interaction tools for collaboration, based on multi-point interaction. In that context we have studied the efficiency of one instance of our Reconfigurable Tangible Device, the RTD-3, for collaborative manipulation of 3D objects compared to state of the art metaphors [9]. We have setup an experiment for collaborative distant co-manipulation (figure 1) of a clipping plane inside for remotely analyzing 3D scientific data issued from an earthquake simulation.

### 6.3. Collaborative work in CVE’s

#### 6.3.1. The immersive interactive virtual cabin (IIVC)

**Participants:** Thierry Duval [contact], Valérie Gouranton [contact], Alain Chauffaut, Bruno Arnaldi, Cédric Fleury.

We are still improving the architecture of our Immersive Interactive Virtual Cabin to improve the user’s immersion with all his real tools and so to make the design and the use of 3D interaction techniques easier, and to make possible to use them in various contexts, either for different kinds of applications, or with different kinds of physical input devices.
The IIVC is now fully implemented in our two VR platforms: OpenMASK 5.1 and Collaviz 7.1.3.

### 6.3.2. Generic architecture for 3D interoperability

**Participants:** Thierry Duval [contact], Valérie Gouranton, Cédric Fleury, Rozenn Bouville Berthelot, Bruno Arnaldi.

Our goal is to allow software developers to build 3D interactive and collaborative environments without bothering with the 3D graphics API they are using. This work is the achievement of the IIVC software architecture. We have proposed PAC-C3D (Figure 11), a new software architectural model for collaborative 3D applications, in order to provide a higher abstraction for designing 3D virtual objects, and in order to provide interoperability, making it possible to share a virtual universe between heterogeneous 3D viewers [17], [16].

![Figure 11. The PAC-C3D software architectural model makes interoperability possible between heterogeneous 3D viewers](image)

We also study how to offer interoperability between virtual objects that are loaded in the same virtual environment but that are described using different formats. This is why we have proposed a generic architecture for enabling interoperability between 3D formats (Figure 12), the Scene Graph Adapter [12]. Our SGA is now able to allow events coming from a 3D format to act upon data provided in another format, such as X3D events operating on Collada data [4].

### 6.4. Immersia Virtual Reality room

**Participants:** Georges Dumont [contact], Alain Chauffaut [contact], Ronan Gaugne [contact], Rémi Félix, Marwan Badawi, Bruno Arnaldi, Thierry Duval, Valérie Gouranton.

The team was the first in France to host a large-scale immersive virtual reality equipment known as Immersia. This platform, with full visual and sound immersion, is dedicated to real-time, multimodal (vision, sound, haptic, BCI) and immersive interaction. The Immersia platform is a key node of the European transnational VISIONAIR infrastructure and will be open in 2012 to the access of foreign research projects. It will accommodate experiments using interactive and collaborative virtual-reality applications that have multiple local or remote users. Our new wall has four faces: a front, two sides and a ground. Dimensions are 9.6 m wide, 2.9 m deep and 3.1 m high. The visual reproduction system combines eight Barco Galaxy NW12 projectors and three Barco Galaxy 7+ projectors. Visual images from Barco projectors are rendered on glass screens. They are adjusted for the user’s position, and this, together with their high resolution and homogeneous coloring, make them very realistic. The ART localization system, constituted of 16 ART-track2 cameras, enables real
Figure 12. Our architecture allows the loading of any 3D graphics format simultaneously in any available rendering engine. The scene graph adapter is an interface that adapts a scene graph (SG) of a given format into a renderer scene graph and which also allows the rendering part to request this scene graph.

objects to be located within the U-shape. Sound rendering is provided by a Yamaha processor, linked either to Genelec speakers with 10.2 format sound or Beyer Dynamic headsets with 5.1 virtual format sound, controlled by the user’s position.
5. New Results

5.1. Introduction

This year we obtained new results in our three sub-objectives and also related to Focuslab platform and software valorization:

1. **Sub-objective 1 - Mining for Knowledge Discovery in Information Systems**: this we get five results (with one achieved PhD thesis).
   
   Let us note that six past works on this sub-objective described in previous AxIS annual reports have been published this year as articles in international journals ([22], [11]) or conferences, one in a national journal [50], two in a french-speaking conference [35], [46], one book [20] and one book chapter [52] at international level. Indeed
   
   - The work in 2009 on mining data streams by Marasci in her thesis [96] has been published in [11] with more details in the algorithms and in the experiments.
   - The work published in 2008 on discovering frequent behaviors [107] has been published in [22] with more details in the algorithms and in the experiments.
   - Our previous work on satellite image mining in 2010 [99] has been published in French at EGC [35].
   - Our past work on Functional data analysis involving data described by regular functions rather than by a finite number of real valued variables has been published as a scientific book chapter [52]. In this paper we propose to use a clustering approach that targets variables rather than individual to design a piecewise constant representation of a set of functions. The contiguity constraint induced by the functional nature of the variables allows a polynomial complexity algorithm to give the optimal solution.
   - In the context of the WRUM project (Morocco) and Zemmouri’s PhD thesis, we have a long paper accepted at JFO 2011 related to past works (2010) on how to integrate domain knowledge in a multi-view KDD process [46].

2. **Sub-objective 2 - Information and Social Networks Mining for Supporting Information Retrieval**: Three results (with one achieved PhD thesis). Let us remind the best paper [15] obtained By E. Smirnova at ECIR 2011 for her research on expert finding.

3. **Sub-objective 3 - Multidisciplinary Research For Supporting User Oriented Innovation**: this interdisciplinary research is dedicated to the design, tailoring and refinement of methodologies and tools for a better users’ involvement in innovation processes. We have seven results this year.

Concerning our activity in terms of **FocusLab Experimental Platform and Software**, a) we first applied ATWUEDA on another context of evolutive data (on system monitoring data at EDF) which is different of Web usage data) to show the genericity of the approach [9], and b) we develop a Web-based version of the FocusLab experimental platform for analysis usage data (hardware and software parts).

5.2. Mining for Knowledge Discovery in Information Systems

5.2.1. **Mining Data Streams: Clustering and Pattern extraction**

**Participant**: Chongsheng Zhang.
In Zhang’s thesis [19] (supervised by F. Masseglia), which was partially funded by ANR MIDAS (cf. 6.2.1), we present our study of the management and mining issues on data streams with evolving tuples, caused by model updates or tuple revisions. For instance, in an online auction system where bids on auction items are streaming, it is possible that some users may bid for more than one item within the user-specified time interval. As a result, the profiles of the users can be updated or revised in such applications. Data streams having evolving tuples bring new challenges as well as research opportunity. In this work, he develops novel and efficient models and methods for managing and mining data streams with evolving tuples. (I) To model data streams with evolving tuples, we propose the Anti-Bouncing Streaming model (ABS) for usage streams. ABS fits data streams with evolving tuples and it enables methods for processing of data streams to handle tuple updates or revisions. (II) To find frequent itemsets from data streams with evolving tuples over pane-based sliding windows, we conduct theoretical analysis and propose theorems which can avoid scanning the past slides to check for possible itemsets that may become frequent. We also design novel data structures which can manage the data streams with evolving tuples efficiently and facilitate the mining of frequent itemsets. Moreover, we devise an efficient counting algorithm to verify the frequentness of the candidate frequent itemsets. We also propose two running frameworks for this problem. (III) To extract important feature set from data streams (including the ones with evolving tuples), based upon ABS, we devise the streaming feature set selection algorithm for data streams which is the first in the literature. This method is based on information theory to extract the informative feature sets. To further accelerate the extraction of the most informative feature set from high-dimensional data, we propose a framework that reduces the huge search space to a rather small subset while still guarantee the quality of the discovered feature sets.

In 2011, Chongsheng Zhang has mainly worked on a data stream mining method, intending to extract frequent itemsets. This method has not been published yet and is described in Chapter 5 (page 79) of his thesis document [19].

5.2.2. Clustering on Multiple Dissimilarity Matrices

Participants: Yves Lechevallier, Francisco de A.T. de Carvalho, Thierry Despeyroux, Alessandra Silva Anyzewski.

In [23] we introduce hard clustering algorithms that are able to partitioning objects taking into account simultaneously their relational descriptions given by multiple dissimilarity matrices [49]. The aim is to obtain a collaborative role of the different dissimilarity matrices in order to obtain a final consensus partition. These matrices could have been generated using different sets of variables and a fixed dissimilarity function or using a fixed set of variables and different dissimilarity functions, or using different sets of variables and dissimilarity functions.

These methods, which are based on the dynamic hard clustering algorithm for relational data as well as on the dynamic clustering algorithm based on adaptive distances, are designed to furnish a partition and a prototype for each cluster as well as to learn a relevance weight for each dissimilarity matrix by optimizing an adequacy criterion that measures the fitting between clusters and their representatives.

These relevance weights change at each algorithm iteration and can either be the same for all clusters or different from one cluster to another. The usefulness of these partitioning hard clustering algorithms are shown on two time trajectory real world datasets.

5.2.3. Clustering of Constrained Symbolic Data

Participants: Marc Csernel, Francisco de A.T. de Carvalho.

In the context of our FACEPE collaboration with Brazil (cf. section 6.4.3.1), we have presented a method which allows clustering of symbolic descriptions constrained by presence rules in a polynomial time instead of a combinatorial one. This method allows to deal with “false missing values”. Such a method can be applied on various classification problems [26].

5.2.4. Web Page Clustering based on a Community Detection Algorithm

Participants: Yves Lechevallier, Yacine Slimani.
Extracting knowledge from Web user’s access data in Web Usage Mining (WUM) process is a challenging task that is continuing to gain importance as the size of the web and its user-base increase. That is why meaningful methods have been proposed in the literature in order to understand the behaviour of the user in the web and improve the access modes to information. In this work [42], we are interested in the analysis of the user browsing behavior. The objective is to understand the navigational practices of users (teachers, students and administrative staff). First we clean the data by removing irrelevant information and noise. During the second step, remaining data are arranged in a coherent way in order to identify user sessions. After we defined a new approach [42] of knowledge extraction. This approach treats the data resulting from the preprocessing phase (first and second steps) as being a set of communities. Our approach extends the Modularity measure, proposed by Newman and Girvan [97], in the Web Mining context in order to benefit from their classifying capacity in the communities discovery.

This work is done in collaboration with the LRIA laboratory – Université Ferhat Abbas, Sétif, Algérie

5.2.5. Critical Edition of Sanskrit Texts

Participants: Marc Csernel, Nicolas Béchet, Ehab Hassan, Yves Lechevallier.

New progresses concerning the computer assisted elaboration of Sanskrit texts have been made. First Nicolas Béchet and Marc Csernel have worked on the problem of moved texts. After an alignment between two versions of the texts, we discover that some parts of the text appears to have been moved according to the technics developed in [48]. Until now, we were not able to discover when a text has been moved in a manuscript.

Now using a words-grams technique proposed in [48], we were able to obtain quite good results on the moved texts problem and we were able to optimize the different possible parameters. A paper on the subject has been submitted to the Cicling 2012 conference (http://www.cicling.org/2012/).

After the new treatment related to the moved text problem, we need to provide an interactive display of the critical edition. During his internship, Ehab Hassan has been working on the subject and obtained good results. These results need to be deeply examined by Sanskritists to see if they always fulfill their needs.

5.3. Information and Social Networks Mining for Supporting Information Retrieval

5.3.1. Clustering of Relational Data and Social Network Data

Participants: Yves Lechevallier, Amine Louati.

The automatic detection of communities in a social network can provide this kind of graph aggregation. The objective of graph aggregations is to produce small and understandable summaries and can highlight communities in the network, which greatly facilitates the interpretation.

Social networks allow having a global view of the different actors and different interactions between them, thus facilitating the analysis and information retrieval.

In the enterprise context, a considerable amount of information is stored in relational databases. Therefore, relational database can be a rich source to extract social network. The extracted network has in general a huge size which makes its analyses and visualization difficult tasks. In [45], we propose a social network extraction approach from relational database.

Often, the network has a large size which makes its analysis and visualization difficult.

The aggregation step is a necessary task, so we offer [33] and [32] an aggregation step based on the k-SNAP algorithm [109] that produces a summary graph by grouping nodes based on attributes and relationships selected by the user.

This work is done in collaboration with Marie-Aude Aufaure, head of the Business Intelligence Team, Ecole Centrale Paris, MAS Laboratory.
5.3.2. Networks Solutions for Expert Finding and People Name Disambiguation
Participants: Elena Smirnova, Yi-Ling Kuo, Brigitte Trousse.

The task of finding people who are experts on a given topic has cently attracted close attention. State-of-the-art expert finding algorithms uncover knowledge areas of candidate experts based on textual content of associated documents. While powerful, these models ignore social structure that might be available. Therefore, we develop a Bayesian hierarchical model for expert finding that accounts for both content and social relationships. The model assumes that social links are determined by expertise similarity between candidates. The results of EGC experiments on UvT expert collection have demonstrated the effectiveness of our algorithm [43].

E. Smirnova visited Intellius, people search technology company (Aug 8 - Oct 5, 2011): the goal of this visit was to validate the research on expert finding in social networks on real dataset and further advance it. As a real dataset, we have taken a sample of United States LinkedIn public profiles. We built an organizational network by connecting a LinkedIn user and his colleagues at different workplaces. We also constructed a geographical network from user’s current location in the United States. We used Amazon’s Mechanical Turk framework (http://aws.amazon.com/code/923) to collect user-oriented judgements for model evaluation. We found that the user-oriented model is statistically significantly preferred to the baseline model on 72.5% of queries.

Her work on name disambiguation done in 2010 has been integrated in an article related to the problem of quick detection of top-k Personalized PageRank (PPR) in [24]. The effectiveness of the chosen approach based on Monte Carlo methods for quick detection of top-k PPR lists has been demonstrated on the Web and Wikipedia graphs.

Yi-Ling Kuo during her internship has worked on Person Name Disambiguation and started by managing the analysis of the very huge Yahoo! Web graph.

This topic has been done in the context of Smirnova’s thesis [18] which has been defended on december 15 (thesis supervised by B. Trousse (AxIS) and K.Avrachenkov (Maestro)).

5.3.3. Towards an On-Line Analysis of Tweets Processing
Participant: Nicolas Béchet.

Tweets exchanged over the Internet represent an important source of information, even if their characteristics make them difficult to analyze (a maximum of 140 characters, etc.). In [25], we define a data warehouse model to analyze large volumes of tweets by proposing measures relevant in the context of knowledge discovery. The use of data warehouses as a tool for the storage and analysis of textual documents is not new but current measures are not well-suited to the specificities of the manipulated data. We also propose a new way for extracting the context of a concept in a hierarchy. Experiments carried out on real data underline the relevance of our proposal.

This work is done inside a collaboration with LIRMM and CEMAGREF.

5.4. Multidisciplinary Research For Supporting User Oriented Innovation

5.4.1. Usability Design and Evaluation Methods
Participants: Dominique Scapin, Yves Lechevallier, Pascal Marie-Dessoude, Claudia Detraux.

We pursued our work on articulation of usage mining approach and human factors expertise for the design and evaluation of information systems. Namely, collaborative clustering techniques were used to analyze data issued from users via a card sorting technique, with respect to an a priori (“expert”) clustering. Considering the difficulties that people have in managing large information sets in their everyday life, for either professional or non-professional purposes (administration, social relationships, leisure, etc.), our recent research focuses on personal information space for which information bits are currently scattered many places.
In this PIMs field, there is little research with a user-centric approach, with the view that users-based knowledge might help specifying computer-based tools and a state-of-the-art [37] showed little work specifically on usability. Studies address a variety of questions from user needs to accessibility (including studies on older people) or user acceptance, among others. In the context of user-centered and long-term studies to understand the evolution of user information practice, we looked in a study at the intuitive way people organize their personal information, with or without computer systems, in order to help the design of future systems. Also, we recently surveyed 15 tools that claim to support personal information management.

5.4.2. Living Lab Landscape

Participants: Marc Pallot, Brigitte Trousse, Bernard Senach, Dominique Scapin.

In order to provide to the research community a comprehensive landscape of research streams in the Living Lab domain, we launched a study on the state-of-the-art about the ubiquitous notion of User Experience. During this continuous study, a landscape [34], [69] has progressively emerged that we organized through 4 main axes: focus granularity (individual/group), user’s role in the design process (observed subject/value creator), collaboration style (structured/unstructured), and evaluation purpose (reliability/adoptability). Our landscape of research streams has been used by Finnish colleagues who conducted an empirical study on the use of the Living Lab research domain landscape as a tool for assessing the maturity level of 16 Finnish Living Labs [36]. The Living Lab research domain Landscape has allowed the study team to identify four categories of Living Labs.

5.4.3. Future Internet Domain Landscape

Participants: Marc Pallot, Brigitte Trousse, Bernard Senach.

There are many different Internet research areas and corresponding technologies that were already investigated, experimented and progressively deployed such as peer-to-peer, autonomous, cognitive and ad hoc networking, that have already demonstrated how to improve network performance and user experience. Peer-to-peer networking for large-scale distributed systems and widely used applications has proved both the feasibility and economic potential for delivering services to millions of users. Others emerged more recently in the context of the future Internet (FI), such as Cloud Computing for transparently sharing among users scalable elastic resources over a limitless network. As it remains difficult to visualise the conceptual evolution and articulate the various Internet research areas, we conducted a study for identifying the appropriate concepts that could populate the FI domain landscape [69], [71] over three different periods of time (1990-1999, 2000-2005 and 2006-2011). Several INRIA research teams are involved in FIRE (Future Internet Research Experimentation) Testbed projects, namely: PlanetLab, OneLab, TEFIS, SensLAB, and BonFIRE whose scientific leaders were interviewed during the development of the FI domain landscape. Four dimensions were used for landscaping the Future Internet research domain: evolution approaches (from incremental evolution design to Clean Slate re-design or radical evolution), Internet routing (from the basic data packet delivery towards more sophisticated content distribution and retrieval capacities such as content Centric Networking), network type (from wired communication to wireless communication networks), evolution trend (from computer network towards network computing). The resulting tentative landscape of FI research areas shown in Figure 1 is intended to provide a faster and broader understanding of the different Internet research streams and related topics.

5.4.4. Future of Internet and User-open innovation for Smart Cities

Participants: Caroline Tiffon, Marc Pallot, Brigitte Trousse, Bernard Senach.

The goal of the Fireball project is to bring together three different constituencies: user driven open innovation, Future Internet, and Smart Cities [34], [39]. It aims at defining a roadmap [29], based on analysis of needs, opportunities and gaps, to benefit a wide scale implementation of the methodologies and concepts elaborated. A first objective in the project was to get a clear picture of the state-of-the-art in each domain. During the review, progressively emerges a landscape [69] that we organized along 4 main axes: wiring (wired/wireless), user’s role (subject/actor), Internet evolution approach (structured/unstructured), evaluation purpose (reliability/adoptability). A large variety of FI research have been engaged. If initial efforts in Future
Internet research have been directed towards the goal of providing the technical infrastructure supporting the next network generation, a rising trend in this research field is to consider now a higher level layer, the layer of services.

5.4.5. Method and Tool for Selection of Open Innovation Software Tools

Participants: Mylène Leitzelman, Brigitte Trouse.

In spite of an important number of tools supporting open innovation, there is few comparative evaluation and no grid or evaluation criteria helping to choose a product. A 2011 review in the Computer Aided Innovation field provides a large overview of available tools in relation with a wide range of innovation cycle features. This useful top-down categorization approach is of little help to choose a specific tool. To find the best OI tool supporting idea exchanges among a community of participants, we built an exploratory method on the Web and we elaborate an assessment grid of OI tools based on the QSOS method which is a method designed to qualify, select and compare free and open source software in an objective, traceable and argued way. It publicly available under the terms of the GNU Free Documentation License. In our QSOS method, evaluation criteria are organized in a tree-hierarchy grid with, a scoring method procedure of each tree-leaf criteria (from 0 - not covered to 2 - completely covered). To achieve the construction of the OI tools criteria assessment grid (first step), we used different mining tools for Web crawling, network analysis, criteria classification and from the 29 top rated OI tools, we finally selected 6 of them from which we extracted the tree-map categorization used to build the reference software criteria sheet. In the following steps, after appropriate weighting, we used the provided OS3 Web application and we were able to compare 4 top selected OI tools [31].

To support our method, we developed a QSOS-based OI Grid for supporting the OI tool selection. The QSOS Grid to compare OI tools, which is in an XML format, has been translated into a MindMap. As a first mock-up, the QSOS Web Interface O3S will be installed at the beginning of 2012 on our server. For the future, we

---

6Häsig, University Regensburg & Khon, Otto Beisheim School of Management
will extend our own QSOS comparative method to other softwares and to other categories interesting for the Living Lab community.

![Figure 2](image)

**Figure 2. Radar Positioning of Selected Open Innovation Tools**

### 5.4.6. Extension of Methods for Idea Generation Process

**Participants:** Anne-Laure Negri, Caroline Tiffon, Bernard Senach, Brigitte Trousse.

Internet of Things (IoT) is supposed to be a physical world where everyday objects, rooms and machines are connected to one another and to the larger digital world. In this web of people and objects, individuals as well as things will have their own unique URL and, according to interactions between all these entities, social networks will be articulated with a ring of connected objects. This mash up of "hyper groups" and "hyper objects" will be the next step towards a deeper level of automation in which the user interface has vanished, the explicit control over the world is no longer the rule and where the users will have to understand the dynamic changes of their environment in relation to their social interactions and to their physical behaviour. The design of IoT based services therefore raises many challenges related to the user experience and requires a deep understanding of users’ needs in their real life environment as well as many field experiments; this is why the Living Lab approach appears as the qualification device which has been lacking so far. As described in a paper accepted for ServDes2012 conference (February 2012) among 80 papers submitted, we developed and tested two methods designed for Internet of Things service idea generation: GenIOT (Generative Technique for the Internet of Things) and Aloha! (Animation Lens: Object/Human Actoring!) methods.

- The **GenIOT** method aims at providing citizen a tool for identifying and sharing examples of situations where they may benefit from an Internet of Things based service. Indeed the objective is to help the introspective effort of analysing one’s own daily behavior and unveil situated data in order to develop grounded internet of things services ideation. Fake sensors are placed by the participants in their daily environment and pictures of the sensors in situ are shared and tagged on a collaborative platform. Other participants are invited to discuss online the ideas behind the pictures. Gamification rules are implemented in the platform in order to enhance collaboration and participation.

- The **Aloha!** method tackles another creative dimension as it asks participants to role or play characters or objects and bodystorm a collaborative scenario meeting the characters need. Participants report it to be a fun and effective method, alternative to traditional brainstorming and integrating serendipity as a creative asset.

For the exploration step in the context of ELLIOT, we developed NiceAir, an android mock-up for visualizing air quality data on Google map in Nice as well as some interest points (free bikes, free park places, bus stop, etc.). Such a mock-up will be available in the co-creation step of green services in the ELLIOT project and during the European Mobility Week.
5.4.7. Mock-ups for two innovation processes (exploration step)

**Participants:** Lucile Gramusset, Guillaume Pilot, Mohamed Gaieb, Bernard Senach, Brigitte Trousse.

In the context of two contracts (TICTAC and ELLIOT) related to user oriented innovation process, we have developed two mock-ups:

- Based on the user feedback from the first experiment in TICTAC ( Cf. section 6.1.1 ), we decided to develop a mock-up *MobilTIC* of a real-time information service related to public transportation (Envibus & SNCF) usable for any smartphone with internet access for Sophia-Antipolis workers. We designed a Web interface the most simple based on PHP/Mysql technologies and accessible from a computer or 3G mobile phone. Usage analysis of MobilTIC has been anticipated by providing rich logs. An exploration task with citizen is planned in January 2012 with a new experiment with an improved version.

- A first mock-up called *Nice Air* has been developed in the context of ELLIOT ( Cf. section 6.3.1.1 ) by L. Gramusset and M. Gaieb on android smartphone for providing information related to air quality and noise on a map of the area of Nice Cote d’Azur with some interest points such as bus stops and Vélib parkings.

5.5. FocusLab Experimental Platform and Software

5.5.1. ATWUEDA based Clustering Approach for System Monitoring

**Participant:** Yves Lechevallier.

Progressive advances in hardware and software technologies have enabled the production and storage of system monitoring data streams in a wide range of fields (e.g. telecommunications, sensor networks, etc.). Traditional clustering methods are unable to deal with data of such a voluminous and dynamic nature. In this work [51], we propose an efficient clustering approach (ATWUEDA) for monitoring massive time-changing data streams. This work considers a real case study on condition monitoring data streams of an electric power plant provided by EDF.

This work is done in collaboration with Alzennyr Da Silva of BILab laboratory (Telecom ParisTech and EDF R&D Common Laboratory).

5.5.2. FocusLab Experimental Platform (CPER Telius 2008-2012)

**Participants:** Xavier Augros, Mohamed Gaieb, Brigitte Trousse, Yves Lechevallier.

The FocusLab platform aims to be a major delivery mechanism of previous and current work in AxIS. It is a way to make methodological contributions (including software) available for the scientific community, but also a way for stimulating further research. This work has slowly started on the software part due to the absence of human resource funding and due to the absence of engineers in the team until 2010. Mid 2011, we started the specification and the development of a first version of a platform with three parts (hardware, software and methods) with the arrival of two engineers on the ELLIOT and TIC TAC contracts. A first version is available since the end of september as a Web portal and a second one is planned for the end of 2012 with advanced features for the software part (cf. service oriented platform, SOA architecture and interoperability).

Related to the software part we are in the process of developing several AxIS methods as Web services: we started with SCDS ( cf. 4.2.3 ) which was demonstrated in the context of the MIDAS project on two applications (Orange labs mobile portal and vehicle trajectories) and in the ELLIOT platform (linked to the San rafaelle Hospital media use case) at the first review meeting ( cf. 6.3.1.1 ) applied on data issued from San Rafaele Hospital use case. ATWUEDA ( cf. 4.3.2 ), GEAR (Marascu’s thesis) are under development as web services.

Our work on mining evolutive data (ATWUEDA) and data streams (such as SCDS) have been used in real applications in the context of Internet of things and sensors: ATWUEDA (system monitoring for EDF (cf. section 5.5.1 ) and SDMS (cf. section 4.2.3 ).
DAHU Project-Team

5. New Results

5.1. Specification and verification of database driven systems

Participants: Serge Abiteboul, Pierre Bourhis, Luc Segoufin, Szymon Toruńczyk, Victor Vianu.

Modelization and verification of data centric systems. We have intensively studied the Active XML model. It is a high-level specification language tailored to data-intensive, distributed, dynamic Web services. Active XML is based on XML documents with embedded function calls. The state of a document evolves depending on the result of internal function calls (local computations) or external ones (interactions with users or other services). Function calls return documents that may be active, so may activate new subtasks. Our first line of result is a comparison of the specification power of various workflow control mechanisms within the Active XML framework and beyond [23]. AXML is very powerful and many static analysis problems are undecidable. We have also introduced a model of automata designed for modeling infinite runs of systems equipped with static relational databases. The automata model is equipped with finitely many variables, each of which can store values from a linearly ordered domain, such as the natural numbers. The transitions of the automata depends on a conjunctive query involving the database and the current values of the variables. For verifying infinite runs of such automata, an extension of temporal logic is considered, capable of comparing values stored in the variables and the database, at different times of the run. The main contribution of the work is the proof that automated verification of such temporal properties of the system can be carried out in PSpace. For more details, see [35].

Static analysis of query languages. XPath is arguably the most widely used XML query language as it is implemented in XSLT and XQuery and it is used as a constituent part of several specification and update languages. Hence in order to perform static analysis on a system manipulating XML data it is important to master the static analysis for XPath. Most of the important static analysis problems reduce to satisfiability checking: does a given query return a non-empty answer on some data. In general, in the presence of data values, the satisfiability of XPath is undecidable. We have shown that when restricted to its vertical navigational power, XPath becomes decidable [30].

5.2. Distributed data management

Participants: Serge Abiteboul, Emilien Antoine, Daniel Deutch, Alban Galland, Wojciech Kazana, Yannis Katsis, Luc Segoufin, Cristina Sirangelo.

Distributed knowledge base. As a foundation for managing distribution, we have proposed a model of a distributed knowledge base, that handles data and meta-data, as well as access control and localization, in a unique integrated setting. To support automatic reasoning on this knowledge base, we also introduced a novel rule-based language supporting the exchange of rules, namely Webdamlog. This work has been presented [21] and demonstrated [26] at major database conferences.

Probabilistic XML. Data from the Web are imprecise and uncertain. To manage this imprecision in a well-principled way, we have made significant advances in the field of probabilistic databases, and specifically, probabilistic XML. We have introduced new tractable probabilistic models for representing uncertain hierarchical information, and carried out in-depth studies of query evaluation, aggregation, and updates in various probabilistic XML models. These results have matured and some of the results are available in journal articles, e.g., [14].
Enumeration of query answers. In many applications the output of a query may have a huge size and enumerating all the answers may already consume too many of the allowed resources. In this case it may be appropriate to first output a small subset of the answers and then, on demand, output a subsequent small numbers of answers and so on until all possible answers have been exhausted. To make this even more attractive it is preferable to be able to minimize the time necessary to output the first answers and, from a given set of answers, also minimize the time necessary to output the next set of answers - this second time interval is known as the delay. We have shown that this was doable with a linear preprocessing time and constant enumeration delay for first-order queries over structures of bounded degree [19].

Data exchange and Web incomplete information. We have addressed the problem of restructuring data exchanged between communicating applications on the Web. We have proposed and analyzed a new language to specify data restructuring rules (schema mappings). This language generalizes existing mapping dependencies, by allowing a more flexible specification mechanism [20].

Jorge. We also invested a lot of effort in a textbook (undergraduate and graduate level) on Web data management (nicknamed Jorge) to be published at Cambridge University Press [38]. The book is already available on the Webdam Web site http://webdam.inria.fr/Jorge

5.3. Tree automata theory

Participants: Stéphane Demri, Florent Jacquemard, Luc Segoufin.

Most of our results for this section concerns data words and data trees. Those are words and trees where each position contains a data value together with the classical label. Data trees can be seen as a model for XML data. We have studied automata model using registers or memory or allowing constraints that can involve data comparisons in its transitions.

Register Automata. These extend the classical model of finite automata with auxiliary registers storing data values for later comparison.

We have introduced a new model of automata over data trees and shown the decidability of its emptiness problem [30]. These automata were used for obtaining decidability results for the static analysis for some fragments of XPath presented in the previous section.

Automata with counters. In [39], a survey chapter on the verification of infinite-state systems is presented that is focused on the verification of counter systems. Verification problems for vector addition systems and recursive Petri nets are considered. Moreover, we introduce subclasses of counter systems for which reachability questions can be solved in Presburger arithmetic viewed as a means to symbolically represent sets of tuples of natural numbers.

Automata with isomorphism tests among subtrees. We have also considered some models described by tree automata enriched with a feature testing for isomorphisms between subtrees. Such constraints could be used for testing monadic key constraints over XML documents. For these models, the main challenge is to establish the decidability of the non-emptiness of the language specified by a given automaton [18].

Rewriting Controlled by Selection Automata. Motivated by the problem of static analysis of XML update programs, we have studied [33] the combination, called controlled term rewriting systems (CTRS), of term rewriting rules with constraints selecting the possible rewrite positions. These constraints are specified, for each rewrite rule, by a selection automaton which defines a set of positions in a term based on tree automata computations. We have established several decidability and complexity results for several cases of the reachability and regular model checking problems for this tree transformation formalism.
DREAM Project-Team

6. New Results

6.1. Diagnosis of large scale discrete event systems

**Participants:** Marie-Odile Cordier, Christine Largouët, Sophie Robin, Laurence Rozé, Yulong Zhao.

The problem we deal with is monitoring complex and large discrete-event systems (DES) such as an orchestration of web services or a fleet of mobile phones. Two approaches have been studied. The first one consists in representing the system model as a discrete-event system by an automaton. In this case, the diagnostic task consists in determining the trajectories (a sequence of states and events) compatible with the sequence of observations. From these trajectories, it is then easy to determine (identify and localize) the possible faults. In the second approach, the model consists in a set of predefined characteristic patterns. We use temporal patterns, called chronicles, represented by a set of temporally constrained events. The diagnostic task consists in recognizing these patterns by analyzing the flow of observed events.

More recently, we started research on interacting with large-scale systems in a decision-oriented way. Scenario patterns were defined for exploring complex systems, based on the use of model-checking techniques.

6.1.1. Distributed monitoring with chronicles - Interleaving diagnosis and repair - Making web services more adaptive

Our work addresses the problem of maintaining the quality of service (QoS) of an orchestration of Web services (WS), which can be affected by exogenous events (i.e., faults). The main challenge in dealing with this problem is that typically the service where a failure is detected is not the one where a fault has occurred: faults have cascade effects on the whole orchestration of services. We have proposed a novel methodology to treat the problem that is not based on Web service (re)composition, but on an adaptive re-execution of the original orchestration. The re-execution process is driven by an orchestrator Manager that takes advantage of an abstract representation of the whole orchestration and may call a diagnostic module to localize the source of the detected failure. It is in charge of deciding the service activities whose results can be reused and may be skipped, and those that must be re-executed. A paper has been submitted to the CAISE conference.

6.1.2. Scenario patterns for exploring qualitative ecosystems

Our work aims at giving means of exploring complex systems, in our case ecosystems. We proposed to transform environmental questions about future evolution of ecosystems into formalized queries that can be submitted to a simulation model. The system behavior is represented as a discrete event system described by a set of interacting timed automata, the global model corresponding to their composition on shared events. To query the model, we have defined high-level generic patterns associated to the most usual types of scenarios. These patterns are then translated into temporal logic formula. The answer is computed thanks to model-checking techniques that are efficient for analysing large-scale systems. Five generic patterns have been defined using TCTL (Timed Computation Tree Logic): WhichStates, WhichDate, Whichstates, Stability, Safety. Three of them have been implemented using the model-checker UPPAAL.

The approach has been experimented on a marine ecosystem under fishing pressure. The model describes the trophodynamic interactions between fish trophic groups as well as interactions with the fishery activities and with an environmental context. A paper has been accepted for publication by the Environmental Modelling Software Journal [4].

We extended the approach to deal with “How to” queries. As before, we rely on a qualitative model in the form of timed automata and use model-checking tools to answer queries. We have recently proposed two approaches to answer questions such as “How to avoid a given situation?” (safety query). The first one exploits controller synthesis and the second one is a “generate and test” approach. We compared these two approaches in the context of an application that motivates this work, i.e the management of a marine ecosystem and the evaluation of fishery management policies. The results have been accepted for publication in [14].
6.2. Machine learning for model acquisition

Participants: Thomas Guyet, René Quiniou.

Model acquisition is an important issue for model-based diagnosis, especially as modeling dynamic systems. We investigate machine learning methods for temporal data recorded by sensors or spatial data resulting from simulation processes. We also investigate efficient methods for storing and accessing large volume of simulations data. Our main interest is extracting knowledge, especially sequential and temporal patterns or prediction rules, from static or dynamic data (data streams). We are particularly interested in mining temporal patterns with numerical information and in incremental mining from sequences recorded by sensors.

6.2.1. Mining temporal patterns with numerical information

We are interested in mining interval-based temporal patterns from event sequences where each event is associated with a type and time interval. Temporal patterns are sets of constrained interval-based events. This year, we have been working on improving the formal setting of the approach as well as its efficiency [8]. We have introduced the notion of \(\epsilon\)-covering of temporal patterns over sequences to cope with the dual nature, symbolic and numerical, of temporal patterns. The parameter \(\epsilon\) specifies the tightness of the similarity used for matching patterns and sequences. It complements the parameter \(\sigma\) representing the minimal support which is used to prune candidate patterns. The \(\epsilon\)-similar occurrences of some pattern, precisely their associated temporal intervals, are classified to characterize the different classes of numerical temporal intervals that correspond to different patterns sharing the same symbolic part. This process have been embedded in two sequential pattern mining algorithms, GSP and PrefixSpan, and we have compared their performance.

6.2.2. Incremental sequential mining

We investigate the problem of mining and maintaining frequent sequences in a window sliding on a stream of itemsets. We propose in [11] a complete and correct incremental algorithm based on a tree representation of frequent sequences inspired by PSP [52] and a method for counting the minimal occurrences of a sequence. Instead of the frequency, to a node representing a pattern is associated the set of occurrences of this pattern. The algorithm updates efficiently the tree representation of frequent sequences and their occurrences by means of two operations on the tree: deletion of the itemset at the beginning of the window (obsolete data) and addition of an itemset at the end of the window (new data). Experiments were conducted on simulated data and on real data of instantaneous power consumption.

6.2.3. Multiscale segmentation of satellite image time series

Satellite images allow the acquisition of large-scale ground vegetation. Images are available along several years with a high acquisition frequency (1 image every two weeks). Such data are called satellite image time series (SITS). In [9], we present a method to segment an image through the characterization of the evolution of a vegetation index (NDVI) on two scales: annual and multi-year. We test this method to segment Senegal SITS and compare our method to a direct classification of time series. The results show that our method using two time scales better differentiates regions in the median zone of Senegal and locates fine interesting areas (cities, forests, agricultural areas).

6.3. Decision aiding with models and simulation data

Participants: Tassadit Bouadi, Marie-Odile Cordier, Véronique Masson, Florimond Ployette, René Quiniou, Karima Sedki.

Models can be very useful for decision aiding as they can be used to play different plausible scenarios for generating the data representing future states of the modeled process. However, the volume of simulation data may be very huge. Thus, efficient tools must be investigated in order to store the simulation data, to focus on relevant parts of the data and to extract interesting knowledge from these data.
6.3.1. Exploring models thanks to scenarios: a generic framework

In the framework of the APPEAU project (see 8.2.1), that ended in December 2010, a paper, describing a generic framework for scenario exercises using models applied to water-resource management, has been written during 2011 in cooperation with all the partners and submitted to Environmental Modelling and Software. It is currently under revision.

6.3.2. A datawarehouse for simulation data

The ACASSYA project aims at providing experts or stakeholders or farmers with a tool to evaluate the impact of agricultural practices on water quality. As the simulations of the deep model TNT2 are time-consuming and generate huge data, we have proposed to store these simulation results in a datawarehouse and to extract relevant information, such as prediction rules, from the stored data. We have devised a general architecture for agro-environmental data on top of the framework Pentaho. An article presenting the principles of this architecture as well as a set of realistic scenarios and their transformation into OLAP queries has been submitted to Compag (Computers and Electronics in Agriculture).

6.3.3. Efficient computation of skyline queries in an interactive context

Skyline queries retrieve from a database the objects that optimizes multiple criteria, related to user preferences for example, or objects that are the best compromises satisfying these criteria. When data are huge such objects may shed light on interesting parts of the dataset. However, computing the skylines (i.e. retrieving the skyline points) may be time consuming because of many dominance tests. This is, especially the case in an interactive setting such as querying a data cube in the context of a datawarehouse. We have worked on how to answer efficiently to skyline queries by the materialization of precomputed skyline queries related to dynamic user preferences. An article has been submitted to the conference SIGMOD 2012.

6.3.4. Influence Diagrams for Multi-Criteria Decision-Making

For multi-criteria decision-making problems, we propose in [6] a model based on influence diagrams able to handle uncertainty, represent interdependencies among the different decision variables and facilitate communication between the decision-maker and the analyst. The model makes it possible to take into account the alternatives described by an attribute set, the decision-maker’s characteristics and preferences, and other information (e.g., internal or external factors) that influence the decision. Modeling the decision problem in terms of influence diagrams requires a lot of work to gather expert knowledge. However, once the model is built, it can be easily and efficiently used for different instances of the decision problem. In fact, using our model simply requires entering some basic information, such as the values of internal or external factors and the decision-maker’s characteristics.

6.3.5. Recommending actions from classification rules

In the framework of the SACADEAU project (see 8.2.1), a paper dedicated to building recommendation actions for a given situation, from the set of classification rules, learnt from simulation results, has been published in the KAIS journal [7].

6.4. Causal reasoning and influence diagrams

Participants: Philippe Besnard, Louis Bonneau de Beaufort, Marie-Odile Cordier, Yves Moinard, Karima Sedki.

This work stems on [23], [24], [25], [26], [27] and, for the logic programming translation, on [53], [54]. It is related to diagnosis (observed symptoms explained by faults).

The previously existing proposals were ad-hoc or, as in [29], [41], they were too close to standard logic in order to make a satisfactory diagnosis. Our proposal starts from a restricted first order logic (of the Datalog kind: no function symbols) and introduces causal formulas, built on causal atoms such as (α causes β) intended to mean: “α causes β”. The system is described thanks to these causal formulas, classical formulas, and taxonomy atoms such as (α IS_A β) (α is of kind β).
The system produces explanation atoms of the kind \((\alpha \text{ explains } \beta \text{ if possible } \{\gamma_1, \cdots, \gamma_n\})\), meaning that \(\beta\) can be explained by \(\alpha\) if all the \(\gamma_i\)'s are possible together in the context of the given data.

This year, we have improved our logic programming translation in ASP. The aim is to improve efficiency and also reduce the work of the programmer, taking advantage of the declarative aspect of this type of programming. We have applied some of these improvements to two classic riddles, in order to illustrate the power and limitations of current answer set programming systems, and we proposed a few improvements which could make the present systems yet easier to use [12], [13].

We are starting a work with some similarities to automatize the treatment of cognitive maps. The aim is to extract relevant information from these maps, which means: building a graph formalism for representing mixed causal and influence relations, and defining a framework (argumentation theory is a good candidate) to aggregate the graphs and provide inference rules in order to infer new information and relations. This work is done in the framework of the RADE2BREST project, involving Agrocampus Ouest and CNRS (GEOMER/LETG), funded by “Ministère de l’Ecologie”\(^1\). The goal of this project is to model shellfish fishing in order to assess the impact of management pollution scenarios on the Rade de Brest. The cognitive maps result from interviews with fishermen.

\(^1\)This project is not mentioned in section 8.1 because DREAM is not a partner of this project.
6. New Results

6.1. Graph Based Knowledge Representation

6.1.1. Knowledge Graph Abstract Machine

Participants: Olivier Corby, Catherine Faron-Zucker, Fabien Gandon.

KGRAM (Knowledge Graph Abstract Machine) is a generic interpreter for W3C SPARQL Query Language that operates not only on RDF graphs but on labelled graphs. The interpreter interacts with the target graph through proxies that implement an interface: *Producer* enumerates edges from the target graph, *Evaluator* evaluates filters and *Matcher* takes entailments into account.

This year, work have been done to leverage KGRAM up to SPARQL 1.1 Query Language & Update. It implements most of current version of the recommendation, except the *service* statement. It passes almost all W3C SPARQL 1.1 test cases.

In addition, the Corese Semantic Web Factory has been redesigned and modularized into release 3.0 entirely based on KGRAM interfaces and proxies. Corese 3.0 is a new lightweight RDF/S implementation with SPARQL 1.1. We ported the former Inference Rule engines (forward and backward engines) onto Corese 3.0. We also ported former SPARQL extensions: approximate search based on ontological distance, SQL and XPath in SPARQL 1.1, edge enumeration and length of Property Path, pragmas.

This new version is already used in several applications among which: cartography at IGN [28], design constraint modeling at CSTB [35], technological watch in ISICIL ANR project. It is also used in several PhD Theses in the team. A list of applications can be found on Corese Web site6.

6.1.2. Semantic Web Graph Visualization

Participants: Olivier Corby, Nicolas Delaforge, Erwan Demairy, Fabien Gandon [contact].

Thanks to an INRIA grant (ADT), we design and develop a Semantic Web Gephi Plugin. This plugin is coupling Corese and the Gephi Open Graph Visualization Platform to provide a framework to query and visualize RDF data taking into account their schemas. See the web pages7.

6.1.3. Semantic Social Network Analysis

Participants: Guillaume Erétéo, Fabien Gandon.

The PhD thesis of Guillaume Erétéo [14] in the context of the ANR project ISICIL allowed us to analyze the characteristics of the heterogeneous social networks that emerge from the use of web-based social applications, with an original contribution that leverages Social Network Analysis with Semantic Web frameworks. Social Network Analysis (SNA) proposes graph algorithms to characterize the structure of a social network and its strategic positions.

Semantic Web frameworks allow representing and exchanging knowledge across web applications with a rich typed graph model (RDF), a query language (SPARQL) and schema definition frameworks (RDFS and OWL). In this thesis, we merged both models in order to go beyond the mining of the link structure of social graphs by integrating two approaches: (1) semantic processing of the network typing and (2) emerging knowledge of online activities.

---

6 http://www-sop.inria.fr/teams/edelweiss/software/corese/
8 https://gforge.inria.fr/projects/segviz-public/
In particular we investigated how (1) to bring online social data to ontology-based representations, (2) to conduct a social network analysis that takes advantage of the rich semantics of such representations, and (3) to semantically detect and label communities of online social networks and social tagging activities. This work was published at [15], [14].

6.1.4. Index Summarizing the Content of RDF Triple Stores

Participants: Adrien Basse, Fabien Gandon, Isabelle Mirbel.

We are interested in designing an architecture to support the distribution of a SPARQL query on a small and fixed number of RDF repositories. To do so, the key stage is to characterize the content of the base of each server in order to be able to predict if a server could contribute or not to the answer of a query. In the context of the PhD Thesis of Adrien Basse we propose an algorithm to extract a compact representation of the content of an RDF store. We improved the canonical representation of RDF graphs based on DFS code proposed in the literature by providing a join operator to reduce the number of generated redundant patterns.

6.1.5. Rules for the Web of Data

Participants: Oumy Seye, Olivier Corby.

In the context of this PhD thesis, the focus is on Rules for the Web of data. We are interested in integrating Rule Interchange Format (RIF) - W3C recommendation for exchanging rules on Web - to others W3C technologies. The aim of this year is to study the integration possibilities of RIF-BLD into semantic Web technologies. RIF-BLD is the dialect of RIF for logic-based systems. Firstly, we have studied the state of the art. Secondly we improved the RIF-BLD parser for presentation syntax and XML syntax. As RIF-BLD can be used with RDF data and OWL ontologies, it is interesting to consider RIF inferences in queries on RDF graph structure. That is why we finally study the integration of RIF-BLD into the Corese Semantic Web engine. In this last step, we have implemented the mapping of abstract syntax tree of RIF-BLD to abstract syntax tree of SPARQL. Thus, we can now execute logic inferences of RIF-BLD in the backward engine of Corese.

We have a paper accepted at EGC 2012 presenting RIF2SPARQL [44], a translation of RIF-BLD statements in SPARQL to perform the logical inferences of RIF-BLD on the Corese Semantic Web Factory. These inferences are implemented in backward chaining approach. We have designed and implemented the mapping of RIF-BLD to SPARQL.

6.1.6. Collaborative Management of Interlingual Knowledge

Participants: Maxime Lefrançois, Fabien Gandon.

We are interested in bridging the gap between the world of natural language and the world of the Semantic Web, in particular to support multilingual access to the Web of Data and management of interlingual knowledge bases. We introduce the ULiS approach, that aims at designing a pivot-based NLP technique called Universal Linguistic System, using Semantic Web formalisms, and being compliant with the Meaning-Text theory. Through ULiS, a user could interact with an Interlingual Knowledge base (IKB) in controlled natural language. Linguistic resources themselves (e.g. dictionary, grammar) are part of a specific IKB, thus, actors may enhance them (i.e. the model of the controlled natural language), through requests in controlled natural language (e.g., add a new lexical units, add grammar rules).

In [30] we proposed a novel approach to define Interlingual Lexical Units classes in the Interlingual Lexical Ontology so that they support the projection of their lexicographic definition on themselves using the OWL formalism. This approach is compliant with the Meaning-Text Theory.

In [31], [40] we introduced three basic interaction scenario for ULiS and we proposed and overviewed the layered architecture of ULiS: meta-ontology, ontology, facts; and ontology, interlingual knowledge, situational knowledge.

We have started a collaboration with the RELIEF project that deals with the construction of a French Lexical Network (Alain Polgùère, CNRS-ATILF).
6.1.7. Reuse of Data Analytics Contents and Processes

Participants: Corentin Follenfant, Fabien Gandon, Olivier Corby.

Industrial Business Intelligence (BI) proposes tools and methods to perform data analysis over heterogeneous enterprise sources. They allow one to harvest, federate, cleanse, annotate, query, organize and visualize data in order to support decision making with human-readable documents such as reports, dashboards, mobile visualizations. Such processes currently require expertise in technical domains like relational modeling in order to produce relevant content.

Users willing to do so without following the learning curve have to reuse existing content to create new one, and need to be guided throughout the workflow. Recommender systems can contribute to easing their progression, but most of them will operate inside walled garden for specific tasks instead of assisting the user throughout his workflow.

Semantic Web tools allow us to provide a common ground for modeling the different operations that compose BI workflows with RDFS vocabularies, capturing usage of the underlying transformations operators within document repositories with RDF graphs, and enabling further composition and reuse of BI operations to achieve new analysis. We introduced with [38] an extension of the RDF Data Cube vocabulary\(^9\) to describe these operations as flexible services that are composed by matching multidimensional data structures interfaces, and validated this model on a production repository containing 900 BI documents decomposed into 8000 documents snippets.

The underlying sequence of operations specific to each snippet was then extracted into a unique RDF graph. Aggregate SPARQL queries allow us to compute basic usage statistics for BI operations that can feed recommender systems such as BI workflows wizards. Besides refining the proposed model, next steps include evaluating the technical usability of SPARQL property paths patterns for data lineage and to identify frequent patterns in sequences of BI operations.

This PhD Thesis is done with a CIFRE industrial grant from SAP Research.

6.2. Interaction Design

6.2.1. Question Answering over Linked Data

Participant: Elena Cabrio.

While an increasing amount of semantic data is being published on the Web, the crucial issue of how typical Web users can access this body of knowledge comes to light. This PostDoc project focuses on the development of methods for a flexible mapping between questions in natural language, and data objects. The main purpose is to allow an end user to submit a query to an RDF triple store in English and get results in the same language, masking the complexity of SPARQL expressions and RDFS/OWL inferences involved in the resolution, but at the same time profiting from the expressive power of these standards. In particular, we address the problem of automatic identification of the relevant relations in Question Answering (QA), to capture the context in which the requests should be interpreted, to be able to determine the constraints on the database query.

We aim at investigating the applicability of the Textual Entailment\(^10\) (TE) approach, recently proposed as a general framework for applied semantics, where linguistic objects are mapped by means of semantic inferences at a textual level [55]. According to such framework, entailment relations can be detected between an input question and a set of relational patterns that represent possible lexicalizations of the relations of interest. Such relations, collected in a pattern repository, can be associated to a SPARQL query to the database. A TE system should therefore first try to establish an entailment relation between an input question and each of the relations in the pattern repository. Then, the SPARQL queries associated to the relations for which the entailed patterns have been found are composed in a single query to the database.

Since this PostDoc research work has just started, our early efforts were directed toward the study of the state of the art on QA over Linked Data. We are currently carrying out a feasibility study on the extraction of the relational patterns from Wikipedia (as the source of free text) and the use of DBpedia\textsuperscript{11} as a linked data resource. For the experimental part, we are considering energy and environment as the reference scenario.

6.2.2. Mobile Access to the Web of Data

Participants: Luca Costabello, Fabien Gandon.

This thesis, directed by F. Gandon and I. Herman (CWI and Semantic Web Activity Lead at W3C) deals with accessing the Web of Data from mobile environments. The first year addressed the multi-faceted relationship between ubiquitous consumption of Linked Data and mobile context. More specifically, focus has been put on RDF adaptive representation and on context-aware SPARQL endpoints access control.

When accessed from devices immersed into ubiquitous environments, RDF resources must be properly adapted to the mobile context in which the consumption is performed. A domain-independent, lightweight vocabulary for displaying Web of Data resources in mobile environments has been designed (PRISSMA, Presentation of Resources for Interoperable Semantic and Shareable Mobile Adaptability [36]). The vocabulary is the first step towards an adaptive rendering engine for RDF data coupled with a declarative framework to share and re-use presentation information for context-adaptable user interfaces for Linked Data.

Another line of research regards the role of mobile context in restricting access to the Web of Data. Ubiquitous connectivity enables new scenarios in consuming Linked Data and access control in such pervasive environments must not ignore the mobile context in which RDF consumption takes place, as uncontrolled access in given situations may be undesired by data providers. The work led to enhance the access control framework for SPARQL endpoints proposed by teammate Serena Villata (see 6.2.3) with the notion of mobile context provided by PRISSMA.

6.2.3. Access Control for the Web of Data

Participant: Serena Villata.

This research activity is mainly focussed on the field of Knowledge Representation. First, we have proposed a new access control model for the Web of Data and the Social Web. In particular, we have introduced the S4AC ontology\textsuperscript{12} where the meaning of the access policies and their components is defined. This access control model proposes, also, a contextual evaluation of the accessors’ information. This model has been applied both to the world of Linked Data and to the world of social networks. This research activity has been addressed in the context of the DataLift ANR project [21], [20].

Second, we have continued a research activity in the area of argumentation theory. In this context, we are exploring the use of argumentation theory for modeling trust in those systems which deal with incomplete knowledge, and for providing explanations about the agents’ choices [22], [19], [23], [25].

6.2.4. ISICIL

Participants: Nicolas Delaforge, Fabien Gandon.

As the leading team of the ISICIL project, we have developed many software components (client-side and server-side) to enrich the ISICIL platform. First, the whole server mechanism was redesigned, in order to improve the server performance, to strengthen and modularize the framework as well. Many semantic REST services were added (activity stream, syndication, subscription/notification, graphs and charts visualizations).

In collaboration with Erwan Demairy, in charge of the SegViz ADT, a Gephi-ISICIL connector was implemented, allowing ISICIL users to visualize the results of their SPARQL queries directly into a dynamic graph. A demo of this work was presented during the ISICIL public seminarium in September. Furthermore, projects such as Datalift and ISICIL had brought out the need of an access control model for the Web of Data. For this purpose, we designed the S4AC model and ontology and we realized a prototype to evaluate it based on the ISICIL dataset Figure 4.

\textsuperscript{11} http://dbpedia.org
\textsuperscript{12} http://ns.inria.fr/s4ac/
Since the Philoweb conference in 2010, a workshop dedicated to *philosophical engineering* was attached to the French IC conference in Chambery. We presented there the advancement of a brand new bookmark model called *Webmarks* which semantically models the user interest on a web resource (Figure 5 & 6). A long paper on this work was accepted in the EGC 2012 conference and will be published in the RNTI journal (Hermann editions) [27].

We also collaborate with the I3S team in a task of *semantization* of a commercial wiki, called Mindtouch. This wiki is enhanced with semantic description of its content (Figure 7), its users are part of the ISICIL social network and their activities on the wiki are reported on the metadata server. This tool represents the editorial layer of the software bundle designed to improve the business intelligence tasks. This work was also accepted as a long paper in the EGC 2012 conference [26].

![Figure 4. S4AC Access Policy Editor](image)

### 6.2.5. Models and Methods for Representing and Identifying Groups of Individuals and Their Activities

#### 6.2.5.1. Models and Methods for Representing and Identifying "Collective Personas"

**Participant:** Alain Giboin.

**Context of the work:** ISICIL project.

As opposed to Individual Personas (which are user models represented as specific, realistic humans), Collective personas are models representing specific, realistic groups of people as such (e.g., teams, communities). Collective personas are aimed to design groupware more closely adapted to groups. In 2010, we updated our review of the existing methods for elaborating collective personas. This year, we published the updated review [17].
Figure 5. Webmark GUI
Figure 6. Webmark Model
Figure 7. SweetDeki Resource Model
6.2.5.2. Models and methods for Representing and Identifying Relationships between Individuals

**Participants:** Alain Giboin, Neji Bouchiba.

Context of the work: AVISICIL project, in collaboration with researchers from the Kewi team (IIS, UNS) and from the Laboratoire de Psychologie Cognitive et Sociale (UNS) who are involved in affective computing design projects (designing systems intended to help elderly people maintain their relationships, or autistic children to build relationships with others).

Digital technologies have been claimed to contribute to prevent elderly people from social isolation or loss of social ties. For example, ubiquitous computing, online social networking and affective computing have been reported to facilitate social interaction [64] or to enhance social connectedness [61] among the elderly.

Participating to a project aimed to design a system for recognizing, through various sensors, the affective states (emotions) that indicate a loss or maintenance of social ties, we conducted a social ergonomic study to provide elements of design and evaluation of such a system. Noting that depressive states are among the most significant signs of an actual or potential loss of social ties (see, e.g., [65]), we focused the study in particular on: (a) the models describing the depressive states and the process of their recognition, and the links between these states and the state of social ties; (b) the sensors that can contribute to this recognition. In order to evaluate our solution (so-called GeREmo) with the elderly, we also identified, from an analysis of existing studies on the acceptability of digital technologies, criteria for assessing the acceptability of the GeREmo solution [50].

6.2.6. Comparing and Bridging Models of Shared Representations and Representation Sharing Processes

**Participant:** Alain Giboin.

Context of the work: GDR CNRS Psycho Ergo, Groupe thématique Coopération homme-homme et Coopération homme-machine. Action de recherche RefCom (Référentiel commun), co-led by Pascal Salembier (UTT).

Sharing representations or shared representations are often claimed to be a key factor for a collaboration to succeed. The notions of shared representations and representation sharing processes are examined in the research literature from several points of view; this variety of viewpoints gave rise to different conceptualizations, which are referred to using such terms as Common Frame of Reference, Mutual Intelligibility, Shared Context, Team/Situation Awareness, etc. In 2010, in order to achieve mutual intelligibility between researchers working on such conceptualizations, we elaborated and asked participants to the RefCom joint research action to test and to apply a grid for collaboratively comparing and bridging the conceptualizations (see Edelweiss activity report 2010). This year, we analyzed and reported the results of the test and application of the grid [39]. This resulted in a revision of the grid.

6.2.7. Frameworks for taking pragmatic dimensions of ontologies into account

**Participant:** Alain Giboin.

Context of the work: Follow-up to the Palette European project. This work was done in collaboration with the Centre de Recherche sur l’Instrumentation, la Formation et l’Apprentissage, ULg (Belgium).

When designing ontologies, ontologists (i.e., knowledge engineers specialized in ontology engineering) most often focus on the semantic dimensions of ontologies (such as expressiveness, level of granularity, etc.). Pragmatic dimensions, i.e. dimensions related to the context of use (including the purpose) of the ontologies, are often neglected whereas they are critical to users: ontologies indeed are used in context. In brief, pragmatic dimensions are not taken seriously into account when engineering ontologies but they have to.

---

We developed a framework to analyze the way we attempted, in the context of the Palette EU project, to contextualize the ontologies underlying a set of semantic knowledge services dedicated to communities of practice. The framework was derived from the Ontology Framework elaborated by members of the Ontology Engineering community during the Ontology Summit 2007 [62]. Both frameworks define a series of “pragmatic dimensions” of ontologies. Because our derived framework did not cover all possible dimensions, we complemented it, by relying on existing work from the Ontology Engineering community in general, and from the Pragmatic Web community in particular [16].

6.2.8. Explanation of Semantic Web Query Results

**Participants:** Rakebul Hasan, Fabien Gandon, Olivier Corby.

This PhD thesis, directed by Fabien Gandon and Olivier Corby, aims at opening the query-solving mechanism to the users, and handling and explaining the distribution of a query over several sources on the Semantic Web. This work is part of the Kolflow ANR project.

The current Semantic Web search engines are not able to explain how a given query result is obtained or why it has failed to obtain a result. The goal of opening the query-solving mechanism is to enable the Semantic Web query engines to explain the query solving process taking into account the inferences used to obtain the results for a given query. In addition, explanation of the performance indicators of the query-solving process contributes to the understanding of the resolution process. These performance indicators can be effectively used to help in formulating queries by suggesting alternative queries based on the history of the performance of the query-solving process. Another focus of this thesis is on how the distribution of the queries can be performed over the distributed sources and how explanation can be used to better understand the queries and their results over the distributed sources.

In the early stage of this thesis, our current focus is on explaining the Semantic Web query results taking the inferences into account. We are working on justification of results for SPARQL query with RDFS entailment. Our next focus will be on the different abstractions of these justifications with different degree of details and different types of presentations depending on different level of user expertise.

6.2.9. Pervasive Sociality through Social Objects

**Participants:** Nicolas Marie, Fabien Gandon.

The work is related to semantic spreading activation algorithm, from idea to first results and visualization. Spreading activation is a method for searching semantic networks by labeling a set of initial nodes with weights (called activation), propagating (spreading) that activation out to other nodes linked to the source nodes and iterating propagation. Previously, at the end of 2010, we designed an ontology called OCSO [42]. This ontology aims at describing in a structured format social objects (content augmented by social functionalities independently of its nature: video, place, text, etc.) and corresponding social activity. Then, the need of powerful and semantic sensitive algorithm to process such data led us to follow the track of semantic spreading activation.

Two posters were published at IC [42] and Web Science [43] presenting OCSO model and research axis about semantic spreading activation. A state of the art about exploitation of semantics in spreading activation and its position in the general context of this algorithm family was written. Then a formal proposition was made and algorithm development started leading to first experimental results. The state of the art, the formal proposition and early results were published at the Social Objects workshop [41]. The end of the year was mainly focused about results visualization through Gephi and knowledge acquisition on algorithm and its behavior through multiple tests.
EXMO Project-Team

6. New Results

In the continuation of our previous work, in 2011 we developed our work on evaluation of ontology matching and especially in running new experiments and generating new tests (§ 6.1.1). We also continued our work on trust in semantic peer-to-peer systems (§ 6.2.2), the use of the $\mu$-calculus for evaluating RDF path queries (§ 6.2.1) and ontology matching for linking data (§ 6.1.2).

6.1. Ontology matching and alignment

We pursue our work on ontology matching and alignment support with contributions to evaluation, data interlinking and multilingual matching.

6.1.1. Evaluation

Participants: Cássia Trojahn dos Santos [Contact], Jérôme Euzenat, Jérôme David.

Evaluation of ontology matching algorithms requires to confront them with test ontologies and to compare the results. Since 2004, we run the Ontology Alignment Evaluation Initiative (OAEI) which organises evaluation campaigns for assessing the degree of achievement of actual ontology matching algorithms [4]. This year, the evaluation campaign had 16 different teams entered the evaluation which consisted of 5 different sets of tests. The participating systems and evaluation results were presented in the 6th Ontology Matching workshop, that was held in Bonn, DE [17][9].

The main activities carried out in 2011 were related to the automation and execution of the OAEI 2011 campaign, in the framework of the SEALS project (see § 8.2.1). This involved the following main tasks:

- describe evaluation processes within the early version of the SEALS platform [11];
- develop a client allowing participants to validate their wrapped tools and evaluate (offline and locally) their tools;
- develop a test generator for automatic generation of systematic benchmarks [12];
- providing participants with a better way to bundle their tools so that they can be evaluated within the SEALS platform; and
- analysis and report of the evaluation campaign results [9].

This work has been used in the OAEI 2011 evaluation campaign. More information on OAEI can be found at http://oaei.ontologymatching.org/.

6.1.2. Ontology matching for linked data

Participants: Zhengjie Fan, Jérôme Euzenat [Contact], Jérôme David.

The web of data consists of using semantic web technologies to publish data on the web in such a way that they can be interpreted and connected together. It is thus critical to be able to establish links between these data, both for the web of data and for the semantic web that it contributes to feed.

In the context of the Datalift project (see § 8.1.1), we are developing a data interlinking module. Based on our analysis of the relationships between ontology matching and data interlinking [13], our goal is to generate data interlinking scripts on from ontology alignments. For that purpose, we have integrated existing technologies within the Datalift platform: the Alignment API, for taking advantage of the EDOAL language and Silk, developed by Frei Universität Berlin, for processing linking scripts. So far we have demonstrated the ability to process simple scripts.

This work is part of the PhD of Zhengjie Fan, co-supervised with François Scharffe (LIRMM), within the Datalift project.
6.1.3. Multilingual ontology matching

Participants: Cássia Trojahn dos Santos [Contact], Jérôme David, Jérôme Euzenat, Giuseppe Pirrò.

We have participated in the creation of a benchmark for multilingual ontology matching, the MultiFarm dataset. This dataset is composed of a set of ontologies translated in different languages and the corresponding alignments between these ontologies. It is based on the OntoFarm dataset, which has been used successfully for several years in the Ontology Alignment Evaluation Initiative. By translating the ontologies of the OntoFarm set into eight different languages – Chinese, Czech, Dutch, French, German, Portuguese, Russian, and Spanish – we created a comprehensive set of realistic test cases. We plan to include this new dataset in the OAEI 2012 campaign.

Finally, in the context of the Cameleon project (see § 8.3.1) we have been working on the creation of a multilingual comparable corpora using as seed a set of multilingual aligned ontologies. These resources will be exploited in the process of populating and enriching ontologies as well as in the process of cross-lingual ontology alignment.

6.2. Ontology networks

Dealing with the semantic web, we are interested in ontology networks, i.e., sets of distributed ontologies that have to work together. One way for these systems to interact consists of exchanging queries and answers. For that reason, we pay particular attention to query systems.

6.2.1. Path queries and $\mu$-calculus

Participants: Melisachew Wudage Chekol [Contact], Jérôme Euzenat, Pierre Genevès, Nabil Layaida.

Querying the semantic web is mainly done through SPARQL [15]. One of its extensions, PSPARQL (Path SPARQL) provides queries with paths of arbitrary length. We study the static analysis of queries written in this language with techniques based on $\mu$-calculus interpretation that have been used for XPATH. We have more specifically considered PSPARQL query containment: determining whether, for any graph, the answers to a query are contained in those of another query [18][14]. To that extent, we proposed an encoding of RDF graphs as transition systems and PSPARQL queries as $\mu$-calculus formulas. We then reduce the containment problem to testing satisfiability in the logic.

This work is part of the PhD of Melisachew Wudage Chekol, co-supervised with Nabil Layaida (WAM).

6.2.2. Trust in peer-to-peer semantic systems

Participants: Manuel Atencia [Contact], Jérôme Euzenat, Marie-Christine Rousset.

In a semantic peer-to-peer network, peers use separate ontologies and rely on alignments between their ontologies for translating queries. Nonetheless, alignments may be incorrect –unsound or incomplete– and generate flawed translations, thus leading to unsatisfactory answers. We have put forward a trust mechanism that can assist peers to select those peers in the network that are better suited to answer their queries [8]. The trust that a peer has towards another peer depends on a specific query and represents the probability that the latter peer will provide a satisfactory answer. In order to compute trust, we exploit both alignments and peers’ direct experience, and perform Bayesian inference. We have implemented our technique and conducted an evaluation. Experimental results showed that trust values converge as more queries are sent and answers received. Furthermore, the use of trust is shown to improve both precision and recall of query answers.

This work has been developed in collaboration with Marie-Christine Rousset (LIG) in the context of the DataRing project (see § 8.1.2).
6. New Results

6.1. Note

Note that we do not include here the results from Souhila Kaci and Tjitze Rienstra since they joined the team in September 2011.

6.2. Ontological Query Answering with Rules

Participants: Jean-François Baget, Marie-Laure Muguier, Michaël Thomazo, Michel Leclère, Eric Salvat, Mélanie König.

In collaboration with: Sebastian Rudolph (Karlsruhe Institute of Technology)

We have developed a framework based on rules that have the ability of generating unknown individuals, an ability sometimes called value invention in databases. These rules are of the form body \( \rightarrow \) head, where the body and the head are conjunctions of atoms (without function symbols except constants) and variables that occur only in the head are existentially quantified, hence their name existential rules hereafter. E.g., \( \forall x (\text{Human}(x) \rightarrow \exists y (\text{isParent}(y, x) \land \text{Human}(y))) \). These rules can be seen as the logical translation of conceptual graph rules, historically a main focus of the team [70] [55]. Existential rules have the same logical form as the well-known Tuple-Generating Dependencies (TGDs) in databases [45]. TGDs have been extensively used as a high-level generalization of different kinds of constraints, e.g., for data exchange [57]. Recently, there has been renewed interest for TGDs seen as rules in the context of ontological query answering. Indeed, the value invention feature has been recognized as crucial in an open-world perspective, where it cannot be assumed that all individuals are known in advance. The deductive database language Datalog allows to express some ontological knowledge but it does not allow for value invention. This motivated the recent extension of Datalog to TGDs (i.e., existential rules), which gave rise to the Datalog +/- family [52], [53], [54]. In KRR and in the Semantic Web, ontological knowledge is often represented with formalisms based on description logics (DLs). However, DLs traditionally focused on reasoning tasks about the ontology itself (the so-called TBox), for instance classifying concepts; querying tasks were restricted to ground atom entailment. Conjunctive query answering with classical DLs has appeared to be extremely complex, hence less expressive DLs more adapted to conjunctive query answering on large amounts of data have been designed recently, namely DL-Lite [51], \( \mathcal{E} \mathcal{L} \) [41], [63], and more generally Horn DLs (see e.g., [60]), cf. also the tractable profiles of the Semantic Web language OWL2. Existential rules cover the core of lightweight DLs dedicated to query answering, while being more powerful and flexible [53], [44],[21]. In particular, they have unrestricted predicate arity (while DLs consider unary and binary predicates only), which allows for a natural coupling with database schemas, in which relations may have any arity; moreover, adding pieces of information, for instance to take contextual knowledge into account, is made easy by the unrestricted predicate arity, since these pieces can be added as new predicate arguments.

Building on our previous work on conceptual graphs, while meeting this new trend, we have developed a knowledge representation framework centered on existential rules, which can be seen both as logic-based and graph-based.

Entailment, hence query answering, with existential rules is not decidable, thus finding decidable classes of rules as expressive as possible is a crucial issue. We have pursued our previous work on better understanding the border between decidability and undecidability. We have also extended rule dependency to k-dependency, which takes into account sequences of rule applications.

\[ \text{Results published in Artificial Intelligence Journal} \ [13] \ (\text{extending the work in} \ [3], \ [44]); \ \text{keynote talk synthesizing this work atRR'2011} \ [20]; \ \text{extension to k-dependency atRR'2011} \ [22] \]
For newly exhibited decidable classes (namely, “frontier-one”, “frontier-guarded” and “weakly-frontier-guarded” rules), the problem complexity was unknown, moreover there was no algorithm for computing entailment. First, we have classified these classes with respect to combined complexity (i.e., usual complexity) with both unbounded and bounded predicate arity, and data complexity (i.e., restricting the input of the decision problem to the facts). An interesting result is that some of the new classes (namely frontier-one and frontier-guarded rules) have a polynomial time data complexity. Secondly, we have provided a generic algorithm for query entailment with a large class of rules including these classes, which is worst-case optimal for combined complexity (with or without bounded predicate arity) as well as for data complexity.

Results partially published at IJCAI’2011 [21]. Long paper in preparation with extended complexity results and all proofs, for submission to a major artificial intelligence journal.

6.3. Processing Conjunctive Queries with Negation

Participants: Marie-Laure Mugnier, Michel Leclère, Khalil Ben Mohamed, Michaël Thomazo.

 Conjunctive queries have long been recognized as the basic queries in database and knowledge-based systems. The fundamental decision problems on these queries, namely query inclusion checking (given two queries $q_1$ and $q_2$, is $q_1$ included in $q_2$, i.e., is the set of answers to $q_1$ included in the set of answers to $q_2$ for all databases) and query entailment (is a given query entailed by the database) are NP-complete. When atomic negation is added to queries and databases, these problems become $\Pi^p_2$-complete (with the open world assumption for the query entailment problem). Note that these problems can be recast as entailment in the FOL fragment of existentially closed conjunction of literals (without function symbols except constants). On the one hand, we have led a theoretical complexity study: we have investigated the role of pairs of literals called “exchangeable” (which generalizes the notion of unifiable literals) in the complexity increase. The main results are that when the number of exchangeable pairs is bounded, say by $k$, then the complexity falls from $\Pi^p_2$-complete to $P^{NP}$-complete for any $k \geq 3$, and is NP-complete for $k \leq 1$ (the case $k = 2$ being open).

In collaboration with: Geneviève Simonet (LIRMM Algeco team)


On the other hand, we have proposed, refined and compared experimentally several algorithms. This study follows first results of us in [61] and is the core of Khalil Ben Mohamed’s PhD thesis defended in December 2010 [64].

- Results published DEXA 2011 [24] (extending our work in RFIA 2010 [48], DEXA 2010 [46], AIMSA 2010 [47]).

Let us point out that both theoretical and practical results still hold when the predicates are preordered, which allows to take very light ontologies into account, i.e., where concepts and relations are organized in a specialization preorder.

6.4. Argumentation Systems for Decision Making

Participants: Rallou Thomopoulou, Madalina Croitoru, Jérôme Fortin, Marie-Laure Mugnier.

In collaboration with: Joël Abecassis (IATE/INRA), Jean-Rémi Bourguet (UM3), Patrice Buche (IATE/INRA), Sébastien Destercke (IATE/CIRAD) Nir Oren (Univ. of Aberdeen, Scotland)

Scientific investigations in this axis are guided by applications of our partners in agronomy (IATE laboratory). Substantial part of the work has consisted of analyzing the proposed applications and the techniques they require in order to select appropriate applications with respect to our team project.
Argumentation is a reasoning model based on the construction and the evaluation of arguments. In his seminal paper, Dung has proposed an abstract argumentation framework [56]. In that framework, arguments are assumed to have the same strength. This assumption is unfortunately strong and often unsatisfied. Consequently, several generalizations of the framework have been proposed in the literature. In [49] and [50], we have led a comparative study of these generalizations. It clearly shows under which conditions two proposals are equivalent. We have also integrated those generalizations into a common more expressive framework.

An instantiation of Dung’s abstract framework with the conceptual graph framework has been proposed. This representation uses default conceptual graph rules, an extension of classical conceptual graph rules (equivalent to existential rules, see Axis 1) with Reiter’s defaults [67] allowing for non-monotonic reasoning, that we developed independently of the argumentation framework [42], [43]. In the conceptual graph representation, arguments are represented as nested graphs, attacks between arguments can be computed from the structure of arguments and default rules allow to compute several kinds of extensions (i.e., maximal sets of arguments jointly acceptable according to a given semantics).

This approach has been applied to agrifood chain analysis, which is a highly complex procedure since it relies on numerous criteria of various types: environmental, economical, functional, sanitary, etc. Quality objectives imply different stakeholders, technicians, managers, professional organizations, end-users, public organizations, etc. Since the goals of the implied stakeholders may be divergent, decision-making raises arbitration issues. Arbitration can be done through a compromise—a solution that satisfies, at least partially, all the actors—or favor some of the actors, depending on the decision-maker’s priorities. We have analyzed a case study concerning risks/benefits within the wheat-to-bread chain. It concerns the controversy about the possible change in the ash content of the flour used for commonly used French bread. Several stakeholders of the chain are concerned, in particular the Ministry of Health through its recommendations in a national nutrition and health program, millers, bakers and consumers.

As already pointed out, the proposed approach is novel both for theoretical and application aspects.

▷ Results presented in [30], [28].

Let us mention additional results related to the applications in agronomy on decision making combining machine learning based on decision trees and ontologies [58], [30], as well as results obtained by our collaborators on semi-automatic data extraction from web data (tables), data reliability, and the representation and flexible querying of imprecise data with fuzzy sets [16], [14], [17], [26], [31], [25], [27], [33], [34]. These investigations are complementary to the above mentioned results on argumentation and generally relate to other aspects in the same applicative projects.

6.5. Semantic Data Integration

Participants: Michel Leclère, Michel Chein, Madalina Croitoru, Rallou Thomopoulos, Léa Guizol.

It often happens that different references (i.e. data descriptions), possibly coming from heterogeneous data sources, concern the same real world entity. In such cases, it is necessary: (i) to detect whether different data descriptions really refer to the same real world entity and (ii) to fuse them into a unique representation. Since the seminal paper [66], this issue has been been studied under various names: “record linking”, “entity resolution”, “reference resolution”, “de-duplication”, “object identification”, “data reconciliation”, etc., mostly in databases (cf. the bibliography by William E. Winckler 1). It has become one of the major challenges in the Web of Data, where the objective is to link data published on the web and to process them as a single distributed database. Most entity resolution methods are based on classification techniques; Fatiha Saïs, Nathalie Pernelle and Marie-Christine Rousset proposed the first logical approach [68]. Many experiments on public data are underway, in France (cf. DataLift 2 and ISIDORE 3 projects) or internationally (e.g., VIAF project 4 led by OCLC 5, whose aim is to interconnect authority files coming from 18 national organizations).

2DataLift, http://datalift.org/
3ISIDORE, http://www.rechercheisidore.fr/
4The Virtual International Authority File, http://www.oclc.org/research/activities/viaf/
Two years ago, we began a collaboration with ABES (National Bibliographic Agency for Universities, which takes part in the VIAF project). The aim of this collaboration is to enable the publication of ABES metadata bases on the Web of Data and to provide an identification service dedicated to bibliographic notices. ABES bibliographic bases, and more generally document metadata bases, appear to be a privileged application domain for the representation and reasoning formalisms developed by the team. This work has an interdisciplinary dimension, as it also requires experts in the Library and Information Science domain. We think that a logical approach is able to provide a generic solution for entity resolution in document metadata bases, even though it is generally admitted in Library and Information Science that “there is no single paradigmatic author name disambiguation task—each bibliographic database, each digital library, and each collection of publications, has its own unique set of problems and issues” [69].

6.5.1. SUDOC Metadata Formalization

The first step of collaboration with ABES was to formalize the SUDOC catalogue, which contains all French academic libraries bibliographic notices, into a knowledge base using a suitable knowledge representation and reasoning language. This required to first analyze SUDOC content, as well as document description standards (CRM-CIDOC, FRBR, Dublin Core). We then designed an ontology expressed in the Semantic Web languages RDFS + OWL, compatible with document description standards, as well as translations from any SUDOC set of notices into a set of RDF facts according to this ontology. These translations have been implemented, which allows to export SUDOC bases into Semantic Web formats. Moreover, using the RDFS to CG second translation mentioned above, we are now able to import SUDOC bases into our tools CoGUI + CoGITaNT.

▶ Technical report [40].

6.5.2. Implementation of an Entity Identification Service

In order to perform entity resolution (for entities restricted to "authors" for now), we have defined a set of rules allowing to enrich Sudoc descriptions; then, using enriched descriptions, authors can be classified according to a proximity criterion. A prototype providing this service has been implemented on top of Cogui. Experiments are currently led in the context of the SudocAd project jointly conducted by ABES and GraphIK. SudocAd aims at enriching the author field of a bibliographic record describing a document with links to Sudoc authorities referring to the authors of the target document. A general description of the implemented approach, an analysis of this approach on a representative sample of bibliographic records and first results on 13400 bibliographic records extracted from a corpus independent from Sudoc catalog are presented in the final report of SudocAd.

▶ Link to SudocAd Final Report: http://www.abes.fr/Media/Fichiers/Sudoc-Fichiers/SudocAD_rapportFinal

Finally, we have defined an extension of our own logical framework (existential rules, constraints, homomorphism-based mechanisms) based on Hector J. Levesque and Gerhard Lakemeyer’s Standard Names [62], and the notion of knowledge base faithfulness with respect to the entity resolution problem (intuitively, the fact that the knowledge base is non-ambiguous). This is still ongoing work.

▶ Research Report [38].

5 Online Computer Library Center, http://www.oclc.org
6. New Results

6.1. Efficient XML and RDF data management

Participants: Mohamed Amine Baazizi, Nicole Bidoit, Dario Colazzo, François Goasdoué, Konstantinos Karanasos, Asterios Katsifodimos, Julien Leblay, Noor Malla, Ioana Manolescu, Alexandra Roatis, Marina Sahakyan, Federico Ulliana.

6.1.1. Materialized views for XML

We have continued our work on optimizing XML queries through materialized view-based rewriting, implemented within the ViP2P system. We published in IEEE ICDE 2011 an algorithm for rewriting XQuery queries using materialized XQuery view, which improves the state of the art in terms of expressive power of the supported XQuery subset, in collaboration with V. Vassalos (AUEB, Greece) [41]. Two follow-up works concern: efficient algebraic algorithms for incrementally maintaining the materialized views when the underlying documents change, in collaboration with A. Bonifati (CNR, Italy) [35], and algorithms for automatically recommending views to materialize for a given XML query workload, with V. Vassalos.

6.1.2. Type-based Update Optimization for XML

XML projection is a well-known optimization technique for reducing memory consumption for XQuery in-memory engines in order to overcome the main-memory limitations of these systems (Galax, Saxon, QizX, and eXist). One of our main research line focuses on a schema-based projection technique for for update optimization. The update language considered is XQuery Update Facility (XUF). The main idea behind this technique is: given a query \( q \) over an XML document \( t \), instead of evaluating \( q \) over \( t \), the query \( q \) is evaluated on a smaller document \( t' \) obtained from \( t \) by pruning out, at loading time, parts of \( t \) that are irrelevant for \( q \). The queried document \( t' \), a projection of the original one, is often much smaller than \( t \) due to selectivity of queries.

The scenario and type-based projection proposed for XML queries, cannot be applied directly for updates. We have proposed a new scenario which is composed of four steps:

1. from the update \( U \) and the DTD \( D \), a type projector \( P \) is inferred;
2. the document \( t \), valid wrt \( D \), is projected following \( P \) in a streaming manner, at loading time;
3. \( U \) is evaluated over the projection \( P(t) \) and produces a partial result \( U(P(t)) \);
4. the initial document \( t \) is merged with \( U(P(t)) \), in a streaming manner, at writing (serializing) time in order to produce the final result \( U(t) \).

The scenario has been first studied and implemented for a kind of type projector which is a good compromise between simplicity and effectiveness, and corresponding results have been published in [31]. Subsequently, we have improved this technique by designing: (i) a new kind of type projector that minimizes the amount of data kept in the projection, and (ii) a new merge algorithm using the improved type projector. This analysis is complicated by the strong interconnection between the two tasks: while minimizing the projection we need to ensure a fast and correct merge process [17].

These results have also been presented in [34], providing an overview on the use of types and constraints from relational to XML data, and in the tutorial [63] focusing on schema-based techniques for safe and efficient XML processing.
6.1.3. XML query-update independence

A query and an update are independent when the query result is not affected by update execution, on any possible input database. Detecting query-update independence is of crucial importance in many contexts: view maintenance, concurrency, access control policies etc. Benefits are amplified when query-update independence can be checked statically. We propose a novel schema-based approach for detecting XML query-update independence. Differently from traditional schema-based analysis for XQuery, our system infers sequence of labels, called *chains*, that are vertically navigated in each schema instance by query and update paths. More precisely, for each node that can be selected by a query/update path in a schema instance, the system infers a chain recording: a) all labels that are encountered from the root to the selected node, and b) the order in which these labels are traversed. The contextual and ordering information provided these chains is at the basis of an extremely precise static independence analysis. We have devised a quite precise chain-inference system, and devised techniques for efficient implementation of the chain-based independence analysis. Results and experiments concerning this line of research have been recently submitted to an international conference.

6.1.4. Precision and complexity of XQuery type inference

A key feature of XQuery is its type system. Any language expression is statically typed and its type is used during program type-checking. In XQuery, types of input data and functions are defined in terms of regular expression types, but it is quite easy to write queries that generate non-regular languages. As a consequence, any type system for XQuery has to rely on a *type inference* process that approximates the (possibly non-regular) output type of a query with a regular type. This approximation process, while mandatory and unavoidable, may significantly decrease the precision of the inferred types. In [37] we study and compare in terms of precision and computational complexity two main existing XQuery type systems.

6.1.5. Managing temporal XML documents

The management of temporal data is a crucial issue in many database applications. We are currently investigating efficient storage and update methods for temporal XML documents, with a focus on compactness of the representation. One of the method developed relies on the type based optimization method developed for updates [31]. First results about this line of research are included in recent publications [29], [30]. Next research activities focus on the design of expressive temporal query and update languages, and on the use of techniques proposed in [29] for query and update optimization.

6.1.6. Materialized view selection for RDF

Syntactically, RDF, the data format of the Semantic Web, resembles relational data. However, RDF query processing is significantly complicated by the irregular nature of RDF data and by its simplistic data model, which leads to syntactically complex queries (involving many joins over the whole triple data set). When a query workload is known, the performance of the workload can be significantly improved by materializing access support data structures such as materialized views. Our efficient algorithms for selecting a set of views to materialize in order to speed up the processing of a set of RDF queries are described in a recent publication [21]. The prototype implementing them has been demonstrated at [53].

6.1.7. Hybrid models for XML and RDF

We have obtained interesting results in the area of jointly managing XML and RDF data. A first direction of work in this area was to support *annotated documents*, that is, XML documents where individual nodes or fragments could be annotated with RDF triples. This model allows to capture, for instance, blog comments, user ratings on social sites etc. We have proposed a general model based on W3C standards for modeling such data [39].

6.1.8. RDF query answering

The current trend for efficiently querying RDF datasets consists of delegating query evaluation to a scalable RDBMS. However, RDF query answering requires in addition to handle – outside the RDBMS – the RDF semantics. We have introduced the *database (DB) fragment of RDF*, encompassing the popular Description
Logic (DL) one with essential RDF features like modeling incomplete information, for which we have devised novel *saturation- and reformulation-based techniques* for answering the *Basic Graph Pattern* (BGP) queries of SPARQL. This extends the state of the art on pushing RDF query processing within RDBMSs.

### 6.2. Models for Web data management

**Participants:** Serge Abiteboul, Emilien Antoine, Meghyn Bienvenu, Alban Galland.

A book on Web Data Management and Distribution [54] was published this year.

#### 6.2.1. A rule-based language for Web data management

We recently proposed [26] a Datalog-style rule-based language (called *Webdamlog*) for web data management. A novel feature of our language is *delegation*, that is, the possibility of installing a rule at another peer. In its simplest form, delegation is essentially a remote materialized view. In its general form, it allows peers to exchange rules, i.e., knowledge beyond simple facts, and thereby provides the means for a peer to delegate work to other peers.

A key contribution of our work is a study of the impact on expressiveness of delegations and explicit timestamps. We showed that both strictly augment the power of the language. In order to validate the semantics of our model, we demonstrated that under certain natural conditions, our semantics converges to the same semantics as the centralized system with the same rules.

#### 6.2.2. Web information management with access control

We investigated the problem of sharing private information on the Web, where the information is hosted on different machines that may use different access control and distribution schemes. Based upon our work on Webdamlog, we introduced a distributed knowledge-base model, termed WebdamExchange, that comprises logical statements for specifying data, access control, distribution and knowledge about other peers. In a demo at ICDE [28], we showed how the model can be used in a social-network context to help users keep control on their data on the web. In particular, we demonstrated how users within very different schemes of data distribution (centralized, DHT, unstructured P2P, etc.) can still transparently collaborate while keeping a good control over their own data.

### 6.3. Ontology-based data and document Management

**Participants:** Meghyn Bienvenu, François Goasdoué, Yassine Mrabet, Nathalie Pernelle, Gianluca Quercini, Chantal Reynaud, Brigitte Safar, Fabian Suchanek.

#### 6.3.1. Semantic Annotation

We have started a work on semantic annotation of public administration data in the setting of the project DataBridges, an ICT Labs activity. We considered public data represented in tables. The tables that we studied were tables created and published by INSEE. They are spreadsheets filled with statistics about geographic locations and are usually composed of multiple columns, of which one, that we term the subject column, contains a list of textual references to geographic entities, or toponyms, while the others contain numeric attributes. We proposed an approach and an algorithm that assigns a type, or header to the subject column of a INSEE table and identifies the geographic entities referred to by the toponyms in the column [64]. An external resource, DBpedia, is used to help to disambiguate the entities mentioned in the tables and a domain ontology ensures that the types are relative to the geographic domain. This work is continued in the setting of a post-doctoral work granted by the ANR project DataBridges. The aim of the project being to enrich a data warehouse, a first work is to automatically build an initial RDF data warehouse from data collected from the web.

#### 6.3.2. Adaptive Ontologies for Information Retrieval

We published the approach supported by the TARGET framework for Web Information Retrieval in the International Journal of Web Portals (IJWP) [22]. This approach was the core of the PhD of Cédric Pruski defended in April 2009.
6.3.3. Querying ontology-based annotations

We have pursued our work on integrating knowledge bases and semantic annotations made on more or less structured tagged documents. We have defined an approach where RDF named graphs are used to distinguish uncertain semantic annotations from rdf triples that are provided by the populated ontology. A user domain query is then reformulated to obtain answers that are ranked according to their provenance (knowledge bases or annotations) [61].

6.3.4. Watermarking for ontologies

Ontologies are usually available under some type of license. The large ontologies of the Semantic Web, e.g., are commonly licensed under a Creative Commons License or a GNU license. These licenses require giving credit to the authors of the ontology if the ontology is ever used somewhere else. However, it can be hard to prove whether an ontology is used somewhere else, because ontologies contain world knowledge. If someone "steals" an ontology and uses it somewhere else, he can always claim that he collected the data by himself from real-world sources. To tackle this problem, we have studied approaches that watermark an ontology [43]. If a watermarked ontology is used somewhere else, the mark proves that the ontology has been stolen. Existing approaches have mainly modified the facts in the ontology to create a mark. This, however, compromises the precision of the ontology. Therefore, we have developed an approach that does not modify, but remove certain facts. Thereby, the precision of the ontology is not affected. We show that only a handful of facts have to be removed from an ontology to protect it against theft.

6.3.5. Consistent query answering in DL-Lite

An important problem which arises in ontology-based data access is how to handle inconsistencies. In the database community, the related problem of querying databases which violate integrity constraints has been extensively studied under the name of consistent query answering. The standard approach is based on the notion of a repair, which is a database which satisfies the integrity constraints and is as similar as possible to the original database. Consistent answers are defined as those answers which hold in all repairs. A similar strategy can be used for description logics by replacing the integrity constraints with the ontology. Unfortunately, recent work on consistent query answering in description logics has shown this problem to be co-NP-hard in data complexity, even for instance queries and the simplest DL-Lite dialect. In light of this negative result, we considered the problem of identifying cases where consistent query answering is feasible, and in particular, can be done using query rewriting, with the aim of better understanding the cases in which query rewriting can be profitably used. In [51], we make some first steps towards this goal by formulating general conditions which can be used to prove that a consistent rewriting does or does not exist for a given DL-Lite TBox and instance query.

6.3.6. Module-based data management in DL-lite

The current trend for building an ontology-based data management system (DMS) is to capitalize on efforts made to design a preexisting well-established DMS (a reference system). The method amounts to extract from the reference DMS a piece of schema relevant to the new application needs – a module –, possibly to personalize it with extra-constraints w.r.t. the application under construction, and then to manage a dataset using the resulting schema. We have revisited the reuse of a reference ontology-based DMS in order to build a new DMS with specific needs. We go one step further by not only considering the design of a module-based DMS (i.e., how to extract a module from a ontological schema): we also study how a module-based DMS can benefit from the reference DMS to enhance its own data management skills. We consider the setting of the DL-LiteA dialect of DL-Lite, which encompasses the foundations of the QL profile of OWL2 (i.e., DL-LiteR): the W3C recommendation for managing efficiently large datasets. We introduce and study novel properties of robustness for modules that provide means for checking easily that a robust module-based DMS evolves safely w.r.t. both the schema and the data of the reference DMS. From a module robust to consistency checking, for any data update in a corresponding module-based DMS, we show how to query the reference DMS for checking whether the local update does not bring any inconsistency with the data and the constraints of the reference DMS. From a module robust to query answering, for any query asked to a module-based
DMS, we show how to query the reference DMS for obtaining additional answers by also exploiting the data stored in the reference DMS.

6.4. Data and Knowledge Integration

Participants: Julio Cesar Dos Reis, Fayçal Hamdi, Rania Khefifi, Yassine Mrabet, Nathalie Pernelle, Chantal Reynaud, Fatiha Saïs, Brigitte Safar, Fabian Suchanek, Danai Symeonidou.

6.4.1. Reference Reconciliation

The reference reconciliation problem consists in deciding whether different data descriptions refer to the same real world entity (same person, same conference etc.) Some of existing approaches, such as LN2R, are declarative and knowledge-based. Different kinds of knowledge can be declared in a domain ontology, like disjointness between classes or key constraints. This knowledge can be exploited to infer reconciliation and non-reconciliation decisions.

Our reference reconciliation work pursues three directions:

- develop an automatic approach of key constraint discovery. We have proposed in [46] KD2R, a method which allows automatic discovery of key constraints associated to OWL2 classes. These keys are discovered from RDF data which can be incomplete. The proposed algorithm allows this discovery without having to scan all the data. KD2R has been tested on data sets of the international contest OAEI and obtains promising results.

- develop a reference reconciliation method for detecting redundant data in case of web data tables that are semantically annotated by an ontology. Each table cell values consists in numerical fuzzy set (NFS) or in symbolic fuzzy set (SFS). We have developed a method which uses ontology knowledge and computes similarity scores to decide the data redundancy. We have also proposed two similarity measures for numerical fuzzy set as well as symbolic fuzzy set. The proposed measures are more flexible than existing ones. This approach has been published in [36], [58]. We are working on its extension to be able to distinguish redundant data from similar ones by using provenance information.

- develop a new approach which addresses the problem of resource discovery in the Linked Open Data cloud (LOD) where data described by different schemas is not always linked. We have proposed an automatic approach in [42], [58] that allows discovery of new links between data. These links can help to match schemas that are conceptually relevant with respect to a given application domain. Furthermore, these links can be exploited during the querying process in order to combine data coming from different sources. In this approach we exploit the semantic knowledge declared in different schemas in order to model: (i) the influences between concept similarities, (ii) the influences between data similarities, and (iii) the influences between data and concept similarities. The similarity scores are computed by an iterative resolution of two non linear equation systems that express the concept similarity computation and the data similarity computation.

6.4.2. Context-aware Personal Information Management

Personal information management (PIM) is the practice and analysis of the activities performed by people to acquire, organize, maintain, and retrieve information for everyday use. PIM is a growing area of interest because everyone is looking for better use of our limited personal resources of time, money and energy. Several research on the topic is being done in different disciplines, including human-computer interaction, database management, information retrieval and artificial intelligence.

The increasingly big amount of personal information (e.g., mails, contacts, appointments) managed by a user is characterized by their heterogeneity, their dispersion and their redundancy. The general goal of this work consists in designing a system, which allows providing the end-users personal data access with services that are relevant to his/her needs, and to access personal data both by mobile devices (smartphone) and Internet-connected Personal Computers. More specifically, we focus here on the problem of defining a common
meta-model for a flexible and homogeneous personal information management. The meta-model that we propose allows users creating personal information and organizing them according to different points of view (ontologies) and different contexts. Contextual queries are defined to allow users to retrieve its personal information using the geographical contexts. The semantic Web languages (OWL, RDF and SPARQL) are used to implement the approach.

6.4.3. Mapping between ontologies

We pursue our work on ontology alignment in the setting of the ANR GeOnto project by aiming to provide full life-cycle support for ontologies.

We investigated how alignment results generated by our alignment tool, TaxoMap, can be used to enrich one ontology with another. We shown that the enrichment process depends on characteristics of the ontology used for enrichment. Three enrichment contexts identified in the setting of the ANR project GeOnto have been studied and enrichment treatments performed. A first context considers ontologies of the same application domain and of a reasonable size. A second context considers small ontologies previously extracted from a generalist one. A third context considers enrichment from a huge, generalist ontology, such as Yago. Early results obtained in the setting of the ANR project GeOnto in the topographic domain have been published in [50], [25].

The module supporting our enrichment approach has been implemented in TaxoMap Framework using patterns. Initially, TaxoMap Framework was composed of our alignment tool, TaxoMap, we are working on for several years in the team and of a mapping refinement module. We extended it in order to obtain a broader framework and an interactive environment by including TaxoPart, a partitioning tool we developed to split two huge ontologies which could not be aligned into two sets of blocks of a limited size, and a module specific to ontology enrichment. Moreover, we re-implemented TaxoMap, our alignment tool, as a web service to make it easily accessible at: http://taxomap.lri.fr:8000/axis2/services/TaxoMapService?wsdl.

We also started a PhD work, joined with CRP Henri Tudor in Luxembourg, to investigate issues dealing with medical knowledge organizing systems evolution. We will define a formal framework to support medical knowledge organizing systems evolution in a consistent way and also to support the maintenance of mappings directly impacted by knowledge organizing systems local evolution.

On a related topic, we have developed a probabilistic framework, PARIS (Probabilistic Alignment of Relations, Instances and Schema), for matching ontologies holistically, thereby exploiting synergies between matches on the instance level and matches on the schema level [57]. The framework is parameter-free and does not require resource-specific tuning. PARIS is fully implemented and has been shown to match some of the largest ontologies on the Semantic Web with a precision of around 90%.

6.4.4. Integration of Web resources

We have pursued our work on integration of resources available on the Web in Adaptive Hypermedia Systems (AHS), allowing creators to define their own adaptation strategies based on their own domain models. The approach is based on a set of 22 adaptation patterns, independent of any application domain and independent of any adaptation engine, published in [59], [47]. These elementary adaptation patterns are organized in a typology in order to facilitate their understanding and their use in the EAP framework to define complex strategies. In [24], we described the whole process to generate complex adaptation strategies and how the generated strategies can be integrated into existing AHSs. The results of an experiment conducted in the e-learning domain is presented. It showed that the pattern-based approach for defining adaptation strategies is more suitable than those based on “traditional” AH languages.

We also pursued our work on the integration of the EAP framework and other AHSs. Our collaboration with A. Cristea from the University of Warwick (UK) led us to a very detailed study of adaptation languages. The first flexible generic adaptation language is the LAG adaptation language. We studied the expressivity of this initial adaptation language in comparison with our newly proposed language, in the EAP framework, and the pros and cons of various decisions in terms of the ideal way of defining an adaptation language. We proposed a unified vision of adaptation and adaptation language. The unified vision is not limited to the
two languages analyzed, and can be used to compare and extend other approaches in the future. Beside this theoretical qualitative study, we also made experimental evaluation and comparison of the two languages, and an article is currently being evaluated.

We have also investigated integration of Web services. The Search Computing project (“SeCo”) at the Polytechnic University of Milan aims to orchestrate Web services to answer user queries. Currently, the project represents Web services by so-called Service Marts. These are frame-like representations of the services, which follow the slot-value paradigm. This representation faces several challenges if more Web services get added to the system, because it is hard to ensure that Web services added by different users can still be joined. Therefore, we have explored a more ontological representation of Web services. In our proposal [55], Web services are represented as sub-graphs of an ontology. This allows users to add new Web services that re-use the vocabulary of existing Web services.

On related topics, together with researchers from the Max-Planck Institute in Saarbrucken, we have worked on extending the YAGO ontology. YAGO already contains dozens of millions of facts. With the present work, we aim to give these facts a temporal and a spatial dimension. For every event and every entity, we want to know where and when these objects existed. For this purpose, we have developed a methodology that extracts these types of facts from Wikipedia. We have also developed a logical reasoning framework that allows propagating these time and space annotations from some facts to others. This has grown YAGO to 80 million facts in total, making it an ontology that is anchored in time and space (Best demo award at the WWW 2011 conference [40]).

6.5. Reasoning over Distributed Systems


6.5.1. Distributed Diagnosis Problems

We pursued the work on distributed algorithms for diagnosing distributed systems. The general framework is consistency-based diagnosis for propositional-logic theories in a P2P setting with privacy constraints. It boils down to distributed implicant finding and is thus in some sense dual to the problem of consequence finding described in next paragraph. Vincent Armant is finishing is PhD and has extended his previous work on more general topics, i.e. focusing on the construction of a good decomposition of the network that will ensure an efficient reasoning mechanism. An important effort has been put in the design of a real-world sized experimentation on distributed systems.

Lina Ye defended her PhD on diagnosability analysis of distributed discrete-event systems, modeled as synchronized labeled automata. The aim of diagnosability is to ensure that a given partially observable system has the property that any fault (taken from a set of faults given a priori) will be detectable and identifiable without ambiguity in a finite time after its occurrence. Distributed diagnosability analysis is optimized by abstracting necessary and sufficient information from local objects to achieve global decision. After having addressed the distribution of the system’s model into local models, we focus in 2011 on the extension to systems where the observable information itself is distributed instead of centralized. Joint diagnosability definition has been provided and undecidability of deciding it has been proved in the general case where communication events are not observable, before proposing an algorithm to test its sufficient condition. In addition, decidability result and algorithm have been given when communications are observable.

Michel Batteux defended also his PhD (led in the framework of a CIFRE thesis with Sherpa Engineering) about diagnosability and diagnosis of technological systems. The work was led in the centralized case, focusing on defining, implementing, testing and validating on a real case study (a fuel cell system) an all-in-one tool to design a diagnosis system for technological systems by integrating representation of the system and its potential faults, off-line diagnosability analysis and automatic generation of the on-line embedded diagnoser.
6.5.2. Distributed Consequence Finding

A major reengineering of the SOMEWHERE platform, for decentralized consequence finding, has been initiated within the DISQUE project. Current efforts have focused on the rewriting of the communication layer, that now relies on the JXTA middleware. A new tool is also being developed, in order to facilitate large scale experimentations on a grid (Grid5000). This tool is designed in a fairly generic way, in order to be reusable for similar projects that require deploying sets of collaborating reasoners in a decentralized setting, and automating collaborative problem solving on various instances. We also expect this tool to be used for automating integration tests during further developments.

6.5.3. Towards distributed architectures for Modern SAT Solvers

If we aim at proposing a new architecture for distributed SAT Solvers, we pursued this year the improvements of Glucose, our centralized SAT solver. Glucose 2 won 3 medals at the SAT 2011 Competition, and one in the category Application SAT+UNSAT. We target to make a massively distributed version of Glucose, for very hard SAT problems.
5. New Results

5.1. Decision Making

5.1.1. Optimizing Automated Service Discovery

**Participant:** Jörg Hoffmann.

Michael Stollberg (SAP Research, Germany) and Dieter Fensel (University of Innsbruck, Austria) are external collaborators.

We completed earlier work, done while all authors were employed at the University of Innsbruck, and published it in the International Journal of Semantic Computing [10]. In a nutshell, the work proposes to use first-order logic for annotating web services to accomplish better precision and recall in service discovery; its core contribution is a technique making such discovery more effective – discovery here involves first-order logical reasoning – by designing a caching technique storing known relationships between available services and possible discovery queries.

5.1.2. Overview of Semantic Web Service Technologies

**Participant:** Jörg Hoffmann.

Stijn Heymans (SemanticBits, USA), Annapaola Marconi (Fondazione Bruno Kessler, Trento, Italy), Joshua Phillips (SemanticBits, USA), and Ingo Weber (University of New South Wales, Sydney, Australia) are external collaborators.

We were invited to write a book chapter about the basic AI technologies underlying semantic Web service discovery and composition. The chapter has been published as part of a book entitled “Handbook of Service Description – USDL and its Methods” in Springer-Verlag [46].

5.1.3. Analyzing Planning Domains to Predict Heuristic Function Quality

**Participant:** Jörg Hoffmann.

The heuristic search approach to planning (cf. the above) rises and falls with the quality of the heuristic estimates. The dominant method, especially in satisficing (non-optimal) planning, is to approximate a heuristic function called $h^+$ – this is used in almost every state of the art satisficing planning system. In earlier work, Jörg Hoffmann showed that $h^+$ has some amazing qualities, in many traditional planning benchmarks, in particular pertaining to the complete absence of local minima. [62] His proofs of this are hand-made, raising the question whether such proofs can be lead automatically by domain analysis techniques. The possible uses of such analysis are manifold, e.g., for automatic configuration of hybrid planners or for giving hints how to improve the domain design. The question has been open since 2002. A serious attempt of Jörg Hoffmann resulted in disappointing results – his analysis method has exponential runtime and succeeds only in two extremely simple benchmark domains. In contrast to this, in our work here we answer the question in the affirmative. We establish connections between certain easily testable syntactical structures, called “causal graphs”, and $h^+$ topology. This results in low-order polynomial time analysis methods, implemented in the Torchlight tool, cf. Section 4.2. Of the 12 domains where Hoffmann proved the absence of local minima, TorchLight gives strong success guarantees in 8 domains. Empirically, its analysis exhibits strong performance in a further 2 of these domains, plus in 4 more domains where local minima may exist but are rare. We show that, in this way, TorchLight can distinguish Hoffmann’s “easy” domains from the “hard” ones. By summarizing structural reasons for analysis failure, TorchLight also provides diagnostic output pin-pointing potentially problematic aspects of the domain. A conference paper on this work was published at ICAPS 2011 [25], and nominated for the best paper award there. A journal paper was published in the Journal of AI Research (JAIR) [9].
5.1.4. Relaxing Bisimulation for State Aggregation in the Computation of Lower Bounds

Participant: Jörg Hoffmann.

Raz Nissim (Ben-Gurion University, Beer-Sheva, Israel) and Malte Helmert (University of Freiburg, Germany) are external collaborators.

Like the previous line of work, this addresses planning as heuristic search, specifically the automatic generation of heuristic estimates. This is also the core question investigated in the BARQ project, see below. In preparation of this project, we are conducting this line of research, which explores some of the most basic ideas behind BARQ. The basic technique under consideration was developed in prior work outside INRIA. [ 61 ] The heuristic estimates are lower bounds generated from a quotient graph in which sets of states are aggregated into equivalence classes. A major difficulty in designing such classes is that there are exponentially many states. Despite this, our technique allows explicit selection of individual states to aggregate, via an incremental process interleaving it with state space re-construction steps. We have shown previously that, if the aggregation decisions are perfect, then this technique dominates the other known related techniques, and sometimes produces perfect estimates in polynomial time. But how to take these decisions? Little is known about this as yet. In the present work, we start from the notion of a “bisimulation”, which is a well-known criterion from model checking implying that the quotient system is behaviorally indistinguishable from the original system – in particular, the cost estimates based on a bisimulation are perfect. However, bisimulations are exponential even in trivial planning benchmarks. We observe that bisimulation can be relaxed without losing any information as far as the cost estimates are concerned. Namely, we can ignore the “content of the messages sent”, i.e., the state transition labels. Such relaxed bisimulations are often exponentially smaller than the original ones. We show to what extent such relaxation can be applied also within our incremental construction process. As a result, in several benchmarks we obtain perfect estimates in polynomial time, and we significantly increase the set of benchmark instances that can be solved with this approach. Indeed, the approach obtained a 2nd place in the optimal track of the 2011 International Planning Competition, and was part of the 1st-prize winning portfolio. A conference paper was published at IJCAI 2011 [ 28 ], and a journal paper is under preparation for submission to the Journal of the ACM.

5.1.5. Relaxing Bisimulation by Choosing Transition Subsets

Participants: Michael Katz, Jörg Hoffmann.

Malte Helmert (University of Freiburg, Germany) is an external collaborator.

This line of work builds on the previous one by designing new methods for relaxing bisimulations. The key idea is to apply the bisimulation property to only a subset of the transitions in the system under consideration. We showed that one can ignore large subsets of transitions without losing any information, i.e., while still guaranteeing to obtain a perfect heuristic. At the same time, such a relaxed bisimulation makes less distinctions and may thus be exponentially smaller. For practical purposes, we designed several approximate strategies relaxing more, obtaining smaller abstractions at the expense of information loss. The techniques are currently being evaluated empirically, and a paper submission is in preparation for ICAPS’12.

5.1.6. Improving $h^+$ by Taking Into Account (Some) Negative Effects

Participants: Emil Keyder, Jörg Hoffmann.

Patrik Haslum (NICTA, Australia) is an external collaborator.

Like the previous lines, this is on planning as heuristic search. As mentioned above in Section 5.1.3, approximating the $h^+$ heuristic is the dominant approach to obtain estimates in satisficing (non-optimal) planning. That notwithstanding, $h^+$ is obtained by ignoring all negative effects, which of course leads to very bad estimates in domains where these domains play a key role, for example puzzle-like domains, e.g. Rubic’s cube, where actions interfere intensively with each other. It has long (for almost 10 years) been an active research issue how to take at least some of the negative effects into account when computing $h^+$. All attempts, however, remained at rather ad-hoc methods, like, counting the number of violated binary constraints (pairs of facts that cannot be true at the same time) within the relaxed plan underlying the estimate. In the present work,
for the first time we provide a well-founded formal approach to the issue. As was suggested in prior work, [60], we design a compiled planning task which introduces constructs allowing \( h^+ \) to correctly handle a subset \( C \) of fact conjunctions. Whereas this prior work requires a compilation exponential in \(|C|\) – and thus allows only to introduce very few conjunctions – in our work we designed a compilation that is linear in \(|C|\). We proved that one can always choose \( C \) so that \( h^+ \) in the compiled task is a perfect heuristic. Of course, in general \( C \) might have to be exponentially large to achieve this. We designed practical methods selecting \( C \) in a way so that the overhead (the size of \( C \)) is kept at bay, while the quality of the heuristic is sufficiently improved to boost search performance. The techniques are currently being evaluated empirically, and a paper submission is in preparation for ICAPS’12.

5.1.7. Accounting for Uncertainty in Penetration Testing

Participants: Olivier Buffet, Jörg Hoffmann.

Carlos Sarraute (Core Security Technologies) is an external collaborator.

Core Security Technologies is an U.S.-American/Argentinian company providing, amongst other things, tools for (semi-)automated security checking of computer networks against outside hacking attacks. For automation of such checks, a module is needed that automatically generates potential attack paths. Since the application domain is highly dynamic, a module allowing to declaratively specify the environment (the network and its configuration) is highly advantageous. For that reason, Core Security Technologies have been looking into using AI Planning techniques for this purpose. After consulting by Jörg Hoffmann (see also Section 6.1.1 below), they are now using a variant of Jörg Hoffmann’s FF planner (cf. Section 4.1) in their product. While that solution is satisfactory in many respects, it also has weaknesses. The main weakness is that it does not handle the incomplete knowledge in this domain – figuratively speaking, the attacker is assumed to have perfect information about the network. This results in high costs in terms of runtime and network traffic, for extensive scanning activities prior to planning. We are currently working with Core Security’s research department to overcome this issue, by modeling and solving the attack planning problem as a POMDP instead. A workshop paper detailing the POMDP model has been published at SecArt’11 [29]. While such a model yields much higher quality attacks, solving an entire network as a POMDP is not feasible. We have designed a decomposition method making use of network structure and approximations to overcome this problem, by using the POMDP model only to find good-quality attacks on single machines, and propagating the results through the network in an appropriate manner. A conference paper is in preparation for submission to ICAPS’12.

5.1.8. Searching for Information with MDPs

Participants: Mauricio Araya, Olivier Buffet, Vincent Thomas, François Charpillet.

In the context of Mauricio Araya’s PhD, we are working on how MDPs—or related models—can search for information. This has led to various research directions that we describe now.

A POMDP Extension with Belief-dependent Rewards — A limitation of Partially Observable Markov Decision Processes (POMDPs) is that they only model problems where the performance criterion depends on the state-action history. This excludes for example scenarios where one wants to maximize the knowledge with respect to some random variables. To overcome this limitation, we have proposed \( \rho \)-POMDPs, an extension of POMDPs in which the reward function depends on the belief state rather than on the state. In this framework, and under the hypothesis that the reward function is convex, we have proved that:

- the value function itself is convex; and
- if the reward function is \( \alpha \)-Hölder, then the value function can be approximated arbitrarily well with a piecewise linear and convex function.

These results allow for adapting a number of solution algorithms relying on approximating the value function.
This theoretical work has been first published in an international conference in December 2010, then in [36], where it has received a best paper award.

We are currently pursuing experimental work about the proposed algorithm.

Active Learning of MDP Models — Reinforcement Learning is about learning how to perform a task by trial and error (no model of the system to control being available). Model-based Bayesian RL (BRL) consists in all RL algorithms that maintain a belief (in the Bayesian sense) about the model of the system to control. In fact, this is a way to turn an RL problem into a POMDP—the unknown model becoming an unobservable part of the state—, thus replacing the exploration-exploitation dilemma by the definition of a prior belief over possible models.

A particular BRL task we have been considering is to actively learn the dynamical model itself, i.e., to act so as to improve the knowledge about the transition function. In a way this means solving a $\rho$-POMDP since the reward depends on a belief, not on a state. To that end, we have proposed several optimization criteria, and derived the corresponding reward functions, making sure that their computational complexity allows for their use in a BRL algorithm. We have also proved that a non-optimistic BRL algorithm—EXPLOIT—could be used in this particular case.

This work, along with experiments, has been published in [36] and [35] (french version).

PAC-BAMDP Algorithms — Exact or approximate solutions to Model-based Bayesian RL are impractical, so that a number of heuristic approaches have been considered, most of them relying on the principle of “optimism in the face of uncertainty”. Some of these algorithms have properties that guarantee the quality of their outcome, inspired by the PAC-learning (Probably Approximately Correct) framework. For example, some algorithms provably make in most cases the same decision as would be made if the true model were known (PAC-MDP property).

We have proposed a novel optimistic algorithm, BOUH, that is

- appealing in that it is (i) optimistic about the uncertainty in the model and (ii) deterministic (thus easier to study); and
- provably PAC-BAMDP, i.e., makes in most cases the same decision as a perfect BRL algorithm would.

First results about this algorithm are currently under review.

5.1.9. Scheduling for Probabilistic Realtime Systems

Participant: Olivier Buffet.

Maxim Dorin, Luca Santinelli, Liliana Cucu-Grosjean (INRIA, TRIO team), and Rob Davies (U. of York) are external collaborators.

In this collaborative research work (mainly with the TRIO team), we look at the problem of scheduling periodic tasks on a single processor, in the case where each task’s period is a (known) random variable. In this setting, some job will necessarily be missed, so that one will try to satisfy some criteria depending on the number of deadline misses.

We have proposed three criteria: (1) satisfying pre-defined deadline miss ratios, (2) minimizing the worst deadline miss ratio, and (3) minimizing the average deadline miss ratio. For each criterion we propose an algorithm that computes a provably optimal fixed priority assignment, i.e., a solution obtained by assigning priorities to tasks and executing jobs by order of priority.

This work has been presented in [26].

We also collaborate on other topics linked to real-time scheduling, as (i) on search algorithms for deterministic, but multiprocessor, problems [38], and (ii) on the problem of which jobs to drop (on-going work).

5.1.10. Adaptive Management with POMDPs

Participant: Olivier Buffet.
In the field of conservation biology, adaptive management is about managing a system, e.g., performing actions so as to protect some endangered species, while learning how it behaves. This is a typical reinforcement learning task that could for example be addressed through BRL.

Here, we consider that a number of experts provide us with one possible model each, assuming that one of them is the true model. This allows making decisions by solving a mixed observability MDP (MOMDPs), where the hidden part of the state corresponds to the model (in cases where all other variables are fully observable).

We have conducted preliminary studies of this approach, using the scenario of the protection of the Gouldian finch, and focusing on the particular characteristics that could be exploited to more efficiently solve this problem. First results have been presented in [39].

5.1.11. Information Gathering with Sensor Systems

**Participant:** Olivier Buffet.

Elodie Chanthery, Matthieu Godichaud (LAAS-CNRS) and Marc Contat (EADS) are external collaborators.

The DOPEC project was a DGA PEA (upstream studies project) on the optimization of the use of sensor systems. In collaboration with EADS (project leader) and the LAAS, we have worked on autonomous sequential decision making problems. We were more particularly interested, on the one hand, in multi-agent problems and, on the other hand, in taking uncertainties into account.

The overall architecture that has been developed in the context of this project was presented in a national and an international conference [40], [23].

5.1.12. How do real rats solve non-stationary (PO)MDPs ?

**Participant:** Alain Dutuch.

Etienne Coutureau and Alain Marchand (Centre de Neurosciences Intégratives et Cognitives (CNIC), UMR 5228, Bordeaux) are external collaborators.

For a living entity, using simultaneously various ways for learning models or representations of its environment can be very useful to adapt itself to non-stationary environments in a Reinforcement Learning setting. In the rats and in the monkey, two different action control systems lie in specific regions of the prefrontal cortex. Neurobiologists and computer scientists find here a common ground to identify and model these systems and the selection mechanisms between them, selection that could depend on uncertainty or error signals. Using real data collected on rats with or without prefrontal lesions, reinforcement learning models are used and evaluated in order to better understand this behavioral flexibility. MAIA is more particularly involved as a reinforcement learning expert in order to suggest and build models of the various learning mechanisms. In particular, we have used an on-policy learning scheme (SARSA) to investigate how well the use of simple or complex representations (with or without memory of the immediate past) can best model the learning behavior of rats in instrumental contingency degradation tasks [7].

This work has led us to investigate in more details the relations between the prefrontal cortex and the basal ganglia and their respective role when rats learn to solve non-stationary tasks. The research is conducted through the PEPII project IMAVO (see 7.2.7).

5.1.13. Developmental Reinforcement Learning

**Participants:** Alain Dutuch, Olivier Buffet.

Luc Sarzyniec and Joël Legrand (M2R Student of UHP Nancy 1) are external collaborators.
The goal of this work is to investigate how reinforcement learning can benefit from a developmental approach in the field of robotics. Instead of having a robot directly learn a difficult task using appropriate but rich (in the number of dimensions) sensory and motor spaces, we have followed an incremental approach. Both the number of perception and action dimensions increase only when the performance of the learned behavior increases. At the core of the algorithm lies a neuronal approximator used to compute the value function of the current policy of the robot. When the perception or action space grow, neurons or networks, initialized from existing neurons and networks, are added to the control architecture.

Thus far, our research focussed on the approximation architecture used to evaluate the Q-function. In simple robotic task, we investigated the use of Multi-Layer Perceptrons, either one approximation for every possible action ([41]) or one unique global approximator with as many outputs as the number of actions (Master Thesis of Joël Legrand). Currently, a reservoir computing architecture is under study as depicted in [16].

5.1.14. Classification-based Policy Iteration with a Critic

**Participant:** Bruno Scherrer.

Victor Gabillon, Alessandro Lazaric and Mohammad Ghavamzadeh (from Sequel INRIA-Lille) are external collaborators.

We study the effect of adding a value function approximation component (critic) to rollout classification-based policy iteration (RCPI) algorithms. The idea is to use the critic to approximate the return after we truncate the rollout trajectories. This allows us to control the bias and variance of the rollout estimates of the action-value function that are strongly related to the length of the rollout trajectories. Therefore, the introduction of a critic can improve the accuracy of the rollout estimates, and as a result, enhance the performance of the RCPI algorithm. We present in [49], [20] a new RCPI algorithm, called direct policy iteration with critic (DPI-Critic), and provide its finite-sample analysis when the critic is based on LSTD and BRM methods. We empirically evaluate the performance of DPI-Critic and compare it with DPI and LSPI in two benchmark reinforcement learning problems.

5.1.15. Linear Approximation of Value Functions

**Participant:** Bruno Scherrer.

Matthieu Geist (Supélec, Metz) is an external collaborator.

In the framework of Markov Decision Processes, we consider the problem of learning a linear approximation of the value function of some fixed policy from one trajectory possibly generated by some other policy. In [30], [42], [51], we describe a systematic approach for adapting on-policy learning least squares algorithms of the literature (LSTD, LSPE, FPKF and GPTD/KTD) to off-policy learning with eligibility traces. This leads to two known algorithms, LSTD(\(\lambda\))/LSPE(\(\lambda\)) and suggests new extensions of FPKF and GPTD/KTD. We describe their recursive implementation, discuss their convergence properties, and illustrate their behavior experimentally. Overall, our study suggests that the state-of-art LSTD(\(\lambda\)) remains the best least-squares algorithm.

We also consider the task of feature selection. A promising approach consists in combining the Least-Squares Temporal Difference (LSTD) algorithm with \(\ell_1\)-regularization, which has proven to be effective in the supervised learning community. This has been done recently with the LARS-TD algorithm, which replaces the projection operator of LSTD with an \(\ell_1\)-penalized projection and solves the corresponding fixed-point problem. However, this approach is not guaranteed to be correct in the general off-policy setting. In [21], we take a different route by adding an \(\ell_1\)-penalty term to the projected Bellman residual, which requires weaker assumptions while offering a comparable performance. This comes at the cost of a higher computational complexity if only a part of the regularization path is computed. Nevertheless, our approach ends up to a supervised learning problem, which let envision easy extensions to other penalties.
5.2. Understanding and mastering complex systems

5.2.1. Complex systems: simulation, control and definition

5.2.1.1. Adaptive control of a complex system based on its multi-agent model

Participants: Vincent Chevrier, Tomas Navarrete.

Laurent Ciarletta (Madynes team, LORIA) is an external collaborator.

We are interested in how to build a control mechanism for a complex/dynamic system. Specifically, we want to evaluate the effectiveness of creating a control mechanism based on a multi-agent model of the system [12]. Multi-agent models can be adapted to that purpose since usual approaches using analytical models as basis can be untractable when dealing with such systems; and because if we consider that the available control actions are meant to be applied locally, a multi-agent model is necessary. We are currently working on a case study within the dynamic networks domain, namely the free-riding phenomenon present in peer-to-peer networks.

We propose an architecture that gathers information from the system and uses it to parametrize and tune a set of multi-agent models. The outcome of simulations is used to decide which control actions have to be applied to the system, in order to achieve a predefined control objective. We consider that we do not have complete information to characterize the state of the system and hence would like to focus on the following two issues of the control problem that we have identified:

1. How to build a multi-agent model that represents the evolution of a dynamic network. That is, what to do when the information given by the simulation of the multi-agent is in contradiction with the information gathered from the system.
2. How to build an adaptive control mechanism based on the multi-agent model of a dynamic network. That is, how to use the information given by the multi-agent model to achieve the control objective.

The architecture we proposed, is designed as a control loop composed of the following steps: estimate the state of the system and instantiate multi-agent models accordingly, simulate different control actions, choose a control action and apply it. From one cycle to another of the control loop, each step can be tuned (in terms of model parameters, control action selection process, sampling strategy, etc.) to overcome the previously mentioned issues of the control problem.

The architecture is currently specified in terms of a formal notation. We have already implemented the architecture within the context of the free-riding problem where we use the PeerSim simulator as the target system to control.

Within our case study, we have conducted two different sets of experiments to investigate under which conditions our control architecture can achieve its goal and to investigate the efficiency of different sampling methods to estimate the state of the network. We have effectively managed to drive the system to a state where the majority of the peers share, when the initial conditions, without intervention from our architecture, would drive the system to a state where no peer would share.

The elements of the architecture having an impact on the performance of the control obtained have been identified. These are: the initialization of the parameters of the models used to estimate the state of the system, predict the evolution of the system and test the possible control actions, as well as the strategy used to observe the system and the different time horizons to consider within the architecture.

The next steps are to better identify the advantages and limits of the proposed architecture and to widen the problem family in the free riding problem.

5.2.1.2. Multi Modeling and multi-simulation

Participants: Vincent Chevrier, Julien Siebert.

This work is undertaken in a joint PhD Thesis between MAIA and Madynes Team. Laurent Ciarletta (Madynes team, LORIA) is co-advisor of this PhD.
Complex systems generally require to use different points of view (abstraction levels) at the same time on the system in order to capture and to understand all the dynamics and the complexity. Being made of different interacting parts, a model of a complex system also requires simultaneously modeling and simulation (M&S) tools from different scientific fields.

Building a model and a simulation of a complex system from the interaction of the different existing M&S tools present in each scientific field involved, is also a complex task. To represent a complex system, we need to couple several models (multi-modeling) that each represents a part of the whole system. Each model could have been designed by and for a specific scientific domain. Making different models interact raises hard issues on model interoperability (semantic coherence, formalism compatibility). As many simulators exist in the scientific fields involved, a possible approach to make a simulation of a complex system is to reuse and to make interact these existing simulators. Since each simulator has been developed for specific purposes, making them interact (multi-simulation) raises simulation issues (interoperability, synchronization).

The multi-agent paradigm is an homogeneous solution both for multi-modeling and multi-simulation of complex systems. On the one hand, a multi-agent model per se is a multi-model: a multi-agent model is made of interacting agent models and environment models. On the other hand, agent oriented software engineering (AOSE) allows designers to create complex softwares as a set of autonomous, heterogeneous and interacting softwares (i.e. as a multiagent system). Robustness, scalability, openness, modularity and interoperability are some of the properties that AOSE allows to achieve.

This work explores the contribution of multiagent paradigm to the fields of multi-modeling and multi-simulation of complex systems.

The first contribution of this work is to propose an homogeneous multiagent meta-model (called AA4MM[4]) that provides solutions both for multi-modeling and multi-simulation of complex systems by reusing existing and heterogeneous M&S tools. The core idea in AA4MM is to build a society of models, simulators and simulation softwares that solves the core challenges of multimodelling and simulation coupling in an homogeneous perspective. AA4MM has been implemented and used both for proof of concept and for a real case study. A proof of concept has been made by coupling different models together to develop a multi-model of a prey-predator model. This has permitted us to show both conceptual and operational properties of AA4MM such as interaction of heterogeneous models, modularity, interoperability.

This multiagent meta-model has been applied to model complex systems that are ubiquitous networks. Ubiquitous networks are highly dynamic computer networks that are composed of a great number of interacting and sometimes mobile nodes which can join or leave the system, interact together and where the environment plays a significant role either on radio communications or on the behavior of users. Modeling and simulation is the approach to evaluate these technologies or to build new ones.

5.2.1.3. Robustness of Cellular Automata and Reactive Multi-Agent Systems

Participants: Olivier Bouré, Vincent Chevrier, Nazim Fatès.

Our research on emergent collective behaviours focuses on robustness analysis, that is the behavioural resistance to perturbations in collective systems. We progressed in the knowledge of how to tackle this issue in the case of cellular automata (CA) and multi-agent systems (MAS).

We focused on the specific case of a perturbation of the updating scheme in CA, that is, changing the way cells are updated. Using similar ideas to the Influence-Reaction principle developed to resolve conflicts related to simultaneous actions, we created a new type of asynchronism, called beta-synchronism, which aims at disrupting the transmission of information about states between cells. We found out that the different types of asynchronism may induce radical change of behaviour for particular a value range of the synchrony rate [15].

More recently, our interest focused on a bio-inspired discrete dynamical system. Using the formalism of a subclass of cellular automata, lattice-gas CA, we study a model of swarming which displays qualitatively different behaviours under certain experimental conditions. We discussed these observations by relating them to the potential links with certain attributes of the model [48].
We studied a phase transition that occurs in the Greenberg-Hastings CA reaction-diffusion [5].

The density classification problem was taken as a typical framework for studying how decentralised computations can be carried out with simple cells. Although it is known that this problem can not be solved perfectly, we showed that using randomness provides a solution with an arbitrarily high success rate [17]. We also studied how to extend this result to the infinite-space case citerefPapier ???.

We studied the behaviour of the amoebae aggregation model [33] and applied the aggregation scheme on a robotic case (ALICE robots and Khepera III with Romea interactive table).

5.2.1.4. Ant algorithms for multi-agent patrolling

**Participants:** Olivier Simonin, François Charpillet, Olivier Buffet, Arnaud Glad.

We proposed in 2007 an ant algorithm, called EVAP, to deal with multi-agent patrolling, which is based on the marking and the evaporation of a digital pheromone. During the simulations carried out to measure the performances of EVAP, we identified that the system can self-organize towards stationary cycles (a periodic attractor). These cycles correspond to an Hamiltonian or quasi-Hamiltonian covering of the environment, which is an optimal or quasi-optimal solution to the multi-agent patrolling problem. We then established the mathematical proof that the system can stabilize only in cycles, one per agent, having the same length (cf. publication in ECAI’2008). Moreover, we introduced new heuristics in the agent behavior that improve dramatically the time for convergence, and we proved that under deterministic hypotheses the system always converges to stable cycles (these results have been published in SASO 2009, AAMAS’10). Results of 2011 are:

- Defense of Arnaud Glad’s PhD. thesis (November 15th) synthesising theoretical and experimental studies of the EVAP algorithm. The writing of a journal article is also in progress.
- EVAP has been adapted to continuous space in the context of the SUSIE project, which consider the surveillance of an area with a set of autonomous aerial robots.

5.2.2. Multi-robot systems: swarm intelligence, cooperation, navigation

5.2.2.1. Multi-robot exploration and mapping

**Participants:** Olivier Simonin, François Charpillet, Antoine Bautin.

In the context of the ANR Cartomatic project, introduced in Sec. 7.2.3, we study multi-agent models for multi-robot deployment and mapping. This work is in line with the PhD thesis of Antoine Bautin, started in November 2009. New results of 2011 are

- A new frontier assignation algorithm for multi-robot exploration has been proposed. It relies on counting the number of robots towards a frontier rather than considering only distances between robots and frontiers. We measured on benchmarks that the approach outperforms the two classical algorithms closest frontier and Greedy approach. Results are presented in [37], [43] and are submitted to ICRA’2012.
- We implemented and experimented the approach with autonomous mobile robots in the context of the ANR Carotte challenge (June 2011, Bourges). Our team “cartomatic” obtained one of the best map of the contest, while deploying several robots.

5.2.2.2. New experimental device: the Interactive Table

**Participants:** Olivier Simonin, François Charpillet, Nicolas Beaufort.

*Olivier Rochel (INRIA research engineer, SED Nancy) is an external collaborator.*

During 2010 we developed with the Nancy INRIA SED6 (Olivier Rochel) a new experimental device dedicated to swarm robotics study. It is composed of two independent components: an interactive table able to display and to compute any active environment and a set of autonomous mobile robots able to read and write information on the environment.

---

6 Service d’Expérimentation et Développement
Studies using the Table in 2011 are:

- We revisited the Drogoul & Ferber Foraging model, inspired by ants and also called “robot dockers” as the agents exchange the transported resources when they meet. From this simulated model we examined how it can be implemented with real mobile robots on an interactive environment, by considering that robots drop pheromones as ants. We defined a model extending the docker model with the robots on the Table, and studied its robustness to perception failure/mistakes. This work, done with Thomas Huraux (Master 2 Recherche internship), has been published in ICTAI 2011 Int. Conference [31].

- Several students (from Science Cog. Nancy 2 Master) implemented and explored pheromone-based foraging behaviors and flocking-based navigation models (supervised by François Charpillet and Christine Bourjot).

5.2.2.3. Local control based platooning

**Participants:** Alexis Scheuer, Olivier Simonin, François Charpillet, Jano Yazbeck.

We consider decentralised control methods to operate autonomous vehicles at close spacings to form a platoon. We study models inspired by the flocking approach, where each vehicle computes its control from its local perceptions. We investigate different decentralised models in order to provide robust and scalable solutions. Open questions concern collision avoidance, stability and multi-platoon navigation.

- **Coupling lateral and longitudinal controls.** A first work [67] focused on longitudinal control, which aims at computing velocities to avoid collision when all the vehicles are moving along a fixed path. When vehicles move in a two dimensional space, a lateral controller is needed to steer the vehicles. While lateral and longitudinal controls can be considered separately, the longitudinal control should be done after the lateral control: while turning, a higher inter-vehicle distance is needed to avoid collisions.

  An innovative approach to improve the quality of lateral control has been proposed during Jano Yazbeck’s internship at LORIA (03/10–07/10), entitled “Decentralised local approach for lateral control of platoons” and supervised by A. Scheuer and O. Simonin. This allows to reduce the distance between each vehicle’s path and the path of the previous vehicle, by using only embedded sensors such as a laser rangefinder. It relies on memorizing and computing in real time the previous vehicle relative trajectory. This work has been published in 2011 IEEE-RSJ International Conference on Intelligent Robots and Systems (IROS’2011) [34].

- **Finding an efficient lateral control.** To obtain an even better lateral control, and to drive each vehicle exactly in the trace of the previous one, we are developing a more efficient lateral control law. This law is defined in order to reduce exponentially the tracking error (which is more or less the distance between each vehicle’s path and the path of the previous vehicle). Once again, as for the longitudinal control [67], the formula of the control law is obtained through the proof of its property: necessary conditions are simplified in order to get the final result.

5.2.2.4. Adaptation of autonomous vehicle traffic to perturbations

**Participants:** Mohamed Tlig, Olivier Simonin, Olivier Buffet.

In the context of the European InTraDE project, one problem is to handle the displacements of numerous IAVs in a seaport. Here we assume a supervisor planning the routes of the vehicles in the port. However, in such a large and complex system, different unexpected events can arise and damage the traffic: failure of a vehicle, human mistake while driving, obstacle on roads, local re-planning, and so on.

We started focusing on a first important sub-problem of space resource sharing among multiple agents: how to ensure the crossing of two opposed flows of vehicles on a road when one of the two paths is blocked by an obstacle, e.g., a disabled vehicle. To overcome this problem, blocked vehicles have to coordinate with vehicles of the other side to share the road and manage delays. The objective is to improve traffic flow and reduce the emergence of traffic jam.

---

7 Intelligent Autonomous Vehicle
Solving this problem with reactive coordination methods is a major challenge of the PhD thesis of Mohamed Tlig (started in December 2010).

- We started by formalizing the problem and the possible actions of agents (vehicles) following a STRIPS formalism. We adapted this model dedicated to planning to the description of local rules in reactive coordination.
- We then defined and studied in simulation two decision rules that produce two different strategies: the first one alternates between two vehicles from each side of the road, and the second one gives priority to the vehicle with the highest delay. We are preparing a publication of these first results.

5.2.3. Ambient intelligence and Actimetry

5.2.3.1. Robotics and spatial computing: the iTiles - intelligent tiles - model

**Participants:** Olivier Simonin, François Charpillet, Lionel Havet.

*Olivier Rochel (INRIA research engineer, SED Nancy) is an external collaborator.*

In the context of intelligent home and assistant robots, we explore the definition and use of an active floor based on a cellular network approach. We aim at exploring spatial calculus models when considering physical cells augmented with sensors where robots and humans can evolve. Since 2009, we study a model consisting in paving the floor with interconnected tiles. Each tile can communicate with its neighbors and can sense the presence of a robot or a human. A first Tile model has been defined and evaluated using a tiles emulator and real mobile robots (Kheperas III), which validated the interest of the approach. See CAR’2010 publication [68].

In 2011 we designed with the help of INRIA Grenoble SED a prototype of 9 physical tiles embedding a WSN node able of computation and communication with other tiles. From this experimental device we explored several questions:

- How to follow a person walking on such a discrete and sensitive floor? We proposed a set of distributed algorithms allowing tiles to track a person or a robot (cooperation between neighboring tiles).
- How to make communications between a robot and the tile(s) it occupies? We developed a set of functions using the wifi communication of the tiles and the robot.
- From the work mentioned in the previous items we propose the definition of a new tile network based on the SensLab technology (wire connections between tiles and wireless communications between tiles and robots/humans). This prototype have been ordered and will be installed in the beginning of 2012.

5.2.3.2. Bayesian 3D Human Motion Capture Using Factored Particle Filtering

**Participants:** Abdallah Dib, Cédric Rose, Amandine Dubois, François Charpillet.

The gait deterioration of elderly people is an important factor in loss of autonomy and it increases the risk of falls. In order to evaluate this risk the MAIA team has been developing since 2003 a markerless human motion capture system that estimates the 3D positions of the body joints over time. The system uses a dynamic Bayesian network and a factored particle filtering algorithm. This year, we have evaluated the impact of using different observation functions for the Bayesian state estimation: chamfer distance, a pixel intersection and finally a pseudo-observation of the subject direction calculated from the previous output of the system. We also compared two methods for the factored generation of the particles. The first one uses a deterministic interval exploration strategy whereas the second one is based on an adaptive diffusion. The capacity of the system to recover after occlusion by obstacles was tested on simulated movements in a virtual scene [57].
An other achievement of the year has been the assessment of the accuracy and precision of this system, especially for measuring the step length of a walking human. This has been realized by Amandine Dubois during her research master [58]. An experiment with young subjects has been designed and realized. Measures of the markerless motion capture system were then compared with real values. These values were obtained through the footprints left by the subjects. Ink swabs placed at the front and rear of the shoes of each subject make it possible to mark a paper strip positioned on the ground. A statistical analysis of the results has been done by Amandine. Thus we were able to determine if the real and measured lengths were significantly different or not.

5.2.3.3. Automatic Evaluation of Vascular Access in Hemodialysis Patients

Participants: Cédric Rose, François Charpillet.

The vascular access that allows to perform the extra-corporeal circulation, is usually a vein of the arm that has been enlarged by a surgical creation of a fistula. The prevention of complications such as stenosis or thrombosis of the vascular access is a key issue in hemodialysis treatment. Many dialysis machines measure ionic dialysance by conductivity measures on the dialysate fluid. Ionic dialysance is an indicator of small molecules transfers through the dialysis membrane. Previous works have shown that the follow-up of the dialysance and the pressures along the extra corporeal circuit can help to detect at an early stage a potential complication on the vascular access. The difficulty of automating the follow-up is the large variability of the measures and the need to detect tendencies. Dynamic Bayesian networks (DBN) allow to formalize expert knowledge as a graphical stochastic model adapted to reasoning under uncertainty. In a DBN the state of the patient and the measurements are represented by interconnected temporal random variables. The relations between those variables are described using probability distributions. The proposed approach [64] is based on a supervised learning of a DBN for classifying the dialysis sessions according to a risk score describing the medical situation (0: no risk, 1: mild risk, 2: severe risk). The training of the system was performed using a dataset labeled by a medical expert. The evaluation of the results was done by performing a double-blind analysis of real data. The result was an 85% agreement rate between the human expert and the automated analysis. The purpose of the system is to assist the human expert by reporting abnormalities. The results show that a score 2 reported by the human is rarely missed by the automated analysis (only 1 case) whereas the opposite is more frequent (8 cases). The final decision to further investigate a case is taken by the human expert.
6. New Results

6.1. Modeling XML document transformations

Participants: Joachim Niehren, Sophie Tison, Sławek Staworko, Aurélien Lemay, Anne-Cécile Caron, Yves Roos, Shunichi Amano, Camille Vacher, Benoît Groz, Antoine Ndione, Tom Sebastian.

Query answering on XML streams. In [16], Gauwin and Niehren introduce the notion of finite streamability for query languages, and classify fragments of XPath that are finitely streamable or not. They show that if a query language is finitely streamable, then its satisfiability problem can be solved in polynomial time, which in turn is known to fail for mostly all fragments of XPath. They also show that FXP, the fragment of ForwardXPath with child and descendant axis, conjunction, and negation becomes finitely streamable if bounding the number of conjunctions. Since 3 conjunctions are enough in many practical applications, FXP is most relevant in practice. Without any bound, FXP is not finitely streamable, since its satisfiability problem is DEXPTIME hard. The positive result for FXP with a bounded number of conjunctions is obtained by compilation of FXP to deterministic nested word automata. The compiler is in exponential time in the number of conjunctions, and thus polynomial if this parameter is bounded.

Answer enumeration for \(n\)-ary queries. Bagan, Filiot, Gauwin, and Niehren investigated answer enumeration algorithms for dialects of XPath with variables. The problem with \(n\)-ary queries is that answer sets may grow exponentially in \(|t|^n\), so that algorithms depending polynomially on the size of the answer set might still be unfeasible. In such case, it might still be possible to enumerate elements of answer sets on need. The questions is then whether enumeration can be done efficiently without duplicates and failures, that is with constant delay between subsequent answers and polynomial time preprocessing in the size of the query and the tree. We obtained positive results on answer enumeration with constant delay enumeration for acyclic conjunctive queries over so called X-doublebar structures that we introduce [24]. These subsume tree structures with child, next-sibling and next-sibling* axis, but not the descendant axis. Our result can be lifted to a dialect of ConditionalXPath with variables, that is FO-complete on trees of bounded depth, so that the descendant axis is not needed.

Tree automata global constraints. TAGEDs are a new class of tree automata with constraints that currently receive much interest from top conferences on theoretical computer science. During its postdoc in Lille, Vacher improved complexity bounds for some fragments of decidability results [12].

Sequential tree-to-word transducers. Laurence, Lemay, Niehren, Staworko, Tommasi considered deterministic sequential top-down tree-to-word transducers (ST\(\omega\)), that capture the class of deterministic top-down nested-word to word transducers. While reordering and copying are not allowed, ST\(\omega\)s are nevertheless very expressive because they allow concatenation of outputs, deletion of inner nodes and they can produce context free languages as output. Their expressiveness is incomparable with DT\(\omega\)Ps (plus concatenation, but minus copying). While objecting for learning algorithms, they study normalization of ST\(\omega\)s in a first step and then develop unique minimalization algorithms for normalized ST\(\omega\)s in a second step in [19]. The idea of normalization is to produce the output in an earliest manner, when reading the input in document order. This works only on binary trees, but can be lifted to unranked trees modulo the binary top-down encoding. The normalization algorithm is by far nontrivial. The natural continuation of this approach will be toward learning algorithms for earliest ST\(\omega\)s.

Access control for XML views. The PhD project of Groz, supervised by Staworko and Tison, is centered on access control for XML databases, and in particular on security of user views over XML documents. He obtained results on query rewriting for read-only queries, and translation for update queries. More precisely, given an XML view definition and a user defined query (resp. update program) \(q\), the problem is to find a source query (resp. update program) that is equivalent to \(q\) on the view. Caron, Groz, Roos, Staworko and Tison study update programs and views represented by recognizable tree languages in [15], and devise algorithms for update translation in different settings, namely without or with constraints on the authorised source updates.

Participants: Jérôme Champavère, Jean Decoster, Jean-Baptiste Faddoul, Antonino Freno, Gemma Garriga, Rémi Gilleron, Mikaela Keller, Grégoire Laurence, Aurélien Lemay, Joachim Niehren, Sławek Staworko, Marc Tommasi, Fabien Torre.

Induction of tree automata. Champavère, Gilleron, Lemay and Niehren proposed to use schemas for improving induction algorithms for monadic queries represented by tree automata [26]. The idea is to use pruning strategies to eliminate useless parts of trees when learning from partially annotated trees such that only the structure of relevant fragments is learned. This allows to avoid generalization errors and to learn from fewer annotations. They define schema-guided pruning strategies. They define stable queries w.r.t. a pruning strategy and show that stable queries are learnable.

Further Results. In [21], Staworko proposed learning twig and path queries. Prioritized repairing and consistent query answering in relational databases was tackled in [13] and Bounded repairability for regular tree languages in [20]. Torre and Terlutte explored the combination of automata and words balls for sequences classification in [14]. Tommasi participated in the writing of a chapter of a book on conditional Markov fields for information extraction [23].

Garriga and collaborators from the Fraunhofer Institute in Bonn, studied fixed parameter tractable algorithms for the discovery of maximal order preserving submatrices in bioinformatic applications in [17].

We begun a new activity on learning for social network and information network supported by the arrivals of Gemma Garriga, Mikaela Keller and Antonino Freno. Freno, Garriga and Keller [22] proposed a model for predicting new links in a network which exploit both the current structure of the network and the content of its node.
ORPAILLEUR Project-Team

6. New Results

6.1. The Mining of Complex Data


Formal concept analysis, together with itemset search and association rule extraction, are suitable symbolic methods for KDDK, that may be used for real-sized applications. Global improvements may be carried on the scope of applicability, the ease of use, the efficiency of the methods, and on the ability to fit evolving situations. Accordingly, the team is working on extensions of such symbolic data mining methods to be applied on complex data such as biological or chemical data or textual documents, involving objects with multi-valued attributes (e.g. domains or intervals), n-ary relations, sequences, trees and graphs.

6.1.1. FCA, RCA, and Pattern Structures

Recent advances in data and knowledge engineering have emphasized the need for Formal Concept Analysis (FCA) tools taking into account structured data. There are a few extensions of FCA for handling contexts involving complex data formats, e.g. graphs or relational data. Among them, Relational Concept Analysis (RCA) is a process for analyzing objects described both by binary and relational attributes [116]. The RCA process takes as input a collection of contexts and of inter-context relations, and yields a set of lattices, one per context, whose concepts are linked by relations. RCA has an important role in KDDK, especially in text mining [85], [84].

Another extension of FCA is based on Pattern Structures (PS) [90], which allows to build a concept lattice from complex data, e.g. nominal, numerical, and interval data. In (major [5]), pattern structures are used for building a concept lattice from intervals, in full compliance with FCA, thus benefiting of the efficiency of FCA algorithms. Actually, the notion of similarity between objects is closely related to these extensions of FCA: two objects are similar as soon as they share the same attributes (binary case) or attributes with similar values or the same description (at least in part). Various results were obtained in the study of the relations existing between FCA with an embedded explicit similarity measure and FCA with pattern structures [48]. Moreover, similarity is not a transitive relation and this lead us to the study of tolerance relations. In addition, a new research perspective is aimed at using frequent itemset search methods for mining interval-based data being guided by pattern structures and biclustering as well [50], [49].

Pattern structures in association with a similarity measure were applied in the field of decision support in agronomy. In this domain, a set of agro-ecological indicators is aimed at helping farmers to improve their agricultural practices by estimating the impact of cultivation practices on the “agrosystem”. The modeling and the assessment of environmental risk require a large number of parameters whose measure is imprecise. The propagation of the imprecision and the different types of imprecision have to be taken into account in the computation of the value of indicators for decision support. Actually, based on pattern structures with a associated similarity measure, this problem has been approached as an information fusion problems with substantial results [34], [35].

6.1.2. Miscellaneous in FCA and Pattern Mining

In the field of medicine, an approach based on a combination of FCA with sequential pattern mining was developed to explore patients care trajectories (PCT) [46]. When PCT are modeled as multidimensional and multilevel sequences [108], the results of a frequent sequential itemsets search feed an FCA step in order to compute interests measures such as concept stability. These measures help the experts to find the most interesting sequential patterns.
In the context of environmental sciences, research work is in concern with the mining of complex hydroecological data with concept lattices. FCA was compared and combined with statistical approaches to deal with multi-valued contexts in hydroecology [31], [27], [39]. Regarding the preparation of agronomical data, we have developed an episode-based analysis about the design of information systems (actually, this work was carried out during the ANR-ADD COPT project between 2005 and 2008). We focused on the experience of persons in charge of building observatoires, i.e. information systems, for the monitoring and the management of rural territories [32]. Moreover, Florence Le Ber –as a member of UMR 7517 Lhyges, Strasbourg– is the scientific head of an ANR project named “FRESQUEAU” (2011–2014) dealing with FCA and data mining and hydroecological data (see http://fresqueau.engees.eu/).

For completing the work on itemset search, there is still on-going work on frequent and rare itemset search, for being able to build lattices from very large data and completing the algorithm collection of the Coron platform. This year, results were obtained on the design of an integrated and modular algorithm for searching for closed and generators itemsets, and equivalence classes of itemsets, thus enabling the construction of the associated lattice [56]. This research aspect is also linked to the research carried on within a the PICS CaDoE research project (see Section 8.1.3).

### 6.1.3. Skylines, sequences and privacy

Pattern discovery is at the core of numerous data mining tasks. Although many methods focus on efficiency in pattern mining, they still suffer from the problem of choosing a threshold that influences the final extraction result. The goal of a study done this current year (2011) is to make the results of pattern mining useful from a user-preference point of view. That is, take into account some domain knowledge to guide the pattern mining process. To this end, we integrate into the pattern discovery process the idea of skyline queries in order to mine skyline patterns in a threshold-free manner. This forms the basis for a novel approach to mining skyline patterns. The efficiency of our approach was illustrated over a use case from chemoinformatics and we showed that small sets of dominant patterns are produced under various measures that are interesting for chemical engineers and researchers [55].

Sequence data is widely used in many applications. Consequently, mining sequential patterns and other types of knowledge from sequence data has become an important data mining task. The main emphasis has been on developing efficient mining algorithms and effective pattern representation.

However, important fundamental problems still remained open: (i) given a sequence database, can we have an upper bound on the number of sequential patterns in the database? (ii) Is the efficiency of the sequence classifier only based on accuracy? (iii) Do the classifiers need the entire set of extracted patterns or a smaller set with the same expressiveness power?

In three different works on sequences, we study the problem of bounding sequential patterns with the combinatorial complexity of sequences and the problem of sequence classifiers with the constraints of optimizing both accuracy and earliness [53], [46].

Orpailleur is one of the few project-teams working on privacy challenges which are becoming a core issue with different scientific problems in computer science. Privacy-preserving data publication has been studied intensely in the past years. In our recent works, we introduce two different data anonymization methodologies based on different usability scenarios [57], [58].

### 6.1.4. KDDK in Text Mining

Ontologies help software and human agents to communicate by providing shared and common domain knowledge, and by supporting various tasks, e.g. problem-solving and information retrieval. In practice, building an ontology depends on a number of “ontological resources” having different types: thesaurus, dictionaries, texts, databases, and ontologies themselves. We are currently working on the design of a methodology and the implementation of a system for ontology engineering from heterogeneous ontological resources. This methodology is based on both FCA and RCA, and was previously successfully applied in contexts such as astronomy and biology. At present, an engineer is in charge of implementing a robust system
being guided by the previous research results and preparing the way for some new research directions involving trees and graphs.

In another work in text mining [19], we propose a method based on a syntactic parsing for extracting rich semantic relationships between pairs of entities co-occurring in a single sentence. The method was applied in pharmacogenomics (study of the impact of individual genomic variation on drug responses) and we obtained a resource encoded in RDF that summarizes pharmacogenomics relationships mentioned in roughly 17 million Medline abstracts. This resource appears to be of major interest since it is used to guide human curation of biomedical databases, and to derive new knowledge about drug-drug interactions [92].

6.2. KDDK in Life Sciences


One of the major challenges in the post genomic era consists in analyzing terabytes of biological data stored in hundreds of heterogeneous databases (DBs). The extraction of knowledge units from these large volumes of data would give sense to the present data production effort with respect to domains such as disease understanding, drug discovery, and pharmacogenomics or systems biology. Research reported here addresses these important issues and shows the spreading of KDDK over such domains.

6.2.1. Ontology-based Functional Classification of Genes

Functional classification involves grouping genes according to their molecular functions or the biological processes they participate in. This unsupervised classification task is essential for interpreting gene datasets produced by postgenomic experiments. As the functional annotation of genes is mostly based on the Gene Ontology (GO), many similarity measures using the GO have been described, but few of them have been used for clustering [107]. We have evaluated a functional classification of genes using our previously described IntelliGO semantic similarity measure with the help of reference sets [38]. The IntelliGO measure computes semantic similarity between genes for discovering biological functions shared by genes and takes into account domain knowledge represented in Gene Ontology [82]. The reference sets consist of genes taken from human and yeast KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways and Pfam clans. Hierarchical clustering and heatmap visualization were used to illustrate the advantages of IntelliGO over several other measures. Because genes often belong to more than one reference set, the fuzzy C-means clustering algorithm was then applied to the datasets using IntelliGO. The F-score method was used to estimate the quality of clustering and the optimal number of clusters. The results were compared with those obtained from the state of the art DAVID (Database for Annotation Visualization and Integrated Discovery) functional classification method. Overlap analysis allows to study the matching between clusters and reference sets, and leads us to propose a set-difference method for discovering missing information [38]. The IntelliGO similarity measure, the clustering tool and the reference sets used for the evaluation are available at http://plateforme-mbi.loria.fr/intelligo.

6.2.2. Use of Domain Knowledge for Dimension Reduction

Data complexity is a major challenge for knowledge discovery approaches. High dimensionality of datasets can impair the execution of most data mining programs and/or lead to the production of numerous and complex patterns, improper for interpretation by the supervising expert. Thus, an important research orientation is dimension reduction as part of the data preparation step [93]. Domain knowledge is essential for achieving such dataset modification with minimal loss of information. The Life Sciences constitute a suitable domain for testing knowledge-guided approaches for dimension reduction because of the continuous increase in the number of both complex datasets and bio-ontologies. Most of these bio-ontologies are used for annotating biological objects leading to high-dimensional datasets. We propose a new approach for reducing dimensions
in a dataset by exploiting semantic relationships between terms of an ontology structured as a rooted directed acyclic graph [40]. Term clustering is performed thanks to the IntelliGO similarity measure and the term clusters are further used as descriptors for data representation. The technique was applied to a set of drugs associated with their side effects collected from the SIDER database. Terms describing side effects belong to the MedDRA terminology. The hierarchical clustering of about 1,200 MedDRA terms into an optimal collection of 112 term clusters led to a reduced data representation. Two data mining experiments were conducted to illustrate the advantage of using such reduced data representation.

Results obtained in the frame of the collaborative Grand Challenge project (see previous report 2009 and 2010) have been published this year. We have designed the HIV-PDI (Protein-Drug Interactions) resource as a decision making tool to propose alternative antiretroviral drugs (ARVs) for personalized antiretroviral treatment [22]. The HIV-PDI is an integrated database in which sequence mutations of viral proteins can be mapped onto three-dimensional structural interactions between these proteins and ARVs. Thus, critical loss of interactions leading to resistance can be detected and serve as indicators for proposing appropriate ARVs escaping the resistance. As a first step, the HIV-PDI was populated with data relating to HIV protease: clinical information on patients, resistance to ARVs treatments, HIV protease structures and mutations, ARV drugs and their 3D interactions with HIV protease models. Possible queries include protein, drug and treatment conditions, coupled with dedicated tools for visualization/analysis of 3D Protein-Drug interactions. Case-studies demonstrate the capabilities of the HIV-PDI resource for retrieving information associated with patients and for analyzing structural data relating proteins and ligands [23].

6.2.3. Mining Agronomical Data with stochastic models

In the framework of agricultural landscape data mining, we have developed an original approach combining two methods used separately so far: the identification of explicit farmer decision rules through on-farm surveys methods and the identification of landscape stochastic regularities through data-mining of the mosaic of agricultural parcels, following preceding work [96]. This approach was assessed in a study on the Niort plain (West of France) database. In this database, provided by the CEBC (UPR CNRS), the land use occupations of the fields covering a 400 km² area are recorded during 12 years. It results a segmentation of the landscape, based on both its spatial and temporal organization and partly explained by generic farmer decision rules. This consistency between results points out that the two modelling methods interact and may be combined for land-use modelling at landscape scale and for understanding the driving forces of spatial organization. Based on farm surveys, we were able to retrieve and measure changes in land use occupation and link some farmer decision and spatiotemporal regularities that were observed in the landscapes.

6.3. Structural Systems Biology and Docking

Participants: Thomas Bourquard, Marie-Dominique Devignes, Anisah Ghoorah, Bernard Maigret, Lazaros Mavridis, Violeta Pérez-Nueno, Dave Ritchie, Malika Smaïl-Tabbone, Vishwesh Venkatraman.

Structural systems biology aims to describe and analyze the many components and interactions within living cells in terms of their three-dimensional (3D) molecular structures. Much of our work in this area has been funded by the ANR Chaires d’Excellence project entitled “High Performance Algorithms for Structural Systems Biology” (HPASSB) which was awarded to Dave Ritchie (January 2009 – September 2011). A related follow-on ANR project entitled “Polynomial Expansions of Protein Structures and Interactions” (PEPSI) has recently started (November 2011). The HPASSB project complements existing competencies in the Orpailleur team represented by Marie-Dominique Devignes (CR CNRS) who is coordinating the MBI project (Modelling Biomolecules and their Interactions, http://bioinfo.loria.fr ), Malika Smaïl-Tabbone (MCU Nancy University) who is working on data integration and relational data-mining approaches, and Bernard Maigret (DR CNRS) who has an extensive experience of molecular dynamics and virtual screening. We are currently developing advanced computing techniques for molecular shape representation, protein-protein docking, protein-ligand docking, high-throughput virtual drug screening, and knowledge discovery in databases dedicated to protein-protein interactions. The PEPSI project is a collaboration with Sergei Grudinin at INRIA Grenoble (project Nano-D) and Valentin Gordeliy at the Institut de Biologie Structurale in Grenoble. This new project will
involve developing further the above techniques and using them to help solve the structures of large molecular systems experimentally.

6.3.1. Accelerating protein docking calculations using graphics processors

We have recently adapted the Hex protein docking software to use modern graphics processors (GPUs) to carry out the expensive FFT part of a docking calculation [115]. Compared to using a single conventional central processor (CPU), a high-end GPU gives a speed-up of 45 or more. This software is publicly available at http://hex.loria.fr. A public GPU-powered server has also been created (http://hexserver.loria.fr) [99]. These advances have facilitated further work on modeling the assembly of multi-component molecular structures using a particle swarm optimization technique [69].

6.3.2. Eigen-Hex: Modeling protein flexibility during docking

Although the Hex protein docking software can often make reasonably good predictions about how two proteins might fit together, a major limitation of many current algorithms, including Hex, is that they assume that proteins are rigid objects. In fact, proteins can be highly flexible, and the internal conformations of their atoms often change on going from the unbound forms in the free proteins to the bound conformations in the complex. We have developed a novel approach to model such flexibility using a principal component analysis (PCA) technique to identify and predict the main atomic motions during a docking calculation. Our approach gives better results than rigid body docking, although the flexible docking problem is still by no means solved. A journal article describing this work has been submitted.

6.3.3. 3D-Blast: A new approach for protein structure alignment and clustering

We recently developed a new sequence-independent protein structure alignment approach called 3D-Blast [102], which exploits the spherical polar Fourier (SPF) correlation technique used in the Hex protein docking software [114]. This approach recently performed very well in a blind shape comparison experiment organized by Orpailleur as part of Eurographics Workshop on 3D Object Retrieval [103]. The utility of this approach has been demonstrated by clustering subsets of the CATH protein structure classification database [106] for each of the four main CATH fold types, and by searching the entire CATH database of some 12,000 structures using several protein structures as queries. Overall, the automatic SPF clustering approach agrees very well with the expert-curated CATH classification, and ROC-plot analysis of database searches show that the approach has very high precision and recall. We recently proposed that the 3D-Blast approach could ultimately provide a novel way to enumerate and index protein fold space (major [7]).

6.3.4. KBDOCK: Protein docking using Knowledge-Based approaches

Protein docking is the difficult computational task of predicting how a pair of three-dimensional protein structures come together to form a complex. Historically, there has been considerable interest in developing \textit{ab initio} docking algorithms such as the Hex docking program developed by Dave Ritchie. However, as structural genomics initiatives continue to populate the space of protein 3D structures, and as several on-line databases of protein interactions have recently become available, using structural database systems to perform docking by homology will become an increasingly powerful approach to predicting protein interactions. In order to explore such possibilities, Anisah Ghoorah has recently developed the KBDOCK system as part of her doctoral thesis project. KBDOCK combines residue contact information from the 3DID database [117] with the Pfam protein domain family classification [89] together with coordinate data from the Protein Data Bank [86] in order to describe and analyze all known protein-protein interactions for which the 3D structures are available. In a recent publication [24] we demonstrated the utility of this approach for template-based docking using 73 complexes from the Protein Docking Benchmark [94]. KBDOCK is available at http://kbdock.loria.fr.

6.3.5. V-Dock: scoring protein-protein interactions using Voronoi fingerprints
There is growing interest in using machine learning techniques to analyze and populate protein-protein interaction (PPI) networks [104]. The aim of this project is to investigate the use of Voronoi fingerprints [16] as a way to distinguish cognate and non-cognate pairs of protein-protein interfaces. In collaboration with colleagues in the INRIA AMIB and INRA Bios teams, we recently applied our Voronoi fingerprint representation (V-Dock) to re-score rigid body docking predictions from Hex [60], and we demonstrated that it could be used to improve the ranking of 7 out of 9 docking targets from the CAPRI protein docking experiment [60]. This approach was also used to predict the stability of engineered protein structures for another recent CAPRI target [21].

6.3.6. DOVSA: Developing new algorithms for virtual screening

In 2010, Violeta Pérez-Nueno joined the Orpailleur team thanks to a Marie Curie Intra-European Fellowship (IEF) award to develop new virtual screening algorithms (DOVSA). The aim of this project is to advance the state of the art in computational virtual drug screening by developing a novel consensus shape clustering approach based on spherical harmonic (SH) shape representations [110]. The main disease target in this project is the acquired immune deficiency syndrome (AIDS), caused by the human immuno-deficiency virus (HIV) [109]. However, the approach will be quite generic and will be broadly applicable to many other diseases. So far, good progress has been made on calculating and clustering spherical harmonic “consensus shapes” which represent rather well the essential features of groups of active molecules [30]. Recent progress on this project has been presented orally at the 5th Journée Nationale de Chémoinformatique in Cabourg, the 9th International Conference on Chemical Structures in Noordwijkerhout, and at 3rd International Conference on Drug Discovery and Therapy in Dubai. A review of the state of the art in drug promiscuity was also recently published [29].

6.4. Around the Kasimir research project

Participants: Nicolas Jay, Jean Lieber, Bart Lamiroy, Amedeo Napoli, Thomas Meilender.

This special research project involves researchers working around the Kasimir project and Bart Lamiroy who was attached to the Orpailleur Team during his “INRIA délégation” (2010-2011) and at the same time was a visiting scientist at Lehigh University, USA. The background of Bart Lamiroy is in document and image analysis. Recently he was interested in investigating the application of KDDK to numerical and structural data including document images. The objective is to extend mining tools towards complex and semi-structured multi-media data on the one hand, and to associate image analysis with KDDK techniques on the other hand.

The main research direction which is followed at the moment is in concern with the Kasimir project. Actually, oncology protocols are mainly documented and represented in diagram formats. The classification and CBR techniques used in the Kasimir project require that the ontologies and decision protocols have to be represented in OWL. Based on previous work, we started modeling the mapping of visual features in diagram charts with semantics of the medical domain ontology. The mapping between the visual ontology and the domain ontology should guide a more complete extraction of the protocols from the diagrams for completing the domain ontology of the Kasimir system.

Moreover, during his stay at Lehigh University, Bart Lamiroy developed a new approach for recovering useful information within image data. By recording a wide range of “provenance information” related to complex image analysis processes, the DAE platform (http://dae.cse.lehigh.edu) provides a large set of metadata that can be used by KDDK methods. For example, this allows the correlation and combination of numerical and symbolic aspects, e.g. relating image aspects and domain symbolic representations (within domain ontologies). This work bridges the gap between formal knowledge representation and signal-based pattern recognition and offers a robust experimental environment for further application of KDDK on image data.

6.5. Around the Taaable research project

The Taaable project (http://taaable.fr) has been originally created as a challenger of the Computer Cooking Contest (CCC, organized during the ICCBR Conference). A candidate to this contest is a system whose goal is to solve cooking problems on the basis of a recipe book (common to all candidates), where each recipe is a shallow XML document with an important plain text part. The size of the recipe book (about 1500 recipes) prevents from a manual indexing of recipes: this indexing is performed using semi-automatic techniques.

After being ranked twice second, in the 2008 and 2009 CCCs organized during the ICCBR conference, Taaable won the first price and the adaptation challenge, in 2010 (note that no contest was organized in 2011). Beyond its participation to the CCCs, the Taaable project aims at federating various research themes: case-based reasoning, information retrieval, knowledge acquisition and extraction, knowledge representation, minimal change theory, ontology engineering, semantic wikis, text-mining, etc.

The most important original features of this version are:

A module for refining the domain ontology for improving the case retrieval. In Taaable, the retrieval of similar cases is based on a query generalization using an ontology of the cooking domain. In order to make the case retrieval more progressive and more precise, an enrichment of the domain ontology, and especially the ingredient hierarchy, has been studied and implemented [42]. The refinement process consists in inserting intermediate classes into the initial hierarchy of the system for better distinguishing classes that were initially not distinguishable. In order to introduce new classes into the initial hierarchy, the initial classes of ingredients have been characterized with additional properties. These additional properties are cooking actions that can be applied to ingredients and that have been extracted from the texts of recipes. FCA has been used on these new properties for restructuring the initial hierarchy.

A module for computing adaptation knowledge. Adaptation knowledge discovery has been performed for better adapting cooking recipes to user constraints. This paper extends the approach proposed in 2009 [80] for extracting this kind of adaptation knowledge. The adaptation knowledge comes from the interpretation of closed itemsets whose items correspond to the ingredients that have to be removed, kept, or added. An original approach focusing on a restrictive binary context building and on a specific ranking based on the form of the closed itemsets has been proposed [47].

Several theoretical studies have been carried out that should be applied to some future versions of Taaable:

- The representation of preparations in temporal representation formalisms [63].
- An algorithm for adapting cases defined in the expressive description logic $\mathcal{ALC}$ [43], [11].
- The study of the relations between adaptation based on belief revision and other approaches to adaptation [61], [11].
- The study of the extension of the domain ontology to make the retrieval step of a case-based reasoning system more accurate [42].
- The study of adaptation knowledge discovery based on variation of ingredients between pairs of recipes [42].
6. New Results

6.1. Embedded data management

Participants: Nicolas Anciaux, Luc Bouganim, Yanli Guo, Lionel Le Folgoc, Philippe Pucheral, Shaoyi Yin.

Inspired by low cost economic models, this work draws the idea of a one-dollar database machine, with the objective to disseminate databases everywhere, up to the lightest smart objects. In contrast to traditional database machines relying on massively parallel architectures, the one-dollar database machine considers the cheapest form of computer available today: a microcontroller equipped with GBs size (external) Flash storage. Designing such a database machine is very challenging due to a combination of conflicting RAM and NAND Flash constraints. To tackle this challenge, this work proposes a new paradigm based on database serialization (managing all database structures in a pure sequential way) and stratification (restructuring them into strata when a scalability limit is reached). We show that a complete DBMS engine can be designed according to this paradigm and demonstrate the effectiveness of the approach through a performance evaluation.

This work capitalizes on previous results related to the indexing of Flash resident data [39] and has also obvious connections with the more general study we are conducting on Flash-based data management (see Section 6.2). Partial elements of this solution have been demonstrated at [28].

6.2. Flash-based Data Management

Participant: Luc Bouganim.

Bimodal flash devices. While disks have offered a stable behavior for decades, thus guaranteeing the timelessness of many database design decisions, flash devices keep on mutating. Many researchers have proposed to adapt database algorithms to existing flash devices. However, today, there is no reference DBMS design based on solid assumptions of flash devices behavior, precisely because flash device behavior varies across models, across firmware updates and possibly over time for the same model: database researchers are running after flash memory technology. In this study, we took the reverse approach and defined how flash devices should support database management. We advocated that flash devices should provide guarantees to a DBMS so that it can devise stable and efficient IO management mechanisms. Based on the characteristics of flash chips, we defined a bimodal FTL that distinguishes between a minimal mode where sequential writes, sequential reads and random reads are optimal while updates and random writes are forbidden, and a mode where updates and random writes are supported at the cost of sub-optimal IO performance. This work started at the end of 2010 and was published at CIDR’11 [19], in cooperation with the IT University of Copenhagen. DBMS/Flash device co-design considerations were the focus of a tutorial on flash devices given recently at VLDB 2011 [20].

6.3. Privacy-Preserving Data Publishing

Participants: Tristan Allard, Benjamin Nguyen, Philippe Pucheral.

While most PPDP works make the assumption of a trusted central publisher, this study advocates a decentralized way of publishing anonymized datasets. More precisely, our work concerns the proof of feasibility of adapting traditional PPDP schemes, such as \(k\)-anonymity, \(\ell\)-diversity or differential privacy to encompass the use of secure portable devices. In the applications we consider, each secure device is a data provider with weak computing capacities and weak connectivity (frequency and duration of connections are unpredictable)\(^1\). Weak

\(^1\)E.g., in the e-health context, patients may have their medical folder embedded in a secure device and connect it sporadically when they visit their physician or when they want to consult it at home.
connectivity precludes any P2P solution to the problem. A server allowing asynchronous communications between the devices becomes necessary to implement a distributed PPDP mechanism but this server does not benefit from the same trustworthiness as the participating devices. Our work aims to provide a generic method to adapt an important subclass of PPDP algorithms to this context, using both the limited secure computation capacities of each device (but taking advantage of their number) and the powerful computation abilities of an untrusted server available 24/7. Our proposal is based on a meta algorithm divided in three phases: (1) a collection phase where encrypted data is collected by the untrusted server, (2) a construction phase where the untrusted server performs a sound computation of a given privacy mechanism to generate sanitization rules and (3) a sanitization phase where the encrypted data is decrypted then sanitized by the devices to produce a final clear-text result. The last phase can be distributed using many different devices for better efficiency. In [15], [17], we showed how it is possible to transform existing anonymity mechanisms into decentralized ones using secure devices, while maintaining equivalent security guarantees against honest-but-curious and weakly malicious adversaries. In [16], we studied the (unlikely) event that some secure devices might be compromised, and can collude with the untrusted server. We provided schemes to detect the compromised devices with a probability that can be fixed as close to 1 as desired (the trade-off being the latency of the protocol).

6.4. Minimal Exposure

Participants: Nicolas Anciaux, Benjamin Nguyen.

When users request a service, the service provider usually asks for personal documents to tailor its service to the specific situation of the applicant. For example, the rate and duration of consumer’s loans are usually adapted depending on the risk based on the income, assets or past lines of credits of the borrower. In practice, an excessive amount of personal data is collected and stored. Indeed, a paradox is at the root of this problem: service providers require users to expose data in order to determine whether that data is needed or not to achieve the purpose of the service. We currently explore a reverse approach, where service providers would publicly describe the data they require to complete their task, and where the applicants would confront those descriptions with their own data to determine themselves the minimal subset of information to expose. We have first investigated solutions for simplistic tasks (e.g., evaluating a decision tree to determine the loan rate and duration a given applicant can claim), and we plan to address more complex ones (e.g., building the profile of customers, mining association rules, etc.) in the short term. The work on Minimal Exposure has just started and a first paper is under evaluation.

6.5. Experiment in the medical field

Participants: Nicolas Anciaux, Luc Bouganim, Lionel Le Folgoc, Philippe Pucheral, Alexei Troussov.

The PlugDB engine is being experimented in the field since September 2011 to implement a secure and portable medical-social folder. The objective is to improve the coordination of medical care and social services provided at home for dependent people. Details related to this experiment conducted with about 120 practitioners and patients are given in Section 7.2. While this action did not generate new academic results (though it helped us validating some previous results), it imposed us a strong investment in terms of test and optimization for our prototype and in terms of communication to promote this experiment at the regional level.
6. New Results

6.1. Multimedia Models and Formats

6.1.1. SMIL timesheets

With the advent of HTML5 and its support in most popular browsers, HTML is becoming an important multimedia language. Video and audio can now be embedded in HTML pages without worrying about the availability of plugins. However, a major issue is to specify the dynamic behavior of documents (user interactions, timing and synchronization with continuous contents). This is done usually by writing (often complex) scripts, which require programming skills from the authors.

To address this issue, we have created the timesheets.js library, a scheduler that allows HTML documents to be animated and synchronized in a purely declarative way. This work is based on the SMIL Timing and Synchronization module and the SMIL Timesheets specification, with a few extensions.

The library is implemented in JavaScript, which makes it usable in any browser. Authors can specify the dynamic behavior of HTML5 (+CSS3) documents. They can thus develop multimedia applications without writing a single line of JavaScript. Timesheets can also be used with other XML document languages, such as SVG for instance. This approach was validated in a class with students learning web multimedia.

6.1.2. Multimedia content adaptation

Multimedia documents may have to be played on multiple devices such as mobile phones, tablets, desktop computers, set-top boxes, etc. Usage and platform diversity requires documents to be adapted according to execution contexts, sometimes unpredictable at design time. In a joint work with project-team Exmo, we have designed a semantic adaptation framework for multimedia documents. This framework captures the semantics of document composition and transforms the relations between media objects according to adaptation constraints.

6.2. XML Processing

Mature results about XML processing were obtained along three main directions: the formalization and implementation for checking the impact of schema evolution on validation and queries; logical extensions supporting a notion of counting and the shuffle operator in trees; and the decision of a subtyping relation for a very expressive type algebra supporting a notion of polymorphism.

In addition, preliminary results were obtained on the definition of a rigorous logical framework for the static analysis of semantic web languages, on the static analysis of cascading style sheets, and on the equipment of an IDE with new static analysis features for XQuery.

6.2.1. Impact of XML schema evolution

In the ever-changing context of the web, XML schemas continuously change in order to cope with the natural evolution of entities they describe. Schema changes have important consequences. First, existing documents valid with respect to the original schema are no longer guaranteed to fulfill the constraints described by the evolved schema. Second, the evolution also impacts programs manipulating documents whose structure is described by the original schema.

We have proposed a unifying framework for determining the effects of XML Schema evolution both on the validity of documents and on queries [2]. The system is very powerful in analyzing various scenarios in which forward/backward compatibility of schemas is broken, and in which the result of a query may not be anymore what was expected. Specifically, the system offers a predicate language which allows one to formulate properties related to schema evolution. The system then relies on exact reasoning techniques to perform a fine-grained analysis. This yields either a formal proof of the property or a counter-example that can be used for debugging purposes. The system has been fully implemented and tested with real-world use cases, in particular with the main standard document formats used on the web, as defined by W3C. The system identifies precisely compatibility relations between document formats. In case these relations do not hold, the system can identify queries that must be reformulated in order to produce the expected results across successive schema versions.

6.2.2. Counting in trees

A major challenge of query language design is the combination of expressivity with effective static analyses such as query containment. In the setting of XML, documents are seen as finite trees, whose structure may additionally be constrained by type constraints such as those described by an XML schema. We have considered the problem of query containment in the presence of type constraints for a class of regular path queries extended with counting and interleaving operators [1]. The counting operator restricts the number of occurrences of children nodes satisfying a given logical property. The interleaving operator provides a succinct notation for describing the absence of order between nodes satisfying a logical property. We have proposed a logic supporting these operators, which can be used to solve common query reasoning problems such as satisfiability and containment of queries in exponential time [4].

6.2.3. Typing higher-order programs

We have considered a type algebra equipped with recursive, product, function, intersection, union, and complement types together with type variables and universal quantification over them. We have defined the subtyping relation between such type expressions, and have proved its decidability [9].

This has solved an open problem that was attracting a considerable research effort. The novelty, originality and strength of our solution reside in introducing a logical modeling for the semantic subtyping framework. We have modeled semantic subtyping in a tree logic and use a satisfiability-testing algorithm in order to decide subtyping. We have shown how the subtyping relation can be decided in EXPTIME. We have reported on practical experiments made with a full implementation of the system. This has provided a powerful polymorphic type system aiming at maintaining full static type-safety of functional programs that manipulate trees, even with higher-order functions, which is particularly useful in the context of XML.

6.2.4. Detection of inconsistent paths and dead code in XML IDEs

One of the challenges in web software development is to help achieving a good level of quality in terms of code size and runtime performance, for increasingly popular domain specific languages such as XQuery. We have presented an IDE equipped with static analysis features for assisting the programmer [8]. These features are capable of identifying and eliminating dead code automatically. The tool is based on newly developed formal programming language verification techniques, which are now mature enough to be introduced in the process of software development.

6.2.5. Static analysis of semantic web languages

We work with the Exmo project-team on the static analysis of semantic web languages such as RDF, OWL and SPARQL by investigating modal logics over graphs. We seek to build a rigorous logical reasoning framework based on $\mu$-calculus adapted for the web semantic languages [7] [11]. In particular, we studied the containment problem for SPARQL queries: determining whether, for any graph, the answers to a query are contained in those of another query. Our approach consists in encoding RDF graphs as transition systems and queries as $\mu$-calculus formulas and then reducing the containment problem to testing satisfiability in the logic.
6.2.6. Static analysis of style sheets

Developing and maintaining cascading style sheets (CSS) is an important issue to web developers as they suffer from the lack of rigorous methods. Most existing means rely on validators that check syntactic rules, and on runtime debuggers that check the behavior of a CSS style sheet on a particular document instance. However, the aim of most style sheets is to be applied to an entire set of documents, usually defined by some schema. To this end, a CSS style sheet is usually written w.r.t. a given schema. While usual debugging tools help reducing the number of bugs, they do not ultimately allow to prove properties over the whole set of documents to which the style sheet is intended to be applied.

We have proposed a novel approach to fill this lack [14] by analyzing CSS style sheets using the same logic and compile-time verification technique we use for other XML problems. We have developed an original tool based on our XML Reasoning Solver (see section 5.2). The tool is capable of statically detecting a wide range of errors (such as empty CSS selectors and semantically equivalent selectors), as well as proving properties related to sets of documents (such as coverage of styling information), in the presence or absence of schema information. This new tool can be used in addition to existing runtime debuggers to ensure a higher level of quality of CSS style sheets.

6.3. Multimedia Authoring

6.3.1. C2M project

The C2M project (see section 7.2.2) aims at developing industrial solutions that allow multimedia developers to achieve mass production with high quality results. It uses the SCENARI platform for document production and we have proposed a solution for dealing with multimedia content in such a framework [16]. Indeed, automatic tools are not always sufficient for generating high quality documents; manual editing of documents in their publishing format is often necessary to tune a number of details.

Our approach consists in providing a post-editing service to allow authors to adjust their multimedia presentations directly on the final form of documents. The first step is to provide a web rendering engine based on the latest advances in web standards, as described in section 6.1.1. The second step consists in designing web-aware authoring tools based on this library, thus providing authors with direct editing services for producing high quality multimedia documents while preserving the advantages of using an XML production workflow. We have developed a prototype of this authoring tool in which all editing templates are described with XUL (XML User interface Language) and XBL (XML Binding Language) elements that we have defined for handling time-based content and widgets (timeContainer, timeNode, timeLine, etc.).

With such a solution, we can combine two worlds: a semantic-oriented authoring approach, as provided by an XML document workflow, and a direct web-based editing system. The first guarantees homogeneous rendering while the latter enables direct adjustments on the final form of the document. Bridging these two worlds is made possible by using declarative web languages (namely HTML5, SMIL and CSS) and implementing their timing part in the browser (with the Timesheets.js library, see section 6.1.1). The authoring components are directly mapped to the document structures.

6.3.2. On-line editing of multimedia web content

In cooperation with EPFL (Lausanne) we have continued to explore the concept of template-driven editing for XML multimedia contents (see section 3.3.2). This year, we have carried out more experiments with very different types of contents, including structured documents, factual data, and multimedia objects [17].

These experiments have been done with the AXEL library developed by EPFL, based on our joint work on template languages. AXEL is an innovative client-side authoring tool that runs in the browser for editing XML documents, driven by an XTiger template. It allows average web users to easily edit XML content on web servers with no specific knowledge of XML. Our experiments have shown that the template-based approach significantly enhances the ability for web users to directly feed various applications with structured content.
6.4. Augmented Environments

A large part of the research on augmented environments specializes in the use of visual media. In WAM, we focus on the use of audio media and we put a strong emphasis on mobility.

We have developed the first indoor augmented reality audio navigation application running on personal AR devices such as mobile phones. The main idea behind the development of this application was a joint use of three concepts:

- Continuous localization by using embedded sensors together with physiological models of walking and assumptions about walking in structured indoor environments.
- OpenStreetMap Indoor Mapping used for map-aided positioning, assistive routing for visually impaired people, and environmental queries through audio panoramics.
- Guidance and navigation through AR audio, both 3D and environmental, with mixing of synthetic and natural sounds and support for timely audio information presentation.

We have demonstrated that these concepts are inter-dependent [12], and that bringing them together is a way to find new solutions to problems which are difficult to tackle when looking at them separately. These three concepts are implemented with web technologies we use XML languages and XML tools for interactive audio, building modeling, and personal navigation module configuration. This enables easy authoring of sound objects or audio icons used for building sonification [15], indoor navigation maps and panoramics, and walking models. Adaptability of navigation to preferences of people is based on the concept of audio stylesheets for OpenStreetMap data rendering, XML configuration of the Pedestrian Dead Reckoning module, and assistive routing specification.

We have developed two mobile browsers and a framework for generic navigation:

1. The Mixed Reality Browser [10] that we have developed can display PoI content either remotely through panoramics with spatialized audio, or on-site by walking to the corresponding place. MRB is the only browser of geolocalized data to use a declarative XML format for PoIs, panoramics, 3D audio and to be based on HTML5 both for the iconic and full information content of PoIs. MRB can be used for any kind of augmented reality visits. A cultural heritage visit of Grenoble (see section 7.1.2) has been realized with the tourist office of Grenoble and the CCSTI (Centre de Culture Scientifique Technique et Industrielle de Grenoble).

2. The Pedestrian Way Browser that we have developed can be used for indoor-outdoor navigation with assistive audio technology for visually impaired people on pedestrian ways with precise geospatial description. Its main characteristic is to be based entirely on the OpenStreetMap XML format for the representation of the route. We anticipate that in the context of the european project Venturi (see section 7.3.1), we will have a convergence of the MRB and PWB, allowing visually impaired people to undertake cultural heritage visit. An demonstration showing the use of the PWB in a structured outdoor environment is available online: http://www.youtube.com/watch?v=h2b8yfCauZ8

3. We have created an extensible client-server framework named TARA which allows navigation on an OpenStreetMap XML graph (indoor and outdoor) by computing routes in real-time. User preferences, like stairs versus lift, are supported through a ponderation of paths in the routing algorithm. The client is an HTML5 running in the browser on mobiles and desktop computers. It can therefore be used for simulation, to test or learn a route before the navigation in the real world. The user interface is based on three modalities, touch, audio and visual and can be operated by visually impaired people through VoiceOver using only touch and audio. Localization through embedded or external sensors is not mandatory as step by step instructions can be accessed through touch modality. The server is a full REST server (Sling-Apache) giving priority to the representation of geospatial resources and allowing environmental queries through the use of XQuery.
6. New Results

6.1. Data and Metadata Management

6.1.1. Uncertain Data Management

Participants: Reza Akbarinia, Patrick Valduriez, Guillaume Verger.

Data uncertainty in scientific applications can be due to many different reasons: incomplete knowledge of the underlying system, inexact model parameters, inaccurate representation of initial boundary conditions, inaccuracy in equipments, etc. For instance, in the monitoring of plant contamination, sensors generate periodically data which may be uncertain. Instead of ignoring (or correcting) uncertainty, which may generate major errors, we need to manage it rigorously and provide support for querying.

In [46], we address the problem of aggregate queries that return possible sum values and their probabilities. This kind of query which, we call ALL-SUM, is also known as sum probability distribution. The results of ALL-SUM can be used for answering many other type of queries over probabilistic data. In general, the problem of ALL-SUM query execution is NP-complete. We propose pseudo-polynomial algorithms that are efficient in many practical applications, e.g. when the agg attribute values are small integers or real numbers with small precision, i.e. small number of digits after decimal point. These cases cover many practical attributes, e.g. temperature, blood pressure, needed human recourses per patient in medical applications.

We have started to develop a probabilistic database prototype, called ProbDB (Probabilistic Database), on top of an RDBMS. ProbDB divides a query into two parts: probabilistic and deterministic (i.e. non probabilistic). The deterministic part is executed by the underlying RDBMS, and the rest of work is done by our probabilistic query processing algorithms that are executed over the data returned by the RDBMS. In [51], we demonstrated the efficient execution of aggregate queries with the first version of ProbDB.

6.1.2. Metadata Integration

Participants: Zohra Bellahsène, Rémi Coletta, Duy Hoa Ngo.

Due to the various types of heterogeneity of ontologies, ontology matching must exploit many features of ontology elements in order to improve matching quality. For this purpose, numerous similarity metrics have been proposed to deal with ontology semantics at different levels: elements level, structural level and instance level.

Elements level metrics can be categorized in three groups: (1) terminological, (2) structural and (3) semantic. Metrics of the first group exploit text features such as names, labels and comments to compute the similarity score between entities. Whereas metrics of the last two groups exploit the hierarchy and semantic relationship features. Our approach consists in first using terminological metrics. Then, during the matching process, mappings discovered by terminological metrics are used as input mappings to other metrics of the second and third groups. Obviously, the more precise results terminological metrics are, the more accurate results structural and semantic metrics have.

However, finding a good combination of different metrics is very difficult and time consuming. We proposed YAM++ (not Yet Another Matcher), an approach that uses machine learning to combine similarity metrics. Our main contributions are: the definition of new metrics dealing with terminological and context profile features of entities in ontologies [37], and the use of a decision tree model to combine similarity metrics [38].

To improve matching quality of YAM++, we exploit instances accompanying ontologies. We then apply similarity flooding propagation algorithm to discover more semantic mappings. At the 2011 competition of the Ontology Alignment Evaluation Initiative (http://oaei.ontologymatching.org), YAM++ achieved excellent results: first position on the Conference track and second position on the Benchmark track [39].
6.2. Data and Process Sharing

6.2.1. Social-based P2P Data Sharing

Participants: Hinde Bouziane, Michèle Cart, Esther Pacitti, Didier Parigot, Guillaume Verger.

This work focuses on P2P content recommendation for on-line communities. In [20], we propose P2PRec, a recommendation service for P2P content sharing systems that exploits users’ social data. Given a query, P2PRec finds peers that can recommend high quality documents that are relevant for the query. A document is relevant to a query if it covers the same topics. It is of high quality if relevant peers have rated it highly. P2PRec finds relevant peers through a variety of mechanisms including advanced content-based and collaborative filtering. The topics each peer is interested in are automatically calculated by analyzing the documents the peer holds. Peers become relevant for a topic if they hold a certain number of highly rated documents on this topic. To efficiently disseminate information about peers’ topics and relevant peers, we proposed new semantic-based gossip protocols. In our experimental evaluation, using the TREC09 dataset, we showed that using semantic gossip increases recall by a factor of 1.6 compared to well-known random gossiping. Furthermore, P2PRec has the ability to get reasonable recall with acceptable query processing load and network traffic. P2PRec was demonstrated in [31] and [47].

In [30], we exploit social relationships between users as a parameter to increase the trust of recommendation. We propose a novel P2P recommendation approach (called F2FRec) that leverages content and social-based recommendation by maintaining a P2P and friend-to-friend network. This network is used as a basis to provide useful and high quality recommendations. Based on F2FRec, we propose new metrics, such as usefulness and similarity (among users and their respective friend network). We define our proposed metrics based on users’ topic of interest and relevant topics that are automatically extracted from the contents stored by each user. Our experimental evaluation, using the TREC09 dataset and Wiki vote social network, shows the benefits of our approach compared to anonymous recommendation. In addition, we show that F2FRec increases recall by a factor of 8.8 compared with centralized collaborative filtering.

6.2.2. Satisfaction-based Query Replication

Participant: Patrick Valduriez.

In a large-scale Internet-based distributed, participants (consumers and providers) who are willing to share data are typically autonomous, i.e. they may have special interests towards queries and other participants’ data. In this context, a way to avoid a participant to voluntarily leave the system is satisfying its interests when allocating queries. However, participants’ satisfaction may also be negatively affected by the failures of other participants. Query replication can deal with providers failures, but, it is challenging because of autonomy: it cannot only quickly overload the system, but also dissatisfy participants with uninteresting queries. Thus, a natural question arises: should queries be replicated? If so, which ones? and how many times?

In [25], we answer these questions by revisiting query replication from a satisfaction and probabilistic point of view. We propose a new algorithm, called S b QR, that decides on-the-fly whether a query should be replicated and at which rate. As replicating a large number of queries might overload the system, we propose a variant of our algorithm, called S b QR+. The idea is to voluntarily fail to allocate as many replicas as required by consumers for low critical queries so as to keep resources for high critical queries during query-intensive periods. Our experimental results demonstrate that our algorithms significantly outperform the baseline algorithms from both the performance and satisfaction points of view. We also show that our algorithms automatically adapt to the criticality of queries and different rates of participant failures.

6.2.3. View Selection in Scientific Data Warehousing

Participants: Zohra Bellahsène, Rémi Coletta, Imen Mami.
Scientific data generate large amounts of data which have to be collected and stored for analytical purpose. One way to help managing and analyzing large amounts of data is data warehousing, whereby views over data are materialized. However, view selection is an NP-hard problem because of many parameters: query cost, view maintenance cost and storage space. In [36], we propose a new solution based on constraint programming, which has proven efficient at solving combinatorial problems. This allows using a constraint programming solver to set up the search space by identifying a set of views that minimizes the total query cost. We address view selection under two cases: (1) only the total view maintenance cost needs be minimized, assuming unlimited storage space (meaning that it is not a critical resource anymore); (2) both storage space and maintenance cost must be minimized. We implemented our approach and compared it with a randomized method (i.e., genetic algorithm). We experimentally show that our approach provides better performance resulting from evaluating the quality of the solutions in terms of cost savings. Furthermore, our approach scales well with the query workload.

6.2.4. Scientific Workflow Management

Participants: Ayoub Ait Lahcen, Eduardo Ogasawara, Didier Parigot, Patrick Valduriez.

Scientific workflows have emerged as a basic abstraction for structuring and executing scientific experiments in computational environments. In many situations, these workflows are computationally and data-intensive, thus requiring execution in large-scale parallel computers. However, parallelization of scientific workflows remains low-level, ad-hoc and laborintensive, which makes it hard to exploit optimization opportunities.

To address this problem, we propose in [23] an algebraic approach (inspired by relational algebra) and a parallel execution model that enable automatic optimization of scientific workflows. With our scientific workflow algebra, data is uniformly represented by relations and workflow activities are mapped to operators that have data aware semantics. Our workflow execution model is based on the concept of activity activation, which enables transparent distribution and parallelization of activities;

We conducted a thorough validation of our approach using both a real oil exploitation application and synthetic data scenarios. The experiments were run in Chiron, a data-centric scientific workflow engine implemented at UFRJ to support our algebraic approach. Our experiments demonstrate performance improvements of up to 226% compared to an ad-hoc workflow implementation. This work was done in the context of the Equipe Associée Sarava and the CNPq-INRIA project DatLuge.

In the context of SON, we also proposed a declarative workflow language based on service/activity rules [41]. This language makes it possible to infer a dependency graph for SON applications that provides for automatic parallelization.

6.3. Scalable Data Analysis

6.3.1. Massive Graph Management

Participant: Patrick Valduriez.

Traversing massive graphs as efficiently as possible is essential for many scientific applications. Many common operations on graphs, such as calculating the distance between two nodes, are based on the Breadth First Search (BFS) traversal. However, because of the exhaustive exploration of all the nodes and edges of the graph, this operation might be very time consuming. A possible solution is partitioning the graph among the nodes of a shared-nothing parallel system. However, partitioning a graph and keeping the information regarding the location of vertices might be unrealistic for massive graphs because of much inter-node communication. In [28], we propose ParallelGDB, a new graph database system based on specializing the local caches of any node in this system, providing a better cache hit ratio. ParallelGDB uses a random graph partitioning, avoiding complex partition methods based on the graph topology, that usually require managing extra data structures. This proposed system provides an efficient environment for distributed graph databases.

6.3.2. Top-k Query Processing in Unstructured P2P Systems

Participants: Reza Akbarinia, William Kokou Dedzoe, Patrick Valduriez.
Top-k query processing techniques are useful in unstructured P2P systems to avoid overwhelming users with too many results and provide them with the best ones. However, existing approaches suffer from long waiting times, because top-k results are returned only when all queried peers have finished processing the query. As a result, response time is dominated by the slowest queried peer. We proposed to revisit the problem of top-k query processing.

In [29] we address the problem of reducing user waiting time of top-k query processing, in the case of unstructured P2P systems with overloaded peers. We propose a new algorithm, called QUAT, in which each peer maintains a semantic description of its local data and the semantic descriptions of its neighborhood (i.e. the semantic descriptions of data owned locally by its direct neighbors and data owned locally by these neighbors direct neighbors). These semantic descriptions allow peers to prioritize the queries that can provide high quality results, and to forward them in priority to the neighbors that can provide high quality answers. We validated our solution through a thorough experimental evaluation using a real-world dataset. The results show that QUAT significantly outperforms baseline algorithms by returning faster the final top-k results to users.

6.3.3. Top-k Query Processing Over Sorted Lists

Participants: Reza Akbarinia, Esther Pacitti, Patrick Valduriez.

The problem of answering top-k queries can be modeled as follows. Suppose we have $m$ lists of $n$ data items such that each data item has a local score in each list and the lists are sorted according to the local scores of their data items. Each data item has an overall score computed based on its local scores in all lists using a given scoring function. Then, the problem is to find the $k$ data items whose overall scores are the highest. This problem model is a general model for top-k queries in many centralized, distributed and P2P applications. For example, in IR systems one of the main problems is to find the top-k documents whose aggregate rank is the highest wrt. some given keywords. To answer this query, the solution is to have for each keyword a ranked list of documents, and return the k documents whose aggregate rank in all lists are the highest.

In [16], we propose an extension of our best position algorithms (BPA) which had been proposed for top-k query processing over sorted lists model. The BPA algorithms have been shown to be more efficient than the well known TA Algorithm. We propose several techniques using different data structures for managing best positions that are crucial for efficient execution of top-k algorithms. We also provide a complete discussion on the instance optimality of TA algorithm (TA was considered so far as optimal over any database of sorted lists). We illustrate that, the existence of deterministic algorithms such as BPA shows that if we are aware of positions of seen data, then one of the main arguments used for proving the instance optimality of TA is invalidated. Therefore, in this case the proof of TA’s instance optimality is incorrect, and must be revisited.

6.3.4. Satellite Image Mining

Participant: Florent Masseglia.

Satellite Image Time Series (SITS) provide us with precious information on land cover evolution. By studying SITS, we can both understand the changes of specific areas and discover global phenomena that spread over larger areas. Changes that can occur throughout the sensing time can spread over very long periods and may have different start time and end time depending on the location, which complicates the mining and the analysis of series of images. In [45], we propose a frequent sequential pattern mining method for SITS analysis. Designing such a method called for important improvements on the data mining principles. First, the search space in SITS is multi-dimensional (the radiometric levels of different wavelengths correspond to infra-red, red, etc.). Furthermore, the non evolving regions, which are the vast majority and overwhelm the evolving ones, challenge the discovery of these patterns. Our framework enables discovery of these patterns despite these constraints and characteristics. We introduce new filters in the mining process to yield important reductions in the search space by avoiding consecutive occurences of similar values in the sequences. Then, we propose visualization techniques for results analysis (where modified regions are highlighted). Experiments carried out on a particular dataset showed that our method allows extracting repeated, shifted and distorted temporal behaviors. The flexibility of this method makes it possible to capture complex behaviors from multi-source, noisy and irregularly sensed data.
6.3.5. Distributed Approximate Similarity Join
Participant: Alexis Joly.

Efficiently constructing the KNN-graph of large and high dimensional feature datasets is crucial for many data intensive applications involving feature-rich objects, such as image features, text features or sensor’s features. In this work we investigate the use of high dimensional hashing methods for efficiently approximating the full knn graph of large collections, in particular, in distributed environments. We first analyzed and experimented what seems to be the most intuitive hashing-based approach: constructing several Locality Sensitive Hashing (LSH) tables in parallel and computing the frequency of all emitted collisions. We show that balancing issues of classical LSH functions strongly affect the performance of this approach. On the other side, we show that using an alternative data-dependent hashing function (RMMH), that we introduced recently [34], can definitely change that conclusion. The main originality of RMMH hash function family is that it is based on randomly trained classifiers, allowing to learn random and balanced splits of the data instead of using random splits of the feature space as in LSH. We show that the hash tables constructed through RMMH are much more balanced and that the number of emitted collisions can be strongly reduced without degrading quality. In the end, our hashing-based filtering algorithm of the all-pairs graph is two orders of magnitude faster than the one based on LSH. An efficient distributed implementation of the method was implemented within the MapReduce framework (and is the basis of the SimJoin prototype). This work is done in the context of the supervision of a PhD student working at INRIA Imedia (Riadh Mohamed Trad).

6.3.6. Visual objects mining
Participant: Alexis Joly.

State-of-the-art content-based object retrieval systems have demonstrated impressive performance in very large image datasets. These methods, based on fine local descriptions and efficient matching techniques, can detect accurately very small rigid objects with unambiguous semantics such as logos, buildings, manufactured objects, posters, etc. Mining such small objects in large collections is however difficult. Constructing a full local matching graph with a naïve approach would indeed require to probe all candidate query leading to an intractable algorithm complexity. In this work, we first introduce an adaptive weighted sampling scheme, starting with some prior distribution and iteratively converging to unvisited regions [35]. We show that the proposed method allows to discover highly interpretable visual words while providing excellent recall and image representativity. We then focused on mining visual objects on top of the discovered visual words. We therefore developed an original shared nearest-neighbors clustering method, working directly on the generated bi-partite graph. This work is in the context of the supervision of two PhD students, one working jointly with INA and INRIA and who will join the Zenith team next year (Pierre Letessier), one working at INRIA Imedia (Amel Hamzaoui).

6.3.7. Visual-based plant species identification from crowdsourced data
Participant: Alexis Joly.

Inspired by citizen sciences, the main goal of this work is to speed up the collection and integration of raw botanical observation data, while providing to potential users an easy and efficient access to this botanical knowledge. We therefore designed and developed an original crowdsourcing web application dedicated to the access of botanical knowledge through automated identification of plant species by visual content. Technically, the first side of the application deals with content-based identification of plant leaves. Whereas state-of-the-art methods addressing this objective are mostly based on leaf segmentation and boundary shape features, we developed a new approach based on local features and large-scale matching. This approach obtained the best results within ImageCLEF 2011 plant identification benchmark [48]. The second side of the application deals with interactive tagging and allows any user to validate or correct the automatic determinations returned by the system. Overall, this collaborative system allows to enrich automatically and continuously the visual botanical knowledge and thus to increase progressively the accuracy of the automated identification. A demo of the developed application was presented at the ACM Multimedia conference [33]. This work is done in collaboration with INRIA Imedia and with the botanists of the AMAP UMR team (CIRAD). It is also closely related to a citizen science project around plant’s identification that we develop with the support of the TelaBotanica social network.
6. New Results

6.1. Perception and autonomous navigation

Participants: Patrick Rives, Pascal Morin, Andrew Comport, Alexandre Chapoulie, Gabriela Gallegos, Cyril Joly, Maxime Meilland, Glauco Scandaroli.

6.1.1. Indoor SLAM: Self-calibration of the camera frame with respect to the odometry frame

Fusing visual data and odometry information is a standard technique to improve the robustness of the SLAM solution. Odometry data is considered as an input of the motion prediction equation while the visual data constitutes the filter observation. However, such method requires that the system is well calibrated: the pose of the camera frame with respect to the odometry frame has to be known. Usually, this pose is directly obtained by an hand made measurement yielding to incorrect values. We propose a new self calibration method to get these calibration parameters automatically. In practice, the state in the SLAM formulation is augmented with the unknown camera parameters (with respect to the odometry frame). This method requires to adapt a few Jacobians with respect to the original SLAM algorithm which assumes that these parameters are known. The accuracy and the stability of the estimation scheme clearly depends on the observability and the conditioning properties of the new system.

In 2010, we presented results in the case where the camera frame location has only 3 degrees of freedom (two translations and a rotation with respect to the vertical z axis). This year, these results were extended to the full calibration problem. As in the previous case, we assume that the robot is moving on a horizontal ground and observes 3D landmarks from the images delivered by the on board camera. The five parameters introduced by the calibration problem - 2 translations and 3 rotations (only 2 translations since the z component is not observable due to the planar motion of the robot) - are estimated simultaneously in addition to the “classical” SLAM parameters. The implementation of the algorithm is based on a Smoothing And Mapping (SAM) approach which computes a solution by considering the whole trajectory (instead of only the current pose as with the EKF approach).

As a theoretical result, we prove that this augmented system remains observable if and only if the curvature of the robot trajectory changes. This analysis was validated on real data with our indoor robot. Fig. 2 shows the mobile platform and the camera. It can be seen that an important rotational offset was added on the camera to test the capability to deal with large rotational values (the parameters are initialized with identity). Results are provided on Fig. 3 - 4 and table 1. They show that the trajectory and the map seem consistent; moreover, the algorithm was able to correct the odometry drift (green trajectory on Fig. 3). Then, the observability analysis was validated since the estimation of the camera frame parameters begins when there is a significant change in the radius of curvature of the trajectory (see the confidence bounds on Fig. 4). Finally, the estimation of these parameters was consistent with the ground truth (table 1). These results were presented at IROS’11 conference [30].

6.1.2. Outdoor Visual SLAM

Safe and autonomous navigation in complex outdoor urban-like environment requires a precise and real time localization of the robot. Standard methods, like odometry, typically performed by wheel encoders or inertial sensors, are prone to drift and not reliable for large displacements. Low cost GPS stations are inaccurate and satellite masking effect happens too frequently due to corridor-like configurations. We develop a real time and accurate localization method based on vision only without requiring any additional sensor.
Figure 2. Left: Robot used for the experiment – Right: The omnidirectional camera mounted with a rotational offset.

Figure 3. Red: trajectory and map provided by the algorithm — Green: odometry integration — Black cross: end of the trajectory — Blue: 99% confidence region for the last robot position.
Figure 4. Evolution of the curvature and the estimation of the camera parameters

Table 1. Numerical results concerning the camera parameters

<table>
<thead>
<tr>
<th>Camera param.</th>
<th>Reference</th>
<th>Estimation</th>
<th>3σ bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_x ) (m)</td>
<td>-0.16</td>
<td>-0.149</td>
<td>([-0.165, -0.132])</td>
</tr>
<tr>
<td>( t_y ) (m)</td>
<td>-0.35</td>
<td>-0.38</td>
<td>([-0.064, -0.013])</td>
</tr>
<tr>
<td>( \gamma_x ) (deg)</td>
<td>6.05</td>
<td>6.609</td>
<td>([6.276, 6.942])</td>
</tr>
<tr>
<td>( \gamma_y ) (deg)</td>
<td>-7.34</td>
<td>-7.714</td>
<td>([-7.889, -7.539])</td>
</tr>
<tr>
<td>( \gamma_z ) (deg)</td>
<td>0</td>
<td>0.668</td>
<td>([0.330, 1.007])</td>
</tr>
<tr>
<td>Final Pose</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x_{end} ) (m)</td>
<td>-0.06</td>
<td>-0.11</td>
<td>([-1.32, 1.09])</td>
</tr>
<tr>
<td>( y_{end} ) (m)</td>
<td>0.15</td>
<td>0.56</td>
<td>([-1.76, 2.89])</td>
</tr>
<tr>
<td>( \theta_{end} ) (deg)</td>
<td>0</td>
<td>-1.18</td>
<td>([-12.96, 10.58])</td>
</tr>
</tbody>
</table>
Our approach relies on a monocular camera on board the vehicle and the use of a database of spherical images of the urban scene acquired during an offline phase. This geo-referenced database allows us to obtain a robust “drift free” localization. Basically, the database is constituted of spherical images augmented by depth which are positioned in a GIS (Geographic information system). This spherical robot centered representation accurately represents all necessary information for vision based navigation and mapping ([26]). During the online navigation, the current vehicle position is computed by aligning the current vehicle camera view with the closest reference sphere extracted from the database.

A spherical augmented acquisition system has been developed and tested on our Cycab vehicle. This system is composed of six wide angle stereo cameras in overlap, which permits to extract depth information by dense correspondence. Since the depth is available, we are able to construct 360 degrees spherical images with a unique center of projection. Those 3D spheres are then used in an image-based spherical odometry algorithm to obtain the trajectory of the vehicle ([31]), fuse the spheres and construct the database.

During the online navigation, we consider a vehicle equipped with a simple camera (perspective, omnidirectional...). Here the aim is to register the current view on the closest sphere stored in the database. To achieve this we have developed a spherical image-based registration which allows efficient and accurate localization. But since the database of augmented visual spheres can be acquired under different illumination conditions than the online camera is experiencing, a robust algorithm combining model based localization and online visual odometry has been developed [32]. This method performs in real-time (45 Hz), and allows to handle large illumination changes and outliers rejection (see figure 5).

As a part of the ANR CityVIP project, the localization and mapping system has been successfully tested in Clermont Ferrand (France). A database of augmented images has been built along a learning trajectory. The aim was to automatically “replay” the learning trajectory using the database and a monocular camera. To avoid collisions and pedestrians, a laser was mounted on the front of the Cycab. The system was able to autonomously follow large scale trajectories (over 400 meters), in crowded urban environments (see figure 6).

Figure 5. Top left, robust outliers weights. Top center, augmented reference image. Top right, reference depth-map. Bottom left, intensity error after alignment. Bottom center, registered image. Bottom right, original current image.
6.1.3. Loop closure detection for spherical view based SLAM

Although more precise than the odometry computed from the wheels encoders, the visual odometry also suffers from the problem of drift when large displacements are performed. It is possible to correct this drift if the robot is capable to determine if the the place it is visiting has already been visited. re-observes a scene previously observed. This is often referred as the loop closure detection problem and several methods exist in the literature using perspective cameras. We develop new methods more reliable by exploiting the peculiar properties of spherical cameras.

Standard perspective cameras have a limited field of view leading to an incapability to encompass all the surrounding environment. This limitation of the field of view drastically limits the performances of visual loop closure algorithms. We propose to use spheres of vision computed by mosaicing images from 6 wide angle cameras mounted on a ring. Such a representation offers a full 360° field of view and keeps the spherical image invariant to the changes in orientations (Figure 7).

Loop closure detection can be exploited in a SLAM context at two levels: firstly, in the metric representation to retro-propagate along the robot’s trajectory the cumulative errors due to the drift, secondly, in the topological representation, to fusion in the graph representation the nodes corresponding to a same place.

Existing algorithms are not point of view independent: loop closures are detected uniquely when a place is revisited by the robot coming from the same direction but if the robot comes back in a different direction, the algorithms fail. Our solution relies on the presented spherical view and an efficient way of information extraction from it. We extract local information describing the points of interest of the scene. We enhance this local information with a global descriptor characteristic of the distribution of the points of interest over the sphere thereby describing the environment structure. These informations are used to retrieve the already visited places. Our algorithm performs well and is robust to the point of view variation [27]. This has led to an accepted paper at OMNIVIS 2011.
The figure 8 below presents obtained results. The trajectory is corrected (drift reduction) using the loop closure constraint. Red and green dots represent the loop closing places, they are linked by a red line.

6.1.4. Context-based segmentation of scenes

In a topological SLAM framework based on vision, the places are often represented by images gathered at regular time/distance intervals. It is nonetheless a meaningless representation in the context of topology. We
would prefer a definition like "in front of a building" or "entrance of the campus" instead of "image number i". Places are thus a set of images we need to group. This is what we call context based segmentation. In order to achieve this segmentation a criterion for "changing place" is needed, we propose to evaluate the environment structure using a global spatial descriptor (computed on the spherical view) called GIST. The algorithm relies on a statistical process control monitoring for an out-of-control signal involving a changing place event. The algorithm still needs to be improved for better robustness on the localization of the changing place events when we come back on previous visited paths.

The figure 9 presents the preliminary results. On the bottom left is the similarity matrix of the images GIST while on the bottom right is the segmented trajectory followed by our robot.

![Figure 9. Context-based segmentation of scenes](image)

### 6.1.5. Nonlinear observers for visual-inertial fusion with IMU-bias and camera-to-IMU rotation estimation

This work concerns the fusion of visual and inertial measurements in order to obtain high-frequency and accurate pose estimates of a visual-inertial sensor. While cameras can provide fairly accurate pose (position and orientation) estimates, the data acquisition frequency and signal processing complexity limit the capacities of such sensors in the case of highly dynamic motions. An IMU (Inertial Measurement Unit) can efficiently complement the visual sensor due to its high frequency acquisition, large bandwidth, and easy-to-process signals. IMU biases and calibration errors of the displacement between the camera frame and the IMU frame, however, can severely impair the fusion of visual and inertial data. Identification of these biases and calibration of this displacement can be achieved with dedicated measurement tools, but this requires expensive equipment and it is time consuming. We propose instead to address these issues via the design of observers. Last year, we had proposed a nonlinear observer to fuse pose and IMU measurements while identifying additive IMU biases on both gyrometers and accelerometers. We have extended this work to the self-calibration of the
rotation between the pose sensor frame (camera) and the IMU frame. Simulation and experimental results have confirmed that this calibration significantly improves the final pose estimation and allows to process motions with faster dynamics. This work has been presented at the IROS conference in October [35]. It is a joint work with G. Silveira from CTI in Brazil. We are currently extending this result to include the self-calibration of the translation displacement between both sensors.

6.2. Control of mobile robots

Participants: Claude Samson, Pascal Morin, Minh-Duc Hua [Post Doc, I3S, CNRS-Univ. of Nice-Sophia Antipolis], Daniele Pucci, Glauco Scandaroli, Luca Marchetti, Tarek Hamel [Univ. of Nice-Sophia Antipolis].

6.2.1. New developments of the Transverse Function control approach

We are pursuing the development of the Transverse Function control approach for highly nonlinear systems via the application of the approach to challenging mechanical systems with various structural control properties.

6.2.1.1. Control of a redundant wheeled snake mechanism using transverse functions on \( SO(4) \)

The Transverse Function approach is applied to the control of a nonholonomic three-segments/snake-like wheeled mechanism, similar to the planar low-dimensional version of Hirose’s Active Cord Mechanism (ACM) previously studied [65], but with two additional internal degrees of freedom (d.o.f.) whose actuation yields more flexible and efficient control solutions (see figure 10). From a theoretical point of view, these complementary d.o.f. modify the Control Lie Algebra of the system so that only first-order Lie brackets of the control vector fields are needed to satisfy the Lie Algebra Rank Condition (LARC). The fact that four independent (angular velocity) control inputs are used also implies for this system the existence of Transverse Functions (TF) defined on the six-dimensional special orthogonal group \( SO(4) \). Several examples of mechanisms whose control involve TF defined on \( SO(3) \) have been pointed out in the past [54], [64], [65]. Beyond the specific control problem addressed here, a motivation for the present study is to illustrate for the first time how functions defined on the larger set \( SO(4) \) can be determined and used for the control of a physical system. This study is complemented with recalls concerning the parametrization of \( SO(4) \) by pairs of isoclinic quaternions and with the derivation of complementary differential calculus relations associated with this parametrization. The results will soon be submitted for presentation at an international conference.

6.2.1.2. Control of three hooked vehicles with off-axle hitches

An extension of the study [65] performed last year on Hirose’s Active Cord Mechanism (ACM) concerns the case when one of the wheeled-trains (the middle one, for instance) possesses actuators giving it tracing and rotating capacities (alike a unicycle-like vehicle), while the other two vehicles are passively hooked to this tracing vehicle. This type of actuation departs from the one of Hirose’s Active Cord Mechanism for which the tracing capacity of the mechanism relies exclusively on the deformation of the system of vehicles via the control of the inter-connecting angles, and it makes an important difference at the control level. This system may also be seen as a unicycle-type vehicle with two trailers and off-axle hitches. Unlike the simpler hitch-on-axle case commonly addressed in the literature, this system is not differentially flat and “complete” feedback solutions ensuring practical stabilization of any, feasible or non-feasible, trajectory remained an open issue. This actuation allows for the complete alignment of the three vehicles without going through actuation singularities, and for the asymptotic tracking of a reference frame moving along a straight line or a circle. On the other hand, in order to fully take advantage of the extra possibilities offered by it, one has to consider higher-order Lie bracket maneuvering motions that significantly complicate the feedback control design. The Transverse Function approach is applied using the fact that a dynamic extension of this two-control-inputs system is left-invariant on a 6-dimensional Lie group. Transverse functions calculated as the group product of “elementary” functions defined either on toruses or on \( SO(3) \), and yielding feedback controls ensuring asymptotic stabilization of “feasible” reference trajectories under common ”persistent excitation” properties (as in the case of classical feedback control solutions based on a linear approximation of the associated tracking error system) are proposed. As usual, the superiority of the transverse function solution over more classical solutions comes from that it also applies to the case of non-feasible reference trajectories for which (practical) stabilization involves complex maneuvers. The results of this study will be submitted next year for presentation at an international conference.
Figure 10. Three segments snake robot with two steering wheels. View from above
6.2.1.3. Control of an extended trident-snake vehicle

This study is part of a thesis work on the control of non-standard nonholonomic mobile robots by W. Magiera under the dual supervision of Prof. K. Tchon (Wroclaw University of Technology) and C. Samson. This collaboration involves several long term visits of the PhD student at INRIA, starting this year (2 months), and for the next two years. This year’s objective is to address a particularly challenging control problem and evaluate the possibilities offered by the Transverse Function approach to solve it. The system under consideration is based on the "common" trident snake mechanism \cite{54} complemented with one, two, or three additional "passive" wheeled extensions, each of them subjected to the rolling-without-slipping constraint (see figure 6.2.1.3). Transverse Functions solutions tested so far involve a mixt (product) of functions defined either on the torus, or on special orthogonal groups, and future improvements may involve the search for new transverse functions.

![Figure 11. Trident-snake mechanism with passive extensions. View from above](image)

6.2.2. Control of aerial vehicles

6.2.2.1. Vehicles subjected to lift forces

The development of a general theory for the control of underactuated (ground, marine, and aerial) vehicles whose main propulsion relies on a thrust force exerted in a single (vehicle’s related) direction was continued
this year. Part of this program, more specifically devoted to aerial vehicles, is the subject of D. Pucci’s thesis research project. This year’s focus was the prolongation of the work initiated last year on the modelling of lift forces and on their effects on the flight and control of aerial vehicles. Among the new results obtained on the subject, an extension and generalization of a previous feedback control strategy developed for spherically-shaped vehicles only subjected to drag forces [52], based on an "ideal" generic model of lift and drag forces associated with bi-symmetric wings, has been accepted for presentation at an international conference ([34]). The proposed solution involves a change of thrust control input in order to render the dynamics of the transformed system independent of the angle of attack associated with the vehicle’s main wing. A weakness of the aforementioned model is that it does not account for the so-called stall phenomenon, which is an abrupt loss of lift when the angle of attack increases beyond a certain value called stall angle. Taking it into account adds considerable complexity to the vehicle’s dynamics, especially in the case of a vehicle moving within a fluid endowed with a large Reynolds number for which the stall phenomenon can no longer be neglected. We showed that, although this phenomenon never forbids the existence of an attitude equilibrium given a desired reference velocity, the uniqueness of this equilibrium is not always granted. As a consequence, modifications of the desired velocity may result in the abrupt disappearance of an equilibrium so that the asymptotic stabilization of a desired velocity profile may become an ill-conditioned problem. To avoid this complication a possibility consists in characterizing "good" velocity profiles—associated, for instance, to transition maneuvers between hovering and high-velocity cruising— for which the existence of continuously changing equilibria is ensured. First results on this topic and research direction have been submitted for presentation to an international conference.

6.2.2.2. Nonlinear control of VTOL UAVs with uncertain position measurements

This work concerns the feedback control of VTOL UAVs (Vertical Take-Off and Landing Unmanned Aerial Vehicles). The objective is to asymptotically stabilize a reference equilibrium configuration with a "semi-global" convergence domain, i.e global convergence domain in position and semi-global in orientation, knowing that a global convergence domain in orientation cannot be obtained with continuous feedback laws due to the topology of the rotation space $\mathbb{SO}(3)$. Several solutions to this problem have been proposed in the past years, under the assumption that the pose (i.e. position and orientation) is completely known. This work concerns the case when the relation between the "position measurements" and the true position vector is uncertain. In practice, such uncertainties are related, e.g., to ill-calibrated sensors or to incomplete knowledge of the environment in the case of proximetry sensors. It is assumed that position measurements are given by $\overline{p} = Mp$ where $p$ is the true position error with respect to the reference position, and $M$ is an unknown invertible matrix. As a first contribution, we propose a class of feedback laws that achieve semi-global stability of the equilibrium $p = 0$ for any matrix $M$ that satisfies the stability criteria $\|M - I_3\| < \delta(k)$ where $I_3$ is the $3 \times 3$ identity matrix, $\delta$ is a strictly positive function, and $k$ is the vector of control parameters. An explicit expression of the function $\delta$ is provided, thus relating the control parameters to the stability margin. The second contribution of this work is the application of this control approach to the visual servoing of VTOL UAVs with respect to a planar vertical structure (wall, etc). From the homography matrix that relates the current camera image to a reference image (taken at the reference pose), we derive a signal output of the form $\overline{p} = Mp$. The matrix $M$ typically depends on unknown parameters but we show that a very rough knowledge of them is sufficient to design a stable controller based on the above-mentioned stability criterion. These results have been submitted for publication at an international conference. This is a joint work with H. de Plinval and P. Mouyon from ONERA Toulouse.

6.2.3. Development of an autonomous shopping cart

This work, which consists in developing a shopping cart with autonomy capabilities (automatic user following, obstacle avoidance, etc), is part of the national INRIA PAL project (Personally Assisted Living) which aims at developing robotic tools for disabled persons or elderlies.

The architecture of “Autonomous Shopping Cart” has been developed in three layers. The first one is responsible for connecting the services layer to physical (or virtual) devices. During this year, all necessary components to access the devices have been implemented:
the Phidgets library wraps the API of Phidgets devices and abstracts the access to the peripherals on the wheelchair robot;
the Hannibal library wraps the interface to access the Hannibal robot (through Carmen library);
the Simulator library wraps the simulator interface.

All of them expose a common interface to the software modules. Thus, the higher components do not have to be changed if the test platform changes.

The second level is the core of this year’s work. It is composed by the Control module and Modeling module. The Control module aims at stabilizing the trajectory of the robot w.r.t a given reference motion. In practice, this reference motion corresponds to the cart user that needs to be followed, but it could be any virtual reference motion. A first implementation of the Control module has been made using Matlab software. The result is a Control library that contains two different methods for controlling the trajectory:

- position control only;
- full-state control (position+orientation).

These methods have been implemented in Matlab language and then converted in C++. The resulting library has been utilized within the Control module deployed on the robotic platform.

The main objective of the Modeling module is to detect the cart user within the sight of the sensors. This task is generally non-trivial, due to noise in the sensor signals and variations of the environment. For this reason a Multiple Hypothesis Tracker has been used to allow for the presence of several persons in the environment. The method uses the laser scans to extract potential persons and then a Selector algorithm extracts the best hypothesis for the cart user. This hypothesis is then converted into a virtual reference point given to the Control module for trajectory tracking.

The third layer is represented by the Behavior module. This component manages the other modules, starts and stops services on request, enables the initialization procedure and so on. As for now, it starts all modules and initiates the starting procedure. In particular, it selects the first person to be tracked, among possible candidates. Experiments have been successfully conducted both on the mobile robot Hannibal and on the wheeled walking aid ANG (Assistive Navigation Guide) developed by the EPI Coprin.
6. New Results

6.1. Robotics

6.1.1. Calibration and identification

6.1.1.1. Calibration of a cable-driven robot

Participants: David Daney, Julien Alexandre dit Sandretto, Jean-Pierre Merlet, Gilles Trombettoni.

To improve the accuracy of a cable manipulator, it is necessary to identify the uncertainties of its model. The robots, studied in Cogiro, an ANR National initiative, are redundantly actuated: the number of powered wires is larger than the number of degrees of freedom of the manipulator. Under some cable properties hypothesis, this over-constraint mechanism allows to perform a self-calibration - i.e. the identification of the parameters does not need additional external measurement. A first experimentation, done in Montpellier, validated a novel approach which consists in a simultaneous identification of parameters and robot position (unknown in self-calibration process).

6.1.1.2. Cable properties

Participants: Julien Alexandre dit Sandretto, Gilles Trombettoni, David Daney.

The majority of researches done on cable-driven robot modeling need to take into account that the mass and the elasticity of wires are neglectable. However, they can not prove that these hypotheses are acceptable regarding these objectives. We have proposed an algorithm based on interval analysis to judge the validity of these assumptions for a cable-robot in a specific workspace. This method have been tested on the Tecnalia/LIRMM’s prototype and used for the construction of the Cogiro robot.

6.1.1.3. Optimal calibration poses of a 3-RPR planar parallel robot

Participant: David Daney.

The choice of the measurement configuration is crucial to improve the robustness of the calibration for measurement uncertainties. This year, a geometrical approach has been used to determine formally the set of the optimal poses for the identification of the kinematic parameter of a 3-RPR planar parallel robot. This result is important because it explains the influence of the location of some particular robot poses in their workspaces during the model identification process. A generalization is explored to construct automatically an optimal set for robot calibration and moreover, to improve experimental design algorithms. The aim is now to obtain similar results for cable driven robots calibration.

6.1.1.4. Geometric calibration of a space telescope

Participants: Thibault Gayral, David Daney, Jean-Pierre Merlet.

In October 2010 begun a collaborative work with Thales Alenia Space on the calibration of the mechanical structure of a space telescope. Its architecture is based on a parallel manipulator (type active wrist 6-PUS) used to correct the relative position of two mirrors. The aim is to identify the parameters of this robot, to improve its accuracy and then increase the quality of the images provided by the telescope. Thus, a geometric calibration procedure was considered and a campaign of photogrammetry was performed on the telescope. Using a kinematic description, a final accuracy of at worst 10 \(\mu m\) was reached on the position of the platform of the telescope. The aim of these measures were also to valid or not the model of comportment of the flexible parts of the device. This campaign brought to light the necessity to consider forces and torques acting on the structure in the deformation of the flexible parts in order to reach a submicrometric accuracy.

6.1.1.5. Modelization of flexible articulations of the telescope

Participants: Thibault Gayral, David Daney, Jean-Pierre Merlet.
In order to improve the final accuracy of the above-mentioned space telescope, a novel model including the statics equations in order to calculate deformations of the flexible articulations is currently under study. The main difficulty is to identify parameters (stiffness matrix and geometric parameters) that have different units and are not of the same order of magnitude. To solve this issue, we are focusing our effort to write the problem in a better robust form.

6.1.1.6. Interval Identification

Participants: David Daney, Julien Alexandre dit Sandretto, Gilles Trombettoni.

There are many approaches to identify the parameters of a model. In most cases, it consists in providing a particular solution of an over-constraint set of equations which must be robust to measurement to errors: in least square sense, with some statistical properties.... However, the interpretation and the validity of the result can be difficult and prone error. We propose to investigate some interval approaches in order to associate to the result some information and a certification of solutions.

6.1.2. Rehabilitation and biomechanics

Participants: Sami Bennour, David Daney, Mandar Harshe, Jean-Pierre Merlet [correspondant].

The focus of the work is on analyzing knee joint motion during a walking activity. The measurement system is based on the wire actuated parallel robot architecture. To increase the reliability of our analysis, and decrease the influence of Skin Tissue Artifacts (STA), we also incorporate a passive wire measurement system, IR camera based motion capture system, accelerometers, and force sensors to measure human motions.

The main principle of the system is to observe relative motions of the collars attached to tibia and femur. These are connected to the base by wires and also hold the other sensors. Measurements in the global frame and collar specific local frames give precise data to reconstruct collar (and thus, knee joint) motion.

Over the past year we have finalized the experimental setup, by calibrating the collars and the sensor systems, and adapting the existing wire robot system (MARIONET-REHAB) to work along with the other sensors. The software developed uses a single unified input file to specify all sensor configurations, streamlining experiments. We performed our preliminary experimental trials for walking motion on three subjects using the wire sensors, accelerometers and optical motion capture system.

We began work on processing the data obtained from these trials. Post-processing functions have also been developed to calculate additional collar properties, perform sensor data processing (filtering, noise removal and estimation) and access files in the C3D file format, which is used a binary file format used by the motion capture system.

The main challenge we are working on is to perform sensor data fusion and increase reliability of results. For this we must identify parameters that correlate the different sensor measurements and perform error analysis. Possible solutions include using interval analysis methods to address the uncertainties.

6.1.3. Kinematics of wire parallel robots

Participant: Jean-Pierre Merlet.

The kinematics of wire robot is a complex problem because a solution is possible only if the tension in the wire is positive. Hence the static equilibrium has to be taken into account. This problem is not well addressed in the literature. Curiously the forward kinematics of robot (i.e. finding the possible poses of the platform for given wire lengths) with at least 6 wires is straightforward: the distance equations allows to determine all poses and then we use the static equations to calculate the wire tensions and discard the one having at least one negative tension. For robot having less than 6 wires we have to consider simultaneously the distance equations and the static equations in order to get a square system (of \( n + 6 \) equations for a \( n \)-wires robot). We have investigated the case of a 3-wires robot with all wires attached at different points on the platform and have shown that all solutions can be computed provided the solving of an univariate polynomial of degree 158 \([17]\). Although we are not able to guarantee that the degree of this polynomial cannot be decreased, we believe that nevertheless the order will be too high for robust determination of the solutions and can only be used to determine an upper bound for the maximal number of solutions. We have also investigated theoretically and experimentally the
kinematics of a $n$-wires ($n \geq 4$) robot with all wires attached at the same point (i.e. only the position of the center of the platform can be controlled). Although this robot is apparently redundant, we have shown that in any pose at most 3 wires will be simultaneously under tension and therefore that the redundant wires cannot be used to control the wire tensions.

As the wire length measurements are not sufficient to determine the current pose of the platform (which is necessary for control purposes) we are investigating the use of additional sensors. Our prototypes MARIONET-ASSIST and MARIONET-VR are instrumented to measure wire directions, but with a large uncertainties. We have started a theoretical investigation to determine under which conditions these uncertainties may lead to a non-unique solution and we will validate the results on the two prototypes.

6.1.4. Rehabilitation robots for the immersive space

Participants: Michael Burman, Jean-Pierre Merlet.

The on-site immersive room provides 3D visualization but is lacking of haptic feedback and motion capabilities. We plan to implement in this room a movable system, constituted of:

- a 6 degrees-of-freedom motion base: the motion system 710-6-500-220 by Servos Simulation Inc. has been selected and is now operational and fully calibrated. If necessary the user may stand on this motion base
- the MARIONET-VR wire-driven parallel robot: this robot uses the same actuation principle than the MARIONET-REHAB robot (linear actuator with a pulley system for coiling and uncoiling of the wires), but is able to lift a person. The prototype is basically functional but its installation in the immersive room has been delayed because of lack of appropriate fixing elements

The full system will be installed in the immersive room at the beginning of 2012.

6.1.5. Assistance robotics

Participants: Michael Burman, David Daney, Jean-Pierre Merlet.

As mentioned earlier in the report we have started in 2008 a long term strategic move toward assistance robotics, with the objectives of providing low-cost, simple to control, robotized communicant devices that may help disabled, elderly and handicapped people in their personal life, with the credo that they have to be adapted to the end-user and to its everyday environment (by contrast with the existing trend of focusing on a “universal” robot, to which the end-user and its environment have to adapt) [18], [14], [21]. We have started last year the development of a simulated flat in order to explore various full scale scenarii that cover a part of the daily life of an elderly, to develop specific assistance devices and to test them. We describe in the following sections several devices that have been developed/improved during this year1. Note that our demonstration in assistance are highlighted during the visit of Sophia (275 visitors have attended our demonstration during 14 visits) and have received serious press coverage (5 papers, 2 TV interviews).

6.1.5.1. Walking aids

Wheeled walking aids are usually the first tools that are used when motricity problems occur. We are developing the family of robotized Assistive Navigation Guide (ANG), which are based on commercially available Rollators, with several objectives (we mention only a few of them):

- fall prevention/detection: fall is a major problem for elderly (it is estimated that fall is the main cause of 10 000 elderly deaths per year in France).
- mobility help: provide an on-demand mobility help
- gait pattern monitoring: we believe that being able to monitor the trajectory of the walking aid will provide useful information on the gait pattern of the user

1 pictures of this assistive flat are available at http://www-sop.inria.fr/coprin/developpements/main.html
For reaching these objectives we have developed two walking aids:

- **ANG-light**: a walking aid with encoders in the wheels, 3D accelerometer, gyrometer. These sensors allow to measure the trajectory of the walking aid and several features of the user’s gait (step pattern, gait symmetry,...). ANG-light has been tested by the CHU of Nice-Cimiez that was willing to perform an in-depth investigation of its use. For that purpose we have asked in September 2009 for the necessary formal authorization to the local CPP, which has been granted only in December 2011. To prepare this study we have organized a large scale experiment at INRIA, where 24 users were asked to perform the trajectories of the protocol twice, with and without the aid. When not using the aid the users were equipped with 3D accelerometer on the wrists and knees and were using specific shoes with force sensors in the sole. Initial analysis of the records shows that indeed we are able to obtain significant information on the gait pattern, that are not available using the existing tools, and detect differences in the gait pattern for user having even a light pathology in the lower limb. The experiment with elderly patients at CHU will take place in January 2012.

- **ANG-II**: this aid is an evolution of the motorized walker ANG, with a lower weight and better integration.

### 6.1.5.2. MARIONET-ASSIST

This wire-driven parallel robot is installed in the ceiling of the flat. It has been used this year in the 4-1 configuration (4 wires attached at the same point), which allows for controlling the position of the platform, but not its orientation. Several platforms have been developed, all of them incorporating a webcam and allowing for a free rotation around the vertical axis, while an accelerometer measure the tilt angle of the platform (which is used to determine in which direction the end-user is willing to move). One of the platform incorporate a 4 d.o.f. robot that may grasp light object (one of our objective is to use also the robot as a manipulator for bringing object back to the user in a more or less autonomous way, which is the subject of the PhD thesis of R. Ramadour).

We have shown that the the robot can be used for sit-to-stand transfer and for lifting handicapped people. A specific attention has been devoted to propose very simple control interface: joystick, remote TV set, control box whose tilt determine the motion axis.

### 6.1.5.3. Other flat equipments

Our scenario includes the management of emergency situations such as the fall. Fall detection can be performed by the ANG walkers but we have also started investigating the inclusion of fall detection system in the clothes of the end-user either through a GEO-300 devices or by incorporating an Arduino Lilypad processor. When a fall is detected indoor an alert is transmitted to a coordinator (a Nabaztag) which will order the walker and the MARIONET-ASSIST robot to move close to the user to provide a support. At the same time two mobile robots will converge to the same location: a remote-controlled, webcam equipped ROVIO (which can provide images of the end-user to a rescue center) and a Pekee II, that we will equip to provide first aid.

An important point in assistance is to be able to have at all time a rough idea of the localization of the patient. Although we plan to use a Kinect for that purpose, we will also investigate the use of non-vision sensors (which are much less intrusive and therefore can be more easily accepted) such as RFID tag (ANG-II has a RFID tag reader), directive distance sensors and light barriers.

Another axis for assistance is to reduce the risks of fall by using the principle that the objects has to come to the hand of the user (or of the robot), not the opposite. This implies instrumenting the environment with drawer openers and doors manipulation and we have started implementing them on drawers and on the fridge of the flat.

### 6.2. Interval analysis

#### 6.2.1. Inner Regions and Interval Linearizations for Global Optimization

**Participants:** Gilles Trombettoni [correspondant], Bertrand Neveu.
Researchers from interval analysis and constraint (logic) programming communities have studied intervals for their ability to manage infinite solution sets of numerical constraint systems. In particular, inner regions represent subsets of the search space in which all points are solutions. Our main contribution is the use of recent and new inner region extraction algorithms in the upper bounding phase of constrained global optimization. Convexification is a major key for efficiently lower bounding the objective function. We have adapted the convex interval taylorization proposed by Lin & Stadtherr for producing a reliable outer and inner polyhedral approximation of the solution set and a linearization of the objective function. Other original ingredients are part of our optimizer, including an efficient interval constraint propagation algorithm exploiting monotonicity of functions.

We end up with a new framework for reliable continuous constrained global optimization. This interval Branch & Bound significantly outperforms the best reliable global optimizers [22], [25], [28].

6.2.2. An Interval Extension Based on Occurrence Grouping

Participants: Bertrand Neveu [correspondant], Gilles Trombettoni.

We proposed last year a new “occurrence grouping” interval extension \([f]_{og}\) of a function \(f\). When \(f\) is not monotonic w.r.t. a variable \(x\) in a given domain, we try to transform \(f\) into a new function \(f^{og}\) which is monotonic w.r.t. two subsets \(x_a\) and \(x_b\) of the occurrences of \(x\): \(f^{og}\) is increasing w.r.t. \(x_a\) and decreasing w.r.t. \(x_b\). \([f]_{og}\) is the interval extension by monotonicity of \(f^{og}\) and produces a sharper interval image than the natural extension does.

This year we have improved the linear program and algorithm that minimize a Taylor-based over-estimate of the image diameter of \([f]_{og}\). We have detailed the proofs of correctness and reliability of this occurrence grouping algorithm [8], [29].

6.3. Miscellaneous results

6.3.1. Equilibrium strategies for linked Electricity and CO2 markets

Participant: Odile Pourtallier.

In collaboration with M. Bossy (INRIA -TOSCA Team) and N. Maïzi (CMA - Mines Paristech) O. Pourtallier the study of equilibrium model for coupled electricity and CO2 allowance exchange markets has been pursued. (see also Section 7.1). A static equilibrium model has been studied under various assumptions on the CO2 market design. All the CO2 market designs do not lead equilibrium, which interferes on the (short term day ahead) electricity market, which in turn interferes on the electricity mix and consequently on the total emission. Together with El-Hadj Dia (INRIA -TOSCA Team) we have also pursued an indifference pricing methodology which is presented in more details in INRIA -TOSCA Team section.

6.3.2. Symbolic tools for modeling and simulation

Participant: Yves Papegay.

This activity is the main part of a long-term ongoing collaboration with Airbus whose goal is to directly translate the conceptual work of aeronautics engineers into digital simulators to accelerate aircraft design. An extensive modeling and simulation platform has been designed which includes a dedicated modeling language for the description of aircraft dynamics models in term of formulae and algorithms, and a symbolic compiler producing as target an efficient numerical simulation code ready to be plugged into a flight simulator, as well as a formatted documentation compliant with industrial requirements of corporate memory.
Implementation of this platform is a modeling and simulation environment based on symbolic computation tools. It contains several components:

- a model editor, that makes it possible and easy to enter the whole set of equations describing large and complex industrial models,
- an highly interactive and modular evaluation workbench allowing to simulate the models and to visualize the results inside the modeling environment with the benefits for the designer of being able to directly use all its computational functionalities.
- a C code generator which, using these models, automatically generates the numerical real-time simulation engines
- a technical documentation generator

During the year 2011 the technology demonstrated by our prototype has been transferred to our industrial partner. A lot of work has been done on our modeling and simulation environment to improve its robustness and its development level of quality toward industrial standards. Final version of our prototype is to be delivered to Airbus at the end of the year.

6.3.3. Multi-agent aircraft design

Participant: Yves Papegay.

The modeling environment described in the previous section is used, in collaboration with other teams at Airbus, in the framework of the ID4CS project founded by ANR and dedicated to multi-agent optimization of large scale system. Several models of aircraft engines and of aircrafts have been developed as user cases for the project. Automatic generation of extended models namely computing first order derivatives of the original models has been implemented.
5. New Results

5.1. Dynamic World Perception and Evolution Prediction

5.1.1. Environment modeling and sensor data acquisition


An overall architecture of our environment-modeling module with the inputs from heterogenous sensors is shown in Fig. 6. The combined use of two lidars and stereo-vision helps mitigate uncertainty and allows for detection of partially occluded objects. The data processing includes the computation of probabilistic occupancy grids for each sensor and their subsequent fusion with the Bayesian Occupancy Filter (BOF). The output of the module is an estimation of the position, velocity and associated uncertainty of each observed object, which are used as input to the risk assessment module.

![Figure 6. Architecture of the environment modeling module.](image)

This architecture is implemented on our experimental platform, a Lexus LS600h car shown in Fig. 7. The vehicle is equipped with a variety of sensors including two IBEO Lux lidars placed toward the edges of the front bumper, a TYZX stereo camera situated behind the windshield, and an Xsens MTi-G inertial sensor with GPS.

The stereo camera baseline is 22 cm, with a field of view of 62°. Camera resolution is 512x320 pixels with a focal length of 410 pixels. Each lidar provides four layers of up to 200 impacts with a sampling period of 20 ms. The angular range is 100°, and the angular resolution is 0.5°. The on-board computer is equipped with 8GB of RAM, an Intel Xeon 3.4 GHz processor and an NVIDIA GeForce GTX 480 for GPU. The observed region is 40 m long by 40 m wide, with a maximum height of 2 m. Cell size of the occupancy grids is 0.2x0.2 m.
Figure 7. Lexus LS600h car equipped with two IBEO Lux lidars, a TYZX stereo camera, and an Xsens MTi-G inertial sensor with GPS.

The Lexus experimental platform provides to acquire sensor data in real traffic environments: eight layers of laser scans, stereo images, IMU data (accelerations), velocity, GPS position, steering angle. The experiments are conducted in various road environments (country roads, downtown and highway), at different time of the day, with various driving situations (light traffic, dense traffic, traffic jams). The datasets are acquired online and are used for testing of our sensor fusion and risk assessment algorithms.

5.1.2. Bayesian fusion of visual and telemetric information

Participants: Igor Paromtchik, Christian Laugier, Mathias Perrollaz, Amaury Nègre.

5.1.2.1. Concept of BOF and obstacle detection in occupancy grids

Obstacle detection is a widely explored domain of mobile robotics. It presents a particular interest for the intelligent vehicle community, as it is an essential building block for Advanced Driver Assistance Systems (ADAS). In the ANR project LOVe (Logiciel d’Observation de Vulnérables) and ArosDyn project, the e-Motion team proposed to perform obstacle detection within the occupancy grid framework. In order to work efficiently with occupancy grids, we have previously developed a probabilistic framework with the Bayesian Occupancy Filter (BOF) [40] [88] (patent 0552736 (2005)), which provides filtering, data fusion, and velocity estimation capabilities while allowing for parallel computation. The Fast Clustering and Tracking Algorithm (FCTA) [73] is then used to identify and track individual objects. The BOF is designed with the intent of its implementation in hardware as a system-on-chip. Like other grid-based approaches, the BOF framework performs sensor fusion at the cell level [40]. The BOF evaluates probabilities of both cell occupancy and cell velocity for each cell in a four-dimensional spatio-temporal grid. The monitoring of traffic scenes includes detection and tracking of objects by the FCTA [73].

Fig. 8 shows examples of occupancy grid mapping with the proposed approach. The arrows indicate the pedestrian, the car, and the bicycle, which appear in the camera images and the occupancy grids. Because the accuracy of stereo-vision tends to become poor at large distance, the corresponding grid has been attenuated beyond 20 m and the system is tuned to give more confidence to the lidars than to the stereo-vision. One of advantages of sensor fusion is a larger viewfield so that the vehicles overtaking the ego-vehicle (they are not seen in the camera images) are correctly mapped on the resulting BOF grid. Moreover, the sensor fusion as well as the Bayesian estimation provide to filter out the laser impacts with the road surface, e.g. right lidar in Fig. 8.

Note that a large number of dynamic objects in the traffic scenes may lead to a failure of object-based fusion because of a large number of association hypotheses. The grid-based approach allows us to avoid the object association problem for sensor fusion.

5.1.2.2. Disparity space approach for a vision based occupancy grid

To use sensors in the BOF framework, it is essential to develop an associated probabilistic sensor model that takes into consideration the uncertainty over measurements. In 2009, we proposed such a sensor model for stereo-vision [79]. The originality of the approach relied on the decision to work in the disparity space, instead of the classical Cartesian space. In 2010, we improved our sensor model, in order to mimic some features of the sensor models used for range finders. Particularly, we worked on managing visible/occluded areas of the scene [81], and on including the information from the road/obstacle segmentation of the disparity image [80]. Our approach was also designed to allow highly parallel computation of the occupancy grid. A. Nègre implemented the approach on GPU using NVIDIA CUDA to enhance the performance. The complete processing of the stereo data can now be done in 6 ms, while more than 150 ms were necessary with the CPU implementation. The complete approach for occupancy grid computation using stereovision is described in [30].

Figure 9 shows an example of the occupancy grid computed by our new approach. We can observe that most objects are detected (light color), even if partially occluded (e.g. the sign on the right). Information from the road surface is also taken into consideration (dark areas). Moreover, similar to a laser scanner, it appears that regions in front of objects are seen as partially unoccupied, while less information is available behind obstacles (occupancy probability is closer to 0.5).

In 2011, we focused on including the approach into the risk estimation framework on our Lexus experimental platform. We implemented a demonstration to estimate a distance measurement to the closer object situated in the future trajectory of the vehicle. The future trajectory is estimated either by using a lane detection algorithm (in the highway) or by combining velocity and steering information of the vehicle. Figure 10 shows the HMI displayed in the car while driving.

5.1.2.3. Processing of multi-layer telemetric data in probabilistic framework

Participants: Mathias Perrollaz, Juan-David Adarve, Alexandros Makris.
The occupancy grid computation based on a laser scanner uses the classical independent beam sensor model [90]. Since our vehicle is equipped with two four-layers laser scanners, it is necessary to merge the data from the multiple layers. In the original BOF framework, the fusion was performed through the classical Bayesian Fusion methodology. As shown in figure 11, this method causes problems of misdetection when some beams go over an object. In 2011, we proposed and implemented another approach. The fusion is now obtained through a weighted sum of the occupancy grids provided by each layer. The weight of each layer is obtained by computing a confidence grid. This confidence depends both on the inclination of the layer and on the possible occlusions. The new approach provides a more precise description of the environment.

5.1.3. Sensor Fusion and parameters estimation

Participants: Agostino Martinelli, Chiara Troiani.

This is the follow up of the research activity started in 2009, when a self-calibration problem for a wheeled robot has been investigated. The main results achieved during that year were published in [69], [71] and [70]. This calibration problem allows us to introduce a general framework able to deal with any estimation problem. This framework is based on a new theoretical concept, the concept of continuous symmetry. Detecting the continuous symmetries of a given system has a very practical importance. It allows us to detect an observable state whose components are non linear functions of the original non observable state. The general theory has been developed during the last two years. Preliminary results have been published in 2010 [72] and a more complete version of these results, which include several extensions, has been published on Transaction on Robotics, in 2011 [9].

In 2011, this general framework has been extensively applied to investigate the problem of the fusion of visual and inertial data in the framework of the European project sFly. Special emphasis has been devoted to the structure from motion problem (SfM) when fusing these data. This problem has particular interest and has been investigated by many disciplines, both in the framework of computer science ([35], [54], [56], [87] and references therein) and in the framework of neuroscience and vision perception ([67], [95] and references therein). Even though prior work has answered the question of which are the observable
Figure 10. a) segmentation of the environment with the stereo-vision algorithm. Blue areas belong to the road surface, while red areas belong to the obstacles. b) HMI shown in the car during the demonstration of risk estimation. The trajectory is estimated by considering the velocity and steering angle of the ego vehicle. Here the car in front is not considered as dangerous because it is more than 2 seconds ahead. c-d) Another example, on the highway. For this example, the trajectory is estimated by considering the road markings.
Figure 11. Occupancy grid computed after fusion of eight layers of laser data. Above: with the previous approach, some objects are not correctly represented (e.g. the barrier on the left). Below: with the new approach, the description is more precise.
modes, i.e. the states that can be determined by fusing visual and inertial measurements [35], [54], [56], the questions of how to compute these states in the absence of a prior, and of how many solutions are possible, were still unanswered. During 2011, we have derived, for the first time, a closed form solution to the SfM problem in this case, allowing the determination of the observable modes without the need for any prior knowledge. The proposed solution analytically expresses all the observable modes in terms of the visual and inertial measurements acquired during a given (short) time-interval allowing the determination of all the observable modes without the need for any prior knowledge. Additionally, we have shown that this problem can have a unique solution or two distinct solutions or infinite solutions depending on the trajectory, on the number of point-features and on the number of monocular images where the same point-features are seen. Our results are relevant in all the applications which need to solve the structure from motion problem with low-cost sensors and which do not demand any infrastructure. Typical examples are the emergent fields of space robotics [77], humanoid robotics and unmanned aerial navigation in urban-like environments [93], where the use of the GPS is often forbidden. Furthermore, our results could play an important role in neuroscience by providing a new insight on the process of vestibular and visual integration. To this regard, we remind the reader that the influence of extra retinal cues in depth perception has extensively been investigated in the last decades. In the case when this extra retinal cue is the motion parallax induced by self-motion relative to a stationary environment, the scale factor is provided by the head velocity [65], [66]. The vast majority of these studies, consider the case when the head motion is active [38], [94]. This prevents the possibility to understand the contribution of the vestibular signals because of efference copy generated by active self movement. However, a very recent study investigates this problem by performing trials with passive head movements [43]. The conclusion of this study is that the combination of retinal image with vestibular signals can provide rudimentary ability to depth perception. Our findings could provide a new insight to this problem of depth perception since by combining retinal image with vestibular signals it is possible to determine the scale factor even without any knowledge about the initial speed. New trials would be necessary in order to verify whether a mechanism reproducing our closed form solution is present in humans and/or in other animals (especially the ones without binocular vision). Our findings also show that it is possible to easily distinguish linear acceleration from gravity. Specifically, our closed form solution perform this determination by a very simple matrix inversion. This problem has also been considered in neuroscience [75], [31]. Our results could provide a new insight to this problem since they clearly characterize the conditions (type of motion, features layout) under which this determination can be performed.

Our results have been published in three conference papers [14], [11], [15] and have been accepted for publication in transactions on robotics (a version is currently available as a technical report, [29]).

In parallel to this theoretical activity an experimental activity has started in order to experimentally validate our findings in the near future and to deploy our technologies to industrial partners. To this regard, a contact with the company Delta Drone in Grenoble has been established and a valorization contract with a SME in the field of civil drone applications is currently in preparation.

5.1.4. Analysis of dynamic scenes for collision risk assessment


The grid-based environment representation is used for dynamic scene analysis in the Arosdyn project [78]. The original idea behind the risk estimation approach developed in the e-Motion team consists in considering the possible behaviors of the vehicles in the scene. Indeed, with the classical TTC(time to collision)-based approach, the risk is estimated based on the prediction of the trajectory, considering the current state of the objects. This is only valid for very short term predictions, and in some cases it can result in a over-estimation of the collision risk. Understanding the intention of the other participants of the road scene allows a longer term, more precise prediction of trajectories.

Our approach is divided into two steps: behavior recognition and behavior realization. The behavior recognition aims at estimating the probability for a vehicle to perform one of its feasible behaviors. The behaviors are semantic representations of driving maneuvers (e.g. turn left, turn right, go straight, ...). The probability
distribution over possible behaviors is obtained by inference using layered HMMs. Driving behavior realization is modeled as Gaussian Process (GP). This model allows us to obtain the probability distribution over the physical realization of the vehicle motion (i.e. trajectories) by assuming a usual driving, for a given behavior. Finally, a complete probabilistic model of the possible future motion of the vehicle is given by the probability distribution over driving behaviors, and by the realization of these behaviors. The risk calculation is performed by sampling of the paths from the corresponding GP. The fraction of the samples in collision gives the risk of collision.

In 2011, we conducted some early experiments on sensor fusion, using real data acquired with our Lexus experimental vehicle [16]. Moreover, the global framework of the Arosdyn project has been presented in [8].

5.1.5. Recognition for intelligent vehicles


We developed a generic object class recognition method. The method uses local image features and follows the part based detection approach. The state-of-the-art visual object class recognition systems operate with local descriptors and codebook representation of the objects. Various local features (e.g. gradient maps, edges) are used to create the descriptors. Then kernel based classifiers are commonly employed to classify the detected features in one of several object classes [32] [45]. The recognition of vehicles or pedestrians from sensors mounted on a moving platform is achieved by different approaches using various types of sensors, e.g. stereo camera, laser [51] [44]. The approaches that perform data fusion from various sensors have proven to be the more robust in a variety of road conditions [86].

This work focuses on the development of an object class recognition system which follows the part based detection approach [64]. The system fuses intensity and depth information in a probabilistic framework. To train the system for a specific object class, a database of annotated with bounding boxes images of the class objects is required. Therefore, extending the system to recognize different object classes is straightforward. We apply our method to the problem of detecting vehicles by means of on-board sensors. Initially, depth information is used to find regions of interest. Additionally, the depth of each local feature is used to weight its contribution to the posterior of the object position in the corresponding scale. The votes are then accumulated in a 3d space-scale space and the possible detections are the local maxima in that space. Figure 12 presents the steps of our approach.

![Figure 12. Detection procedure steps. The stereo information is used to define the regions of interest for the subsequent steps. Intensity and depth features are extracted from a dense grid within these regions. In the following the features are matched with the codebook clusters which are in turn used to estimate the posterior for the object in each position. The detections are the local maxima of the posterior.](image-url)
The novelty of our approach is the fusion of depth and intensity information to form a probabilistic part-based detector. Using depth information is beneficial for the robustness of the approach, because we avoid including many noisy detections resulting from false matches between features of different scales. The method is tested with stereo video sequences captured in an urban environment. Fig. 13 shows some example detections. The proposed method detects side-views of cars in various scales, in cases with partial occlusions, and under significant background clutter.

Figure 13. Car-side detection examples. True and false positive detections are represented by red and yellow bounding boxes respectively. (a) Cars in different scales with significant background clutter and significant occlusions are detected. (b) Precise detection of the un-occluded vehicle, whereas a vehicle that is heavily occluded in the left is not detected. (c) Difficult detection of a vehicle which is far and partially occluded and a false detection in the region between the road surface and the trees. (d) Detection with partial occlusion. (e) Partial detection of a taller than normal vehicle (on the left). The training dataset does not contain vehicles of this type. (f) Successful detection of a partially occluded car and a false positive arising from a bus and a van. Training separate detectors for these type of vehicles as well will help to avoid these false alarms.

5.1.6. Context-aware Bayesian estimation of risk at road intersections for cooperative vehicles

Participants: Stéphanie Lefèvre, Christian Laugier.

The work developed in this PhD is done in collaboration with Renault (CIFRE thesis) and concerns safety applications for cooperative vehicles.

In a few years, car manufacturers will start equipping vehicles with V2X communication devices, which will allow vehicles to share information with other vehicles and with roadside units using a dedicated communication channel. This new sensor on the car opens a whole new world of possibilities for Advanced Driver Assistance Systems (ADAS). In particular, the fact that the vehicle is able to “see” a car before it even enters the field-of-view of the driver allows for a better assistance in the tasks of perceiving, analyzing, predicting, and estimating the risk of a situation.

Early in the PhD we identified safety applications at road intersections as a relevant application domain for V2X technologies. The variety and complexity of scenes at road intersections makes reasoning and interpretation particularly difficult. On the other hand, intersections are a location of many accidents (they represent up to 50% of accidents in some countries), therefore reducing the accident rates in these areas would
have a considerable impact on global traffic safety. We also identified the key issues (and challenges) to be 1) situation understanding and 2) risk assessment, to be carried out from incomplete models and uncertain data.

The focus of the year 2010 was on the first of these two problems. We developed a Bayesian Network that could estimate a driver’s intended exit lane at an intersection based on the current state of the vehicle (position, orientation, turn signal state) and on contextual information extracted from the digital map. The idea was to use the information on the geometry of the road network and on the connectivity between lanes to build a statistical model of the relationship between the position and turn signal of a vehicle and the driver’s intended exit lane. Initial results of this work were published in IEEE CIVTS’11 [12], then in IEEE IV’11 [13] with a more thorough evaluation.

The objective of the work conducted in 2011 was twofold:

1. Extend the initial system: add some filtering and take into account the priority rules.
2. Estimate the risk of a situation, based on the estimated behavior/intention of the drivers in the scene.

We proposed a probabilistic motion model for vehicles approaching and traversing an intersection that incorporates some knowledge about how the context (i.e. the traffic rules, the presence of other vehicles, the geometry and topology of the intersection) influences vehicle behavior. The distinctive features of our algorithm are:

- The explicit use of priority rules
  
  Priority rules are explicitly taken into account in the motion model: the necessity for a driver to stop and/or yield to another vehicle at an intersection is estimated, jointly with the driver’s intention to comply. This allows for a flexible and computationally inexpensive computation of risk. Flexible because depending on the final application one can decide to compute different types of risk, e.g. the probability that a specific vehicle is a violator, or the probability that a crash will occur between two vehicles, or the risk of a specific maneuver for a vehicle. Inexpensive because these can be computed without performing trajectory prediction for the vehicles in the scene.

- The assumption that drivers generally respect traffic rules
  
  Instead of making the classical assumption that vehicles’ trajectories are independent, we model their mutual influences by introducing a prior knowledge that drivers generally respect priority rules. The motion model therefore takes into account the priority rules and the presence of other vehicles to better interpret correctly a vehicle’s behavior. The advantages are twofold. Firstly, we are able to better estimate the maneuver intention of the drivers, which means our situation assessment capabilities are improved. Secondly, risk is estimated with a higher sensitivity. We avoid risk overestimation while still being able to detect dangerous situations as well as the conventional, more conservative, methods.

This reasoning is implemented using a Bayesian filter which estimates the hidden variables M (maneuver intention), D (distance to intersection), H (intention to stop) and H’ (necessity to halt) jointly for all the vehicles in the scene, using the position, speed and heading information shared between the vehicles via V2X communication. Inference on the hidden variables is carried out by a particle filter. The algorithm was described in an INRIA research report [27]. In this report we showed by reasoning on theoretical scenarios that our assumption that drivers tend to respect priority rules should lead to improved situation assessment and risk assessment (see Fig. 14).

Recently, data has been collected at an intersection using the Renault demonstrator vehicles, so that our algorithm can be tested on real data. Preliminary results seem to confirm that the intuitions described in the research report were correct. A Graphical User Interface is in the process of being developed so that demonstrations of the system can be carried out live in the Renault demonstrator vehicles (see Fig. 15).

5.2. Human Centered Navigation in the physical world
Figure 14. Illustration of a scenario where the advantage of taking into account the interactions between vehicles for maneuver prediction is obvious for ADAS applications. The behavior of the red vehicle is interpreted differently depending on whether or not the interactions with the green vehicle are considered.

Figure 15. Graphical User Interface for warning a driver of a violation of priority rules at an intersection (the violator vehicle is displayed in red).
5.2.1. Goal oriented risk based navigation in dynamic uncertain environment


Navigation in large dynamic spaces has been addressed often using deterministic representations, fast updating and reactive avoidance strategies. However, probabilistic representations are much more informative and their use in mapping and prediction methods improves the quality of obtained results.

Since 2008 we have proposed a new concept to integrate a probabilistic collision risk function linking planning and navigation methods with the perception and the prediction of the dynamic environments [47]. Moving obstacles are supposed to move along typical motion patterns represented by Gaussian Processes or Growing HMM. The likelihood of the obstacles’ future trajectory and the probability of occupation are used to compute the risk of collision. The proposed planning algorithm, call Risk-RRT, is a sampling-based partial planner guided by the risk of collision. Results concerning this work were published in [48] [49] [50].

In 2011, our algorithms were integrated into an embedded software architecture for social aware navigation (see fig. 16). For this purpose we started to migrate our algorithms to a new experimental platform. Moreover, we adapted the code to the open source software called ROS (Robot Operating systems) which offers tools to develop robot applications based in state of the art algorithms. Particularly, localization and visualization tools have been used. We have linked the control of our robotic wheelchair, the Risk-RRT planning and the social filter modules described in 5.2.2 into the framework ROS as shown in figure 16. The main objective was to increase the visibility of our approach and develop common libraries with research groups in robotics. In 2011, in the scope of the AEN PAL project, we started a collaboration with the EPI Arobas and complementary developments have been put on the INRIA forge.

Figure 16. Architecture for the social navigation system in ROS

Next two sections are conducted under the french project PAL “Personally Assisted Living” with a goal to enhance the quality of living by providing more autonomy in the daily activities of the disabled.

5.2.2. Social conventions based navigation

Participants: Jorge Rios-Martinez, Anne Spalanzani, Christian Laugier.

\[^2\]Willow Garage Inc., http://www.ros.org
The objectives of this work are to integrate the notion of comfort in the classical safe navigation methods. If one consider that the navigation system transports a person, the integration of social conventions in the navigation strategy starts to be crucial. In this work, we propose to integrate the notions of personal space and interaction between people. We propose to enrich the knowledge the robot has, with a representation of the social conventions. The robot must take into consideration interactions to avoid groups of people (even if passing through the group is the “best” path for a conventional planning algorithm), or to join a group with a behavior close to the one of a human. To understand the behaviors of interaction between humans and the management of space, the works developed in the area of sociology to define some concepts as Personal space, o-space and F-formations are used.

- **Personal Space**
  In [53], Hall describes the use of space between humans, he observed the existence of some rules that conducted people to keep distances from others. He proposed a classification of the space around a person (its Personal Space) in social interaction in four zones:
  - the public zone > 3.6m,
  - the social zone > 1.2m
  - the personal zone > 0.45m
  - the intimate zone < 0.45m
  This is a useful tool for a robot to understand the intentions of the humans. It is well known that these measures are not stricts and that they change depending on age, culture and type of relationship but the categories proposed explain very well reactions like the uncomfortable sense of a stranger invading your intimate zone or the perception of somebody looking social interaction because he is entering to your social zone.

- **F-formation**

![F-formation examples](image)

*Figure 17. Examples of F-formations: (a) Vis-a-vis, (b) L-Shape, (c) C-Shape.*

In [57], Kendon observed that people interacting in groups follow some spatial patterns of arrangement. When people are executing some activity they claim an amount of space related to that activity, this space is respected by other people and Kendon referred it as individual’s transactional segment. This transactional segment can vary depending on body size, posture, position and orientation during the activity. Moreover the groups can establish a joint or shared transactional segment and only the intervenants have permitted access to it, they protect it and others tend to respect it. The o-space is that shared transactional segment. A F-formation system is the spatial-orientation arrangement that people create, share and maintain around their o-space. We can see in fig. (17) three examples of F-formations.

The first stage in order to achieve an integration of social concepts with robot navigation was to include estimations of the risk of disturbing personal space and interaction space in the general risk estimation. A strategy to detect interactions in the environment based in the velocity, position and orientation of humans was implemented.
In fig. 18 we observe the results of the proposed integration, the robot (green rectangle) can use the detections of conversations (light ellipses) between humans (blue circles) for add more risk to paths that invade the space of conversations. When a conversation is detected, a bi-dimensional Gaussian $G$ is created to represent the interaction space, also called o-space, the center of this space is approximated by taking into account the the participants’ poses. Then, $G$ is used to obtain an estimation of risk of disturbing by passing around the conversation. The navigation strategy is based on the Risk-RRT algorithm. Details of this approach were published in [18].

5.2.3. Autonomous Wheelchair for the Elderly People’s Assistance

Participants: Arturo Escobedo-Cabello, Anne Spalanzani, Christian Laugier.

The elderly and the disabled are expected to benefit from the new technologies in the field of autonomous navigation robotics. Normal users of electric wheelchairs will also benefit from the development of more automatic functionalities bringing an extra driving comfort, especially during delicate maneuvers such as narrow door passages. This contribution is similar to the installation of driving assistance on a car. A simple improvement of the classical powered wheelchair can often diminish several difficulties of control.

Comfort defined as a state of ease and satisfaction of bodily wants, with freedom from pain and anxiety, has recently emerged as a design goal in autonomous navigation systems. Designers are becoming more aware of the importance of the user when scheming solution algorithms. The idea of comfort is especially important in the case of wheelchairs where the occupants are weak as result of their age or disease.

For any robot that is designed to transport people, the trajectory should be smooth and correspond to the user’s understanding as much as possible. Since human interpretation of the environment often differs from a robot’s interpretation, the decisions taken by the system might seem incomprehensible to a human observer. For example an autonomous vehicle could refuse to move forward due to some obstacle, while a human user would easily be able to move its way through. This undesirable behaviors may prove irritating and with time may lead to users stopping from using the system.

In 2011 we setup a robotic wheelchair as a trial platform. The wheelchair is a differential drive robot equiped with a SICK LMS-200 lidar to get 2D range information from the environment, odometry sensors, and a velocity controller we have also added a kinect sensor in order to perform some in the field of social

Figure 18. Detecting conversations in the environment lets the robot to take navigation decisions that avoid humans activity interruption
interactions. Some basic functions can be executed including the mapping of the environment using a Rao-
Blackwellized Particle Filter [52], localization using an Adaptive Monte Carlo Localization approach (AMCL) [91], global planning using an A* algorithm [60] and local reactive planning using the Dynamic Window Algorithm [46].

Alongside we started working with the kinect sensor to detect and track people. Using the given tracking
information, the wheelchair is able to follow a human located in front of it. This behavior is aimed to
bring assistance not only to the user but also to the caregiver by allowing him to move without pushing the
wheelchair. The technical implementation of the related approaches has been done on the basis of the ROS
middleware due to easy integration with other opensource robotics software which benefit sharing and testing
developed software.

In 2012 we shall focus on the estimation of the user intentions by learning models of behavior. We’ll then use
these models to propose an adaptive autonomous navigation method that best answer the user needs.

5.2.4. Multi-Robot Distributed Control under Environmental Constraints

Participants: Agostino Martinelli, Alessandro Renzaglia.

This research has been carried out in the framework of the European project sFly. In recent years it is revealed
more and more the importance of using multi-robot systems for security application, otherwise impossible to
be performed by a single robot.

The main problem approached is the optimal surveillance coverage of an unknown and complex environment,
i.e. finding the optimal deployment for the robots and the way to safely reach such configuration. The solution
for the 2D case without obstacles is already known in literature [39]. On the other hand, for the non-convex
case, it is still a difficult problem. In [84] we firstly proposed a possible strategy based on a combination
of the repulsive potential field method and the Voronoi partition. Then, in the last two years we have mainly
approached the coverage problem by using a new stochastic optimization method. This work is in collaboration
with professor Elias Kosmatopoulos, from CERTH (Thessaloniki), and professor Lefteris Doitsidis, from TUC
(Crete), partners in the sFly project.

The Kosmatopoulos’s group has proposed a new adaptive stochastic optimization algorithm for a general
class of multi-robot passive and active sensing applications [59], [58]. This method possesses the
capability of being able to efficiently handle optimization problems for which an analytical form of the
function to be optimized is unknown, but the function is available for measurement at each iteration of the
algorithm employed to optimize it. As a result, it perfectly suits for multi-robot optimal coverage in non-
convex environments, where the analytical form of the function to be optimized is unknown but the function
is available for measurement (through the robots’ sensors) for each multi-robot configuration.

The main results obtained for the 2D case by using this method has been published in [85], [83]. We assume
the robots are equipped with global positioning capabilities and visual sensors able to monitor the surrounding
environment. The goal is to maximize the area monitored by the team, by identifying the best configuration of
the team members. Moreover, in 2011, a distributed version of the algorithm was presented in [17]. In multi-
robot systems, a distributed approach is desirable for several fundamental reasons. The most important are
failure of the central station and limited communication capabilities. The proposed approach has the following
key advantages with respect to previous works:

- it can solve the problem in a distributed way;
- it does not require any a priori knowledge on the environment;
- it works in any given environment, without the necessity to make any kind of assumption about its
topology;
- it can incorporate any kind of constraints, for instance regarding a possible existing threshold on the
maximum distance on the monitored region, or a limited visibility angle;
- it does not require a knowledge about these constraints since they are learnt during the task
execution;
- its complexity is low allowing real time implementations.
Figure 19. (a) Wheelchair used in the emotion team, (b) Two people being tracked using the kinect and the map of the environment done by the wheelchair.
The previous approach has been also extended for the more important and realistic 3D case. Working in collaboration with the ETHZ (Zurich), some simulations using real data, which were collected with the use of a miniature quadrotor helicopter specially designed for the needs of the European project sFly, have been performed (see fig. 20). This work has lead to two joint publications with CERTH and TUC: one conference paper to present (CDC2011) and one journal papers under review, and two joint publications with CERTH, TUC and ETHZ: one conference paper ([10]) and one journal papers under review.

In 2011, this approach has been combined with human aware navigation technics presented in section 5.2.5.

In the next months, the algorithm will be implemented on real MAVs for the final demo of the project. This demo will include experimentation both in indoor and outdoor complex environments.

Finally, a new collaboration with professor Kosmatopoulos has recently begun. The objective of this work is to develop a new efficient and scalable algorithm for multi-robot active control to perform cooperative simultaneous localization and mapping (CSLAM) and target tracking. The main idea is to use a convex optimization algorithm based on Semi-Definite Programming and Sum-of-Squares polynomials. Preliminary simulation results are very promising and a journal paper is under preparation.

5.2.5. Exploring stochastic optimization method to navigate between humans


Suppose that we have a robot navigating in an unknown and complex environment where people are moving and interacting. In such scenario the respect of the humans’ comfort becomes an important goal to achieve. The discomfort concept could be very general but we focus on the one mentioned before, i.e., the discomfort caused by disturbing one interaction or a personal space of humans. The approach here is to minimize the discomfort while the robot is navigating. As we cannot measure directly the value of discomfort, we can
infer it by modeling the concepts presented before using simple equations and after by applying a method of optimization. We propose to exploit a new stochastic and adaptive optimization algorithm (CAO) [59]. This method is very useful in particular when the analytical expression of the optimization function is unknown but numerical values are available for any state configuration. Furthermore, the proposed method can easily incorporate any dynamical and environmental constraints. To validate the performance of the proposed solution, several simulation results are provided.

![Simulation of robot navigating in an environment populated by people at three different times, three humans walking and two in conversation. Above discomfort function, below image of scenario, people represented by circles, robot’s positions represented by small triangle.](image)

In fig. 21 the model for discomfort function is shown together with robot navigation. At each step the robot randomly generate configurations in the environment and selects the one that takes it closer to the goal while minimizing values for the discomfort function of humans in the environment, this is repeated until goal is reached. Several executions of proposed approach in different scenarios can be observed in fig. 22.

The details of this approach have been submitted to ICRA2012.

5.3. Bayesian Modelling of Sensorimotor Systems and Behaviors

Results proposed in this section were done in collaboration with the LPPA collège de France.

5.3.1. Bayesian programming applied to a multi-player video games

**Participants:** Gabriel Synnaeve, Pierre Bessière.

The problem addressed in this work is the autonomous replacement of a human player. It is the continuation of last year’s work on the same topic as well as a follow-up of previous E-Motion Ph.D Ronan Le Hy [61]. This year, we focused on real-time strategy (RTS) games, in which the players have to build an economy, advance technology, produce and control an army to kill the opponents. From a research point of view, multi-player games are interesting because they stand for a good in-between of the real world and simulations. The world is finite and simulated (no sensors problems) but we didn’t wrote the simulation and the other players are humans (or advanced robots in the case of AI competitions).
This year’s research work focused on plan recognition from noisy and incomplete observations. Previous plan recognition works in multiplayer games were mainly based on planning and case-based reasoning (CBR) [92], [68], [55], [76] or HMMs [41]. CBR allows for taking domain knowledge into account easily while not dealing efficiently with uncertainty/incompleteness of information, HMMs deal with uncertainty quite well but domain knowledge is harder to structure. We found different ways to decompose the joint $P(\text{Observation}_1:N, \text{Plan}_1:M)$ which allows for tractable and robust inference. For instance with the help of intermediate variables which can be derived from domain knowledge (as we did) or found automatically (e.g. cross-validation on a HMM). Particularly, we were able to structure dependencies between domain knowledge extracted variables using coherence variables. We then learn the parameters of such joint distributions from data. Supervised (labeled), and semi-supervised learning (when we label automatically from clustering) have led to a publication at CIG (IEEE) 2011 [19] and unsupervised learning (using only raw game data) led to a publication at AIIDE (AAAI) 2011 [20].

On top of the research/evaluation implementation, we also implemented it in our StarCraft: Broodwar’s bot implementation BroodwarBotQ. With this bot, we took part in AIIDE and CIG conferences AI tournaments placing respectively 9th (out of 18) and 4th (out of 10). We also published last year’s result on multiple units control in real-time engagements (see 23) at CIG (IEEE) 2011 [21]. As optimal micro-management is almost always intractable (P-space) in real situations, we considered each unit as a Bayesian sensory motor robot which makes a fusion of its sensory inputs about the world, the enemy units, but also its allies (without explicit communication for less complexity) and higher level directions. So the units only take short term decision on where to go and who to attack, higher level planning is done at a squad (and then army) level and given as a sensory input. Results in micro-management tournaments are state of the art. In the more general case, they could be improved by reinforcement learning of the models parameters.

Figure 22. More simulations with different scenarios, start position in green, goal position in red. In (a) up robot decides to take a path that minimizes discomfort. In (b) robot changes its route to do not disturb human at right. Bottom, a pair of complex scenarios where paths chosen respects people comfort.
We are now working on concurrent goals resources attribution, still in the context of incomplete knowledge about the opponent. We are also working on correlating low-level observations (effects) and high-level inferences (causes) about the enemy strategy to be able to predict its future behavior.

![Figure 23. A real-time engagement: where should we go (we consider only the wide arrows)? Who should we fire on (we can fire only on orange arrows pointer units, while violet units are also potentially interesting targets)?](image)

5.3.2. Bayesian modelling to implement and compare different theories of speech communication

**Participants:** Raphael Laurent, Pierre Bessière, Julien Diard, Jean-Luc Schwartz.

A central issue in speech science concerns the nature of representations and processes involved in communication. The search for phoneme or syllable specific invariants led to three major sets of approaches: motor, auditory and perceptuo-motor theories, which have been widely argued for and against. The debate appears to be stagnating. This work is based on the belief that mathematical modeling of these theories could provide breakthroughs. More precisely, it is proposed that casting these theories into a single, unified mathematical
framework would be the most efficient way of comparing the theories and their properties in a systematic manner.

Bayesian modeling provides a mathematical framework that precisely allows such comparisons. The same tool, namely probabilities, can be used both for defining the models and for comparing them. Moreover, the use of a unified framework implies that common hypotheses would have common mathematical translations. This helps toward more principled studies of the competing theories.

Following this integrative approach, the motor, auditory and perceptuo-motor theories are thus cast into one unifying Bayesian framework in which they all appear as instances of various questions asked to one probabilistic communication model. This allows to compare these theories through quantitative testing in various paradigms. The work is aimed at understanding the differences in the predictions given by the different theories, and from these predictions to suggest experiments involving human subjects.

The model was used first to work on purely theoretical simulations aimed at studying with diverse paradigms the decrease in the performances predicted by the different theories due to communication noise. It was then used to work on plosive syllables production and perception, thanks to VLAM, a vocal tract simulation tool, which allows to map articulatory parameters to acoustic signals.
6. New Results

6.1. Autonomous Development of Representations

6.1.1. Open-ended bootstrapping of new sensorimotor representations

Participants: Alexander Gepperth.

We have explored a novel approach to the open-ended development of internal representations in autonomous agents, addressing in particular the transfer of knowledge between different modalities or abstraction levels. We propose a self-organized neural learning paradigm termed PROPRE (projection-prediction) that is driven by predictability: competitive advantages are given to those feature-sensitive elements that are inferable from activity in a reference representation, which may be innate or previously formed by learning. For generating and adapting the new induced representations, PROPRE implements a bi-directional interaction of clustering (“projection”) and inference (“prediction”), the key ingredient being an easy-to-compute online measure of predictability, by which the projection step is encouraged to favor sensitivity to predictable clusters. We demonstrated the potential of this paradigm by several simulation experiments with synthetic inputs. We showed that induced representations are indeed significantly more sensitive to predictable stimuli, that they are continuously being adapted to changing input statistics and that the behavior under severe resource constraints is favorable.

6.1.2. The contribution of context information to object detection in intelligent vehicles

Participants: Alexander Gepperth, Michael Garcia Ortiz.

In this work package, we explored the potential contribution of multimodal context information to object detection in an “intelligent car”. The used car platform incorporates several sophisticated processing subsystems, both for the detection of objects from local visual patterns as well as for the estimation of global scene properties, e.g., the shape of the road area, or the 3D position of the ground plane (sometimes denoted ”scene context” or just ”context”). Annotated data recorded on this platform is publicly available as the ”HRI RoadTraffic” vehicle video dataset, which formed the basis for this investigation.

In order to quantify the contribution of context information, we investigated whether it can be used to infer object identity with little or no reference to local patterns of visual appearance. Using a challenging vehicle detection task based on the ”HRI RoadTraffic” dataset, we trained selected algorithms to estimate object identity from context information alone. In the course of our performance evaluations, we also analyzed the effect of typical real-world conditions (added noise, high dimensions, environmental variation) on context model performance.

As a principal result, we showed that the learning of context models is feasible with all tested algorithms, and that object identity can be estimated from context information with similar accuracy as by relying on local pattern recognition methods. We also found that the use of basis function representations (also known as ”population codes”) allows the simplest (and therefore most efficient) learning methods to perform best in the benchmark, suggesting that the use of context is feasible even in systems operating under strong performance constraints.

6.1.3. Discovering object concept through developmental learning

Participants: Natalia Lyubova, David Filliat.

The goal of this work is to design a visual system for a humanoid robot. Taking inspiration from child’s perception and following the principles of developmental robotics, the robot should detect and learn objects from interactions with people and from experiments it performs with objects, avoiding the use of image databases or of a separate training phase. In our model, all knowledge is therefore iteratively acquired from low-level features and builds up hierarchical object models, which are robust to changes in the environment, background and camera motion.
In our scenario, people in front of the robot are supposed to interact with objects to encourage the robot to focus on them. We therefore assume that the robot is attracted by motion and we segment possible objects based on clustering of the optical flow. Additionally, the depth information from a Kinect is used to filter visual input, considering the constraints of the robot’s working area and to refine the object contours obtained from motion segmentation.

The appearance of objects is encoded following the Bag of Visual Words approach with incremental dictionaries. We combine several complementary features to maximize the completeness of the encoded information (SURF descriptor and superpixels with associated colors) and construct pairs and triples of these features to integrate local geometry information. These features make it possible to decide if the current view has been already seen or not. A multi-view object model is then constructed by associating recognized views and views tracked during object motion.

This system is implemented on the iCub humanoid robot, which detects objects in the visual space and characterizes their appearance, their relative position and their occurrence statistics. Ten objects were presented in the current experiment; each of them was manipulated by a person during 1-2 minutes. Once the vocabulary reached a sufficient amount of knowledge, the robot was able to reliably recognize human hands and most of objects.

6.1.4. Scaling-up Knowledge for a Cognizant Robot

Participants: Thomas Degris, Joseph Modayil.

A cognizant robot is a robot with a deep and immediately accessible understanding of its interaction with the environment—an understanding that the robot can use to flexibly adapt to novel situations. Such a robot will need a vast amount of situated, revisable, and expressive knowledge to display flexible intelligent behaviors. Instead of relying on human-provided knowledge, we consider the case where an arbitrary robot can autonomously acquire pertinent knowledge directly from everyday interaction with the environment. We study how existing ideas in reinforcement learning theory can be used to formalize knowledge and use reinforcement learning techniques to enable a robot to maintain and improve its own knowledge. We consider robot performing a continual learning process that scales-up knowledge acquisition to cover a large number of facts, skills and predictions. This knowledge has semantics that are grounded in sensorimotor experience and can then be used for more abstract process such as planning. We see the approach of developing more cognizant robots as a necessary key step towards broadly competent robots.

Paper being published: Scaling-up Knowledge for a Cognizant Robot accepted at Designing Intelligent Robots: Reintegrating AI, AAAI Spring Symposium 2012.

6.1.5. Learning parallel combinations of motor primitives from demonstration and linguistic guidance with non-negative matrix factorization

Participants: Olivier Mangin, Pierre-Yves Oudeyer.

We have elaborated and experimented a novel approach to joint language and motor learning from demonstration. It enables discovery of a dictionary of motor and linguistic primitives, that can be combined in parallel to represent training data as well as novel skills in the form of combinations of known skills. These methods and the results of our experiments participate in addressing two main issues of developmental robotics: 1) symbol grounding for language learning; 2) achieving compositionality in motor-learning from demonstration, which enables re-using knowledge and thus scaling to complex tasks. In particular, we are interested in learning motor primitives active in parallel, a less explored way of combining such primitives. To address these challenges we have explored and studied the use of nonnegative matrix factorization to discover motor primitives from histogram representations of data acquired from real demonstrations of dancing movements. Initial results were presented in [30] and further results are presented in an article under review.


6.2.1. The SAGG-RIAC algorithm: competence based active learning of motor skills

Participants: Adrien Baranès, Pierre-Yves Oudeyer.
We have continued to develop and experiment the Self-Adaptive Goal Generation - Robust Intelligent Adaptive Curiosity (SAGG-RIAC) algorithm as an intrinsically motivated goal exploration mechanism which allows high-dimensional redundant robots with various body schemas to efficiently and actively learn motor skills in their task space. The main idea is to push the robot to perform active babbling in the low-dimensional goal/task space, as opposed to motor babbling in the high-dimensional actuator space (possibly defined with motor primitives), by self-generating goals actively and adaptively in regions of the task space which provide a maximal competence improvement for reaching those goal states. Then, a lower level active motor learning algorithm is used to drive the robot to locally explore how to reach a given self-generated goal. We have conducted systematic experiments with high-dimensional continuous sensorimotor spaces related to different robotic setups such as a highly-redundant robotic arm, a quadruped, and an arm controlling a fishing rod with a flexible wire and show that 1) exploration in the task space can be a lot faster than exploration in the actuator space for learning inverse models in redundant robots; 2) selecting goals based on the maximal improvement heuristics creates developmental trajectories driving the robot to progressively focus on areas of increasing complexity and is statistically significantly more efficient than selecting goals randomly, as well as more efficient than different standard active motor babbling methods. These results were published in [13], [15], [17] and a journal publication is in preparation.

### 6.2.2. SGIM-D: Bootstrapping Intrinsically Motivated Learning with Human Demonstration

**Participants:** Mai Nguyen, Pierre-Yves Oudeyer.

We have studied the coupling of internally guided learning and social interaction, and more specifically the improvement owing to demonstrations, of the learning by intrinsic motivation. We have designed Socially Guided Intrinsic Motivation by Demonstration (SGIM-D), an algorithm for learning the mapping between high dimensions in continuous, non-preset, highly redundant environments. We have shown through a robot learning experiment involving a high-dimensional sensorimotor space related to fishing skills that SGIM-D efficiently combines the advantages of social learning and intrinsic motivation to gain a wide repertoire while being specialised in specific subspaces. An article presenting aspects of this work was awarded the second best student paper award in IEEE ICDL/Epirob 2011 [27].

### 6.2.3. Maturationally-Constrained Competence-Based Intrinsically Motivated Learning

**Participants:** Adrien Baranès, Pierre-Yves Oudeyer.

We have continued to develop computational models of the coupling of intrinsic motivations and physiological maturational constraints, showing that both mechanisms may have complex bidirectional interactions allowing the active control of the growth of complexity in motor development which directs an efficient learning and exploration process. The coupling relies on the Self-Adaptive Goal Generation - Robust Intelligent Adaptive Curiosity algorithm (SAGG-RIAC) that instantiates an intrinsically motivated goal exploration mechanism for motor learning of inverse models. Then, we have introduced a functional model of maturational constraints inspired by the myelination process in humans, and showed how it can be coupled with the SAGG-RIAC algorithm, forming a new system called McSAGG-RIAC2. We have then conducted systematic experiments to evaluate qualitative and, more importantly, quantitative properties of these systems when applied to the learning of the forward and inverse kinematic of an unknown robotic arm of up to 60 dimensions, the learning of walking in a 12DOF quadruped controlled with 24 dimensions motor synergies, and learning the control of a fishing rod involving a flexible/rope component. These results were published in [13], [15], [17] and a journal publication is in preparation.

### 6.2.4. Actor-Critic for Parallel Learning

**Participants:** Thomas Degris, Matha White, Richard Sutton.

Parallel learning is necessary for a robot to learn multiple tasks in parallel while executing a behavior in the environment not necessarily directly related to the tasks to learn. In previous existing work, an interesting class of learning algorithms for control are actor–critic. First, these algorithms can be used with high-dimensional action space. Second, they also sometimes provide computational models for biological decision-making systems. At FLOWERS, we work on new actor–critic algorithms suitable for parallel learning, with
theoretical guarantees, applicable and practical to use with robots, and formulated in the general framework of reinforcement learning.

6.2.5. Curiosity for Parallel Learning of Predictions and Tasks from the Continuous Interaction of a Robot with its Environment

**Participants:** Thomas Degris, Adam White, Pierre-Yves Oudeyer.

On one hand, a robot needs a wide variety of knowledge to fully interact with its environment. On the other hand, a robot, like humans or animals, can only perform one behavior at a time in the real world to learn this vast amount of knowledge. A solution to scale up learning while keeping the interaction time with the real world realistic is to learn multiple elements of knowledge simultaneously in parallel. The Horde architecture proposes a set of demons each learning about new policies (i.e. skills) and predictions about these skills (i.e. partial models) off-policy simultaneously. The number of demons learning in parallel is limited only by memory and processing power, and not by the fact that there is only one sensorimotor interaction with the environment to learn from. At FLOWERS, we investigate the question of what the behavior policy of the robot should be to speed-up learning of the demons. Our goal is to test if the Horde scales-up to complex humanoid robots and if, driven by intrinsic motivations, it can autonomously learn building blocks of knowledge for future, more complex, behaviors.

6.2.6. Optimal Teaching on Sequential Decision Tasks

**Participants:** Manuel Lopes, Maya Cakmak.

A helpful teacher can significantly improve the learning rate of an autonomous learning agent. Teaching algorithms have been formally studied within the field of Algorithmic Teaching. These give important insights into how a teacher can select the most informative examples while teaching a new concept. However the field has so far focused purely on classification tasks. In this paper we introduce a novel method for optimally teaching sequential decision tasks. We present an algorithm that automatically selects the set of most informative demonstrations and evaluate it on several navigation tasks. Next, we present a set of human subject studies that investigate the optimality of human teaching in these tasks. We evaluate examples naturally chosen by human teachers and found that humans are generally sub-optimal. Then based on our proposed optimal teaching algorithm we try to elicit better teaching from humans. We do this by explaining the intuition of the teaching algorithm in an informal language prior to the teaching task. We found that this improves the examples elicited from human teachers on all considered tasks. This shows that a simple modification the instructions given to human teachers, has the potential of greatly improving the performance of the agent trained by the human [41].

6.2.7. Inverse Coordinated Reinforcement Learning

**Participants:** Manuel Lopes, Jonathan Sprauel.

Inverse Coordinated Reinforcement Learning

We extended of inverse reinforcement learning to the multi-agent case. Under this formalism a team of agents can learn a task goal, encoded as a reward function, by observing another team executing that task. Our agents behave using local information and limited communication following the coordinated reinforcement learning framework. We show that a team behavior can be learned using this formalism and how well this mechanism can deal with changing initial conditions and number of agents [68].

6.3. Motor Learning and Morphological Computation

6.3.1. Morphological Computation in Acroban the Humanoid: Balance Control and Dynamic Walking

**Participants:** Olivier Ly, Pierre-Yves Oudeyer, Matthieu Lapeyre, Jérome Béchu, Paul Fudal, Haylee Fogg.
We have continued to elaborate and experiment the humanoid platform Acroban and its use to study various scientific topics. Our goal was to study three main issues: 1) Compliance and semi-passive dynamics in the framework of dynamic walking in humanoid robots and more generally its impact in terms of semi-passive interactive motor primitives and their robustness to unknown external perturbations; 2) the advantage of a bio-inspired multi-articulated vertebral column in the dynamics of these motor primitives; The platform uses mechatronic components that allow us to adjust dynamically the compliance of actuators, which combines with the intrinsic mechanical compliance of the structure due to the use of elastics and springs. We have explored how these capabilities can allow us to enforce morphological computation in the design of robust dynamic locomotion. Compliance also allows us to design semi-passive motor primitives using the torso as a system of accumulation/release of potential/kinetic energy. This is made possible by the combination of adequate morphology and materials, full-body compliance, semi-passive and self-organized stable dynamics, as well as the possibility to experiment new motor primitives by trial-and-error thanks to light-weightedness. These results were presented in [25]. A dedicated web page with videos is available at: http://flowers.inria.fr/acroban.php.

6.3.2. Maturational constraints for motor learning in high-dimensions: the case of biped walking
Participants: Matthieu Lapeyre, Pierre-Yves Oudeyer, Olivier Ly.

We have elaborated and began to experiment a new developmental approach to motor learning in very high-dimensions, applied to learning biped locomotion in humanoid robots. This approach relies on the formal modeling and coupling of several advanced mechanisms inspired from human development for actively controlling the growth of complexity and harnessing the curse of dimensionality: 1) Maturational constraints for the progressive release of new degrees of freedoms and progressive increase their explorable ranges; 2) Motor synergies; 3) Morphological computation; 4) Social Guidance. An experimental setup involving a simulated version of the Acroban Humanoid robot, based on the V-REP simulator, has been elaborated, and initial encouraging results were obtained. These results are presented in [23].

6.3.3. Acroban v2: improving morphological computation with dampers
Participant: Olivier Ly.

Theoretical studies and experiments concerning in particular dynamics of passive walkers drove us to design, construct and continue to experiment a new version of Acroban. This new version has two goals both fitting in the study of the impact of morphology in the behaviour of the robot:

- experiment deep structural modifications of the morphology, in order to avoid as much as possible inelastic chocks. Indeed, during the gait, the unstability is mainly due to chock at the landing of the foot.
- improve the global ratio weight/power of the robot in order to get more dynamic movements.

Indeed, this new version uses RX-28 motors which are lighter than the RX-64 motors which are used in the first version of Acroban. The robot is smaller and lighter. First experiments show that the obtained ratio weight/power is better than the first version. Movements of the robot, and in particular amplitude of locomotion movements, are not limited by torque now. Second, we have experimented plastic materials to design the structure in order to makes it naturally flexible comparing to the metal used in the first version. This way, we improve the natural compliance of the robot. Finally, and this is probably the most important change, we used non actuated linear joints in the hip and in the spline. To control these linear joints, instead of servo-motors, we use dampers. This kind of design is new in humanoid robotic. While bringing new control problems (because of the non-controlled joints which makes the robot semi-passive), this design softens chocks in a significant maner. Experiment shows that stability of the whole structure is greatly improved especially during locomotion.

6.4. HRI and Robot Language Teaching

6.4.1. Intuitive and Robust Physical Human-Robot Interaction with Acroban
Participants: Olivier Ly, Pierre-Yves Oudeyer, Pierre Rouanet, Matthieu Lapeyre, Jérome Béchu, Paul Fudal, Haylee Fogg.
We have experimented and shown how the humanoid robot Acroban allows whole-body robust, natural and intuitive physical interaction with both adults and children. These physical human-robot interaction are made possible through the combination of several properties of Acroban: 1) it is whole-body compliant thanks to variable impedance control and also thanks to the use of elastics and springs; 2) it has a bio-inspired vertebral column allowing more flexibility in postural and equilibrium control; 3) it is light- weight; 4) it has simple low-level controllers that leverage the first three properties. Moreover, the capabilities for physical human-robot interaction that we show are not using a model of the human, and in this sense are “model free”: 1) the capability of the robot to keep its equilibrium while being manipulated or pushed by humans is a result of the intrinsic capability of the whole body to absorb unpredicted external perturbations; 2) the capability of leading Acroban by the hand is an emergent human-robot interface made possible by the self-organizing properties of the body and its low-level controllers and was observed a posteriori only after the robot was conceived and without any initial plan to make this possible. Finally, an originality of Acroban is that is is made with relatively low-cost components which lack of precision is counterbalanced with the robustness due to global geometry and compliance. These results were presented in [ 28 ]. A dedicated web page with videos is available at: http://flowers.inria.fr/acroban.php .

6.4.2. A Real World User Study of Different Interfaces for Teaching New Visually Grounded Words to a Robot

Participants: Pierre Rouanet, Pierre-Yves Oudeyer, Fabien Danieau, David Filliat.

We have continued to elaborate and experiment an integrated system based on a combination of advanced Human-Robot Interaction, visual perception and machine learning methods that allows non-expert users to intuitively and robustly teach new visually grounded words to robots. This system is based on the state-of-the-art bags of words technique but focuses on different mediator based interfaces that we can propose to the users. Indeed, we argue that by focusing on interaction we could help users to collect good learning examples and thus improve the performance of the overall learning system. We compared four different interfaces and their impact on the overall system through a real world study where we asked participants to show and teach a robot names for five different objects. Three interfaces were based on mediator objects such as an iPhone, a Wiimote and a laser pointer and provided the users with different kinds of feedback of what the robot is perceiving. The fourth interface was gesture based with a Wizard-of-Oz recognition system included in order to compare our mediator interfaces with a more natural interaction. We showed that the interface may indeed strongly impact the quality of the learning examples collected by users, especially for small objects. More precisely, we showed that interfaces such as the iPhone interface do not only give feedback about what the robot is perceiving but also drive users to pay attention to the learning examples they are collecting. Thus, this interface allows non-expert users to intuitively and easily collect almost as good learning examples as expert users trained for this task and aware of the different visual perception and machine learning issues. Finally, we showed that the mediator based interfaces were judged as easier to use than the a priori more natural gestures based interface. This work was presented in [ 29 ].

6.4.3. Language Acquisition as a Particular Case of Context-Dependant Motor Skills Acquisition

Participants: Thomas Cederborg, Pierre-Yves Oudeyer.

Imitation learning, or robot programming by demonstration, have made important advances in recent years. We have proposed to extend the usual contexts investigated to also include linguistic expressions. We have proposed a modification to existing algorithms within the imitation learning framework so that they can handle learning from the demonstration of several unlabelled tasks (or motor primitives) without having to inform the imitator of what task is being demonstrated or what the number of tasks is, which then allows directly for relatively complex language learning. A mechanism for detecting wether or not linguistic/speech input is relevant to the task has also been proposed. With these additions it becomes possible to build an imitator that bridges the gap between imitation learning and language learning by being able to learn linguistic expressions using methods from the imitation learning community. In this sense the imitator learns a word by knowing that a certain speech pattern present in the context means that a specific task is to be executed. The imitator is
however not assumed to know that speech is relevant and has to figure this out on its own by looking at the demonstrations. To demonstrate this ability to find the relevance of speech non-linguistic tasks are learnt along with linguistic tasks and the imitator has to figure out when speech is relevant (in some tasks speech should be completely ignored and in other tasks the entire policy is determined by speech). A simulated experiment demonstrates that an imitator can indeed find the number of tasks it has been demonstrated, discover what demonstrations are of what task, for which of the tasks speech is relevant and successfully reproduce those tasks. This work is presented in a publication under review.

6.4.4. Robot Learning by Imitation of Internal Cognitive Operations in the Context of Language Acquisition

Participants: Thomas Cederborg, Pierre-Yves Oudeyer.

We have examined the problem of learning socio-linguistic skills through imitation when those skills involve both observable motor patterns and internal unobservable cognitive operations. This approach is framed in a research program trying to investigate novel links between context-dependent motor learning by imitation and language acquisition. More precisely, the paper presents an algorithm for learning how to respond to communicative/linguistic actions of one human, called an interactant, by observing how another human, called a demonstrator, responds. The response of the demonstrator, which depends on the context, including the signs of the interactant, is assumed to be appropriate and the robotic imitator uses these observations to build a general policy of how to respond to interactant actions. In this paper the communicative actions of the interactant is hand signs, and the learnt behavior consists of how to respond to the hand signs of a small and simple sign language, both in terms of adequately focusing attention on the right part of the scene, and in terms of responding physically. As a response to two continuous signs of the interactant, the demonstrator focuses on one out of three objects, and then performs a movement in relation to the object focused on. An algorithm is proposed based on a similarity metric between demonstrations, and a simulated experiment is presented where the unseen “focus on object” operation and the hand movements are successfully imitated, including in situations where there are no demonstrations. This work has been published in [21].

6.4.5. Learning Simultaneously New Tasks and Feedback Models in Socially Guided Robot Learning

Participants: Manuel Lopes, Thomas Cederborg, Pierre-Yves Oudeyer.

We have developed a system that allows a robot to learn simultaneously new tasks and feedback models from ambiguous feedback in the context of robot learning by imitation. We have considered an inverse reinforcement learner that receives feedback from a user with an unknown and noisy protocol. The system needs to estimate simultaneously what the task is, and how the user is providing the feedback. We have further explored the problem of ambiguous protocols by considering that the words used by the teacher have an unknown relation with the action and meaning expected by the robot. This allows the system to start with a set of known symbols and learn the meaning of new ones. We have presented computational results that show that it is possible to learn the task under a noisy and ambiguous feedback. Using an active learning approach, the system is able to reduce the length of the training period. [24], [26].

6.5. Hardware

6.5.1. Ergo-Robots/FLOWERS Fields: Towards Large-Scale Robot Learning Experiments in the Real World

Participants: Jérome Béchu, Fabien Bénureau, Haylee Fogg, Paul Fudal, Hugo Gimbert, Matthieu Lapeyre, Olivier Ly, Olivier Mangin, Pierre Rouanet, Pierre-Yves Oudeyer.
In the context of its participation to the exhibition “Mathematics: A Beautiful Elsewhere” at Fondation Cartier pour l’Art Contemporain in Paris, starting from 19th October 2011 and to be held until 18th March 2012, the team has elaborated and experimented a robotic experimental set-up called “Ergo-Robots/FLOWERS Fields”. This set-up is not only a way to share our scientific investigations with the general public, but attacks a very important technological challenge impacting the science of developmental robotics: How to design a robot learning experiment that can run continuously and autonomously for several months? Indeed, developmental robotics takes life-long learning and development as one of its central objective and object of study, and thus shall require experimental setups that allow robots to run, learn and develop for extended periods of time. Yet, in practice, this has not been possible so far due to the unavailability of platforms adapted at the same time to learning, exploration, easy and versatile reconfiguration, and extended time of experimentation. Most experiments so far in the field have a duration ranging from a few minutes to a few hours. This is an important obstacle for the progress of developmental robotics, which would need experimental set-ups capable of running for several months. This is exactly the challenge explored by the Ergo-Robots installation, which we have approached by using new generations of affordable yet sophisticated and powerful off-the-shelf servomotors (RX Series from Robotis) combined with an adequately designed software and hardware architecture, as well as processes for streamlined maintenance. The experiment is now running for five months, six days a week, in a public exhibition which has strong constraints over periods of functioning and no continual presence of dedicated technicians/engineers on site. The experiment involved five robots, each with 6 degrees of freedoms, which are endowed with curiosity-driven learning mechanisms allowing them to explore and learn how to manipulate physical objects around them as well as to discover and explore vocal interactions with humans/the visitors. The robots are also playing language games allowing them to invent their own linguistic conventions. A battery of measures has been set up in order to study the evolution of the platform, with the aim of using the results (to be described in an article) as a reference for building future robot learning experiments on extended periods of time, both within the team and in the developmental robotics community. More information available at: http://flowers.inria.fr/ergo-robots.php and http://fondation.cartier.com/.

Figure 15. Installation of Ergo Robots at Foundation Cartier.
6. New Results

6.1. High speed autonomous driving on roads

Participants: Fawzi Nashashibi, Laurent Bouraoui, Paulo Lopes Resende.

In this exceptional year, IMARA-INRIA had a very busy agenda with several scheduled demonstrations in the frameworks of several European projects, especially HAVEit, CityMobil and CityNetMobil.

The HAVEit project final event took place at Boras (Sweden) in June. IMARA was tightly involved in the development of the Joint Demonstrator called FASCar-II. In this demonstrator, IMARA was responsible for the development of the Co-Pilot system. This is the main decisional system that handles the manoeuvres to be executed taking into account the multisensor data fusion sub-system, the driver monitoring system and the infrastructure (data provided by the infrastructure-to-vehicle telecommunications). The system also generates the trajectories to be executed as well the control-command laws to be send to the actuators [40] and provides passive or active assistance to the driver according to the active automation level.

![Figure 1. The Co-Pilot system in the Joint system architecture and HAVEit’s FASCar demonstrator](image)

In parallel, IMARA has also developed the “Wireless Infrastructure to Vehicle communication system”. Here, a specific hardware and software systems have been developed to allow V2I and V2V applications integrated on the Joint System demonstrator. This architecture – called the 4GCube – is based on the wireless communications devices and standards (802.11 a/b/g); they are IPv6 / CALM compliant architectures that have been tested with 2.4 GHz and 5 GHz bands, allowing multiple services handling.

The integration and the validation of the Co-pilot system as well as the communication device were done during the Final Event of HAVEit with a tremendous success and with high professional standards. The FASCar was able to demonstrate high driving autonomous skills at high speeds (up to 120 km/h) and was able to demonstrate new advanced features like overtaking mobile and static obstacles.

6.2. New urban transportation platforms: INRIA’s Cybus

Participants: Laurent Bouraoui, François Charlot, Carlos Holguin, Fawzi Nashashibi, Tony Noël, Michel Parent, Paulo Lopes Resende, Jianping Xie, Armand Yvet.
In order to achieve autonomous driving, autonomous systems (robots, intelligent vehicles, UAV’s, UGV’s, ...) must have a decisional system that integrates an advanced perception system that performs sensors data fusion and environment modeling. From perception to control, task planning and path planning algorithms have to plan safe and optimized itineraries while processing sensory data. Motion control is the last link of the processing chain where itineraries and dynamic trajectories are executed by the low level control system. IMARA works on each of the topics mentioned above.

With the European projects CityMobil and CityNet Mobil, IMARA had the opportunity to validate the autonomous driving architecture developed through the projects showcases held in European cities. In 2011, several cities hosted IMARA team in charge of demonstrating autonomous driving and autonomous sensor-based navigation using sensory data (laser scanners) and GNSS (GPS). These events were a total success and were the opportunity to deploy and test the large-scale SLAMMOT system used for environment mapping, vehicle localization and mobile obstacles detection but also the new VMS (Vehicle Management System) developed in order to coordinate the mobility and navigation of several Cybercars (INRIA’s Cybus platforms).

For this purpose, the perception sub-system was based on the generic SLAM-based system that was already presented in detail last year [43]. The proposed localization architecture has been implemented in two different vehicle platforms. AGV is a fully autonomous vehicle equipped with two IBEO Alasca-XT laser scanners (left and right front corner). Cycab is a prototype of smart car mounted with a single IBEO-ML laser scanner in front.

This year, because of delivery problems related to an industrial partner, IMARA had to design and develop its own cybercars. Thus the Cybus are the newest prototyping and demonstration platforms designed at INRIA. Apart from the chassis and engines, the whole hardware and software systems were developed thanks to IMARA’s researchers and engineers talents. These electric vehicles are based on a Yamaha chassis but the embedded intelligence is the result of this year’s IMARA developments. Much of the perception and control software is now registered.

The system developed can be seen as an experimental platform for a new public mobility transportation system operating in mixed environments. This electric vehicle is a “clean” transportation mean that is capable to achieve the well known last-mile itinerary in urban areas where classical transportation means are inefficient or simply non profitable. In order to demonstrate the feasibility of the system, a 3-months service has been programmed in the City of La Rochelle. The Mayor of the City authorized the Cybus to operate during 3-months providing free transport service to the inhabitants of the city. For that purpose, a 2 kilometers route has been defined on which 5 calling stations where installed. The users were able to call the Cybus from any station to reach any other station. The evaluation of the system has shown an acceptability of more than 95% among population; very low failures or technical leakages were reported. Following this successful operation,
the European Commission asked INRIA to extend this operation as well as its technical capabilities in order to achieve a simultaneous multiple Cybus navigation. This was achieved last December and was a real success.

The platforms developed here (Cybus) were exploited and demonstrated in the context of the EU-CityNetMobil project. The cities of Antibes (France) and Reggio di Calabria (Italy) hosted the team for 2 weeks respectively in order to experiment this new mobility and transportation mean in the heart of their cities.

Following this success, a new proposal of CityMobil-2 project is under submission with the objective this time to extend real operational mobility services to 6-12 months in selected European cities!

6.3. Communications Management in Cooperative Intelligent Transportation Systems

Participants: Thierry Ernst, Manabu Tsukada.

Cooperative Intelligent Transportation Systems (Cooperative ITS) are systems where the vehicles, the roadside infrastructure, central control centers and other entities exchange information in order to achieve better road safety, traffic efficiency and comfort of the road users. This exchange of information must rely on a common communication architecture. The ITS Station reference architecture has thus been specified in ISO and ETSI. It allows vehicles and roadside ITS stations to organize themselves into Vehicular Ad-hoc Network (VANET), presumably though IPv6 GeoNetworking using IEEE802.11p and to connect seamlessly to the
Internet though any available access technology. Several paths may thus be available at a given vehicle ITS
station to communicate with other ITS stations. Paths are of three types: direct path, optimize path and anchor
path. The objective of the study is to optimize the communication between ITS Stations by selecting the best
available communication path. This requires first to gather information available locally at the ITS station
(position, speed, application requirements, media characteristics, capabilities, path status, ...) and collected
from neighbors ITS stations (position, speed, services, ...) and then to process this information through a
decision-making algorithm. First, we define a network module allowing the combination of IPv6 together with
GeoNet-working. Second, we propose a cross-layer path selection management module. Our contributions are
mapped to the ITS station reference architecture by defining the relation between the ITS station network and
transport layer (which hosts our IPv6 GeoNetworking contribution) and the vertical ITS station cross-layer
entity (which hosts the path decision-making algorithm). We specify the functions allowing the exchange of
parameters through the Service Access Point (SAP) between the network layer and the management entity
(MN-SAP). The parameters used at the cross layer ITS station management entity are abstracted in a way
so that they are agnostic to the protocols used at the ITS station network and transport layer, therefrom
allowing easy replacement of protocol elements (e.g. replacing NEMO by other mobility support protocol)
or permutation of the network stack (IPv6 or GeoNetworking, a combination of both or other network stack).

6.4. Managing the system (via probabilistic modeling)

6.4.1. Belief propagation inference for traffic prediction

Participants: Cyril Furtlehner, Jean-Marc Lasgouttes, Arnaud Lewden, Victorin Martin.

This work [41] deals with real-time prediction of traffic conditions in a setting where the only available
information is floating car data (FCD) sent by probe vehicles. The main focus is on finding a good way to
encode some coarse information (typically whether traffic on a segment is fluid or congested), and to decode
it in the form of real-time traffic reconstruction and prediction. Our approach relies in particular on the belief
propagation algorithm.

These studies are done in particular in the framework of the projects TRAVESTI and Pumas.

This year’s highlights are

- A particular effort has been done this year in studying the theoretical aspects of the ways to encode
  real valued variable into an binary Ising model. A publication on the subject is in preparation.
- A review of our work on road traffic inference using methods from statistical physics has been
  published [21].
- The investigation of the effect of various types of normalization in the belief propagation algorithm
  has lead to a technical report [38].
- Arnaud Lewden has specified and implemented the new software BPstruction, which is our contribu-
  tion to the Pumas project. Besides implementing traffic reconstruction from FCD, it is intended as
  a testbench for our research on inference using Belief Propagation.
- Victorin Martin has given a talk at the “Séminaire de Modélisation des Réseaux de Transport” at
  IFSTTAR. He presented there our method for real-time traffic reconstruction and prediction.
- Jean-Marc Lasgouttes also presented this work at the Xerox Research Centre Europe seminar.

6.4.2. Evaluation of dual mode transport system by event-driven simulation

Participants: Arnaud de La Fortelle, Sami Mahari.

The European project CATS — City Alternative Transport System — is developing and evaluating a new
vehicle system using a single type of vehicle for two different usages: individual use or collective transport.
Real experiments will necessarily take place with a limited number of vehicles and stations. Hence there is
a need for evaluation using simulations. We have been developing a discrete events simulator for that purpose,
based on a previous work done for collective taxis [42].
Our model relies on an adapted events/decision graph that extends previous graphs. The new feature of this model is the way we deal with two modes that can be extended to many other modes. This work therefore shows on a concrete example a method to efficiently merge multiple modes into one model.

- This year has seen the design and first implementation of the simulator.
- The results have been presented at a conference [29].

### 6.4.3. Multi-speed exclusion processes

**Participants:** Cyril Furtlehner, Jean-Marc Lasgouttes, Maxim Samsonov.

The slow-to-start mechanism is known to play an important role in the particular shape of the Fundamental diagram of traffic and to be associated to hysteresis effects of traffic flow. We study this question in the context of stochastic processes, namely exclusion and queueing processes, by including explicitly an asymmetry between deceleration and acceleration in their formulation. Spatial condensation phenomena and metastability are observed, depending on the level of the aforementioned asymmetry. The relationship between these 2 families of models is analyzed on the ring geometry, to yield a large deviation formulation of the fundamental diagram (FD)

This work has been presented at the TGF'11 conference [22], and a more extensive article is in preparation for a journal.

### 6.4.4. Dynamics of points of interest in a social game

**Participants:** Guy Fayolle, Jean-Marc Lasgouttes.

*Ma Micro Planète* is a geolocalized video game which entices players to use sustainable means of transport. At the heart of the game are community-driven points of interest (POI’s), or sites, which have a score that depends on the players activity. The aim of this work is to understand the dynamics of the underlying stochastic process.

We examine the system in the thermodynamic limit, as the number of players tends to infinity, the existence of which is proved under general conditions, where the probability of increasing the score of a visited POI is a function of the state of the system. Concerning the existence of a stationary regime, some complete answers are given for particular values of the parameters, and the existence of possible phase transition phenomena is enlightened.

A publication on the subject is in preparation.

### 6.4.5. Random walks in the quarter plane

**Participant:** Guy Fayolle.

In collaboration with K. Raschel (CNRS, Université F. Rabelais à Tours), we pursued the works initiated in 2010 in two main directions.

#### 6.4.5.1. The zero drift case

In several recent studies on random walks with small jumps in the quarter plane, it has been noticed that the so-called group of the walk governs the behavior of a number of quantities, in particular through its order. In the article [11], when the drift of the random walk is equal to 0, we provide an effective criterion giving the order of this group. More generally, we also show that in all cases where the genus of the algebraic curve defined by the kernel is 0, the group is infinite, except precisely for the zero drift case, where finiteness is quite possible.

#### 6.4.5.2. Counting and asymptotics

The enumeration of planar lattice walks, is a classical topic in combinatorics. For a given set $S$ of allowed unit jumps (or steps), it is a matter of counting the number of paths starting from some point and ending at some arbitrary point in a given time, and possibly restricted to some regions of the plane.

Like in the probabilistic context, a common way of attacking these problems relies on the following analytic approach. Let $f(i, j, k)$ denote the number of paths in $\mathbb{Z}^2_+$ starting from $(0,0)$ and ending at $(i, j)$ at time $k$. Then the corresponding CGF
Robotics - New Results - Project-Team IMARA

\[ F(x, y, z) = \sum_{i,j,k \geq 0} f(i, j, k)x^iy^jz^k \]
satisfies the functional equation

\[ K(x, y)F(x, y, z) = c(x)F(x, 0, z) + \overline{c}(y)F(0, y, z) + c_0(x, y), \]

where \( x, y, z \) are complex variables, although the time variable \( z \) plays somehow the role of a parameter. The question of the type of the associated counting generating functions, that is rational, algebraic, holonomic (solution of a linear differential equation with polynomial coefficients), was solved whenever the group is finite (see RA 2010). When the group is infinite, the problem is still largely.

It turns out that the nature of singularities play a deep important role in this classification. Making use of the general and powerful approach proposed in the book [3], a paper entitled Some exact asymptotics in the counting of walks in the quarter-plane has been submitted to AofA (International Conference on Analysis of Algorithms, Montreal, June 2012), in which a new approach is proposed to obtain some exact asymptotics for walks confined to the quarter plane.

6.4.6. Statistical physics and hydrodynamic limits

**Participants:** Guy Fayolle, Cyril Furtlehner.

Having in mind a global project concerning the analysis of complex systems, we first focus on the interplay between discrete and continuous description: in some cases, this recurrent question can be addressed quite rigorously via probabilistic methods.

To attack this class of problems, in touch with many applications domains (e.g. biology, telecommunications, transportation systems), we started from paradigmatic elements, namely the discrete curves subjected to stochastic deformations, as those mentioned for instance in [39]. After convenient mappings, it appears that most models can be set in terms of interacting exclusion processes, the ultimate goal being to derive hydrodynamic limits for these systems after proper scalings. We extend the key ideas of [39], where the basic ASEP system on the torus was the toy model. The usual sequence of empirical measures, converges in probability to a deterministic measure, which is the unique weak solution of a Cauchy problem.

The Gordian knot is the analysis of a family of specific partial differential operators in infinite dimension. Indeed, the values of functions at given points play here the role of usual variables, their number becoming infinite. The method presents some new theoretical features, involving promeasures (as introduced by Bourbaki), variational calculus, functional integration, and the construction of generalized measures. In [20], we present a detailed analysis of the ASEP system on the torus \( \mathbb{Z}/N\mathbb{Z} \). Then we claim that most of the arguments a priori work in higher dimensions (\( \text{ABC}, \) multi-type exclusion processes, etc), leading to systems of coupled partial differential equations of Burgers’ type. In the course of the study, several fascinating multi-scale problems emerge quite naturally, bringing to light some connections with the so-called renormalization in theoretical physics.
6. New Results

6.1. Visual tracking

6.1.1. 3D model-based tracking

Participants: Antoine Petit, Eric Marchand.

Our 3D model-based tracking algorithm [3] was used in various contexts. First, it has been studied and tested on a mock-up of a telecommunication satellite using a 6-DOF robotic arm, with satisfactory results in terms of accuracy of the pose estimation and computational costs [41], [42]. A potential application would be the final phase of space rendezvous mission using visual navigation. Then, it has been considered for designing a visual servoing scheme able to control the walking of a humanoid robot [29].

6.1.2. Omnidirectional stereovision

Participants: Guillaume Caron, El Mustapha Mouaddib, Eric Marchand.

Omnidirectional cameras allow direct tracking and motion estimation of planar regions in images during a long period of time. However, using only one sensor leads to plane and trajectory reconstruction up to a scale factor. We proposed to develop dense plane tracking based on omnidirectional stereovision to answer this issue. The method estimates simultaneously the parameters of several 3D planes along with the camera motion using a spherical projection model formulation [20].

6.1.3. Motion estimation using mutual information

Participant: Eric Marchand.

Our work with Amaury Dame related to template tracking using mutual information as registration criterion has been extended to motion estimation applications. It has been applied to mosaicing from an image sequence [28]. The main advantage is that this approach is robust to noise, lighting variations and does not require a statistically robust estimation process.

6.1.4. Augmented reality

Participants: Pierre Martin, Hideaki Uchiyama, Eric Marchand.

We developed an approach for detecting and tracking various types of planar objects with geometrical features[45]. We combine traditional keypoint detectors with Locally Likely Arrangement Hashing (LLAH) for keypoint matching. In order to produce robustness to scale changes, we build a non-uniform image pyramid according to keypoint distribution at each scale. It demonstrates that it is possible to detect and track different types of textures including colorful pictures, binary fiducial markers and handwritings. This approach was extended to consider non-rigidly deformable markers [46].

6.2. Visual servoing

6.2.1. Micro-manipulation

Participant: Eric Marchand.

We developed an accurate nanopositioning system based on direct visual servoing [43],[17]. This technique relies only on the pure image signal to design the control law, by using the pixel intensity of each pixel as visual features. The proposed approach has been tested in terms of accuracy and robustness in several experimental conditions. The obtained results have demonstrated a good behavior of the control law and very good positioning accuracy: 89 nm, 14 nm, and 0.001 degrees in the $x$, $y$, and $\theta_z$ axes of a positioning platform, respectively.
6.2.2. **Multi sensor-based control**  
**Participants:** Olivier Kermorgant, François Chaumette.

We have designed a generic sensor-based control approach to automatically tune the weights related to the features involved as inputs of a control scheme, allowing to take constraints into account. This scheme has been applied to several configurations, such as fusing the data provided by an eye-in-hand camera and an eye-to-hand camera, ensuring the visibility constraint, and avoiding the robot joint limits [30], [31], [32], [11].

6.2.3. **Visual navigation of mobile robots**  
**Participants:** Eric Marchand, Andrea Cherubini, Fabien Spindler, François Chaumette.

We have developed a visual servoing scheme based on the mutual information between the images acquired by an onboard camera and a visual memory to control the orientation of a vehicle during its navigation [27]. We have also fused the data provided by a pan-tilt camera and a laser range sensor for the autonomous navigation of a mobile vehicle while avoiding obstacles [23], [22]. Real experiments with our Cycab (see Section 5.4) have been conducted on Place de Jaude in Clermont-Ferrand in the scope of the ANR Tosa CityVIP project (See Section 8.2.1).

6.2.4. **Visual servoing for aircrafts**  
**Participants:** Céline Teulière, Eric Marchand, Laurent Coutard, François Chaumette.

A dynamic controller has been designed for the homing of a quadri-rotor aerial vehicle [39]. A color-based tracking algorithm has also been designed and combined with an image-based visual servoing for chasing a moving target from a a flying UAV [44]. Finally, a method has been developed to detect and localize an aircraft carrier in an image sequence, from which visual servoing control laws have been designed for the automatic landing [25], [26].

6.3. **Medical robotics**

6.3.1. **Visual servoing based on ultrasound images**  
**Participants:** Caroline Nadeau, Alexandre Krupa.

We developed a new approach of ultrasound image based visual servoing that directly uses the intensities of the ultrasound image pixels as visual features. This method that spares any segmentation or image processing time consuming step was initially proposed to control the 6 DOF of a conventional 2D probe for positioning and tracking tasks [38], [48]. To increase the tracking performance we also adapted this method by considering a predictive control law based on the periodicity of physiological motions [36]. Rigid motion compensation experiments were conducted in the context of the ANR USComp project (See Section 8.2.3). The method was also improved by estimating on-line the image 3D gradient required for the positioning task and extended for the use of a bi-plane ultrasound probe [37]. Finally, the use of a 3D motorized probe was also considered to compute directly the image 3D gradient and a comparison of the results obtained with the different probes (2D, bi-plan, 3D) was performed [12].

6.3.2. **Autonomous control modes for ultrasound probe guidance**  
**Participants:** Tao Li, Alexandre Krupa.
In the context of the ANR Prosit (See Section 8.2.2), we developed several autonomous control modes in order to assist a doctor during a robotized and teleoperated ultrasound examination (tele-echography). The robotic tasks we proposed concern: an automatic scanning of the patient by a 2D probe, a shared control mode that maintains the visibility of an anatomic element of interest while the doctor teleoperates the slave robot holding the 2D probe, an automatic positioning task that allows the doctor to retrieve a desired anatomic section that was previously captured by the doctor. The two latter modes are based on visual servoing schemes that use as input image moments extracted from the observed 2D ultrasound image. This extraction is performed thanks to an active contour (snake) based on Fourier descriptors that we developed and implemented on GPU in order to provide real-time performance [34],[47]. The proposed autonomous control modes were experimentally validated on the Lagadic medical robotics platform (see Section 5.3) and are now in the process of being integrated on the Prosit robot platform.

6.3.3. Real-time 3D ultrasound image reconstruction and 3D deformation tracking

Participants: Deukhee Lee, Alexandre Krupa.

We developed and implemented on GPU an algorithm that reconstructs in real-time a sequence of dense ultrasound volumes from a set of pre-scan 2D ultrasound images provided online by a motorized ultrasound probe [33]. Then we proposed a dense ultrasound tracking algorithm that estimates in real time both rigid and non-rigid motions of a region of interest observed in the sequence of reconstructed ultrasound volumes [33]. The algorithm consists in estimating in real-time, from intensity-value changes between successive 3D ultrasound images, motions of a set of 3D control points that describe the evolution of 3D Thin-Plate Splines (TPS) modeling the deformation. The estimated rigid motion was then used in a pose-based control scheme to automatically displace the probe held by a robot for soft tissue motion compensation. These works were conducted in the context of the ANR USComp project (See Section 8.2.3).
ARIANA Project-Team

6. New Results

6.1. Optical imagery for remote sensing

6.1.1. Phase Field-Higher Order Active Contours for Object Modelling and Image Segmentation

Participants: Ikhlef Bechar, Josiane Zerubia [contact].

This work is done in collaboration with Dr Ian Jermyn of Durham University (United Kingdom) and was funded by a contract with the EADS foundation [ http://www-sop.inria.fr/ariana/Ikhlef.Bechar/hoacs/index.html ].

The problem of object segmentation from imagery is an essential preliminary task for many applications (target recognition, automated navigation, organ segmentation in medical imaging, etc). The problem of adding prior knowledge about objects to the image segmentation process has received a lot of interest since recently, and active contours [38] provide us with such a tool.

We consider a new class of active contours called Higher Order Active Contours (HOACs) introduced initially in [41], and which consider an optimal contour as the one which minimizes an energy involving three additive terms; namely the length of a contour, its area and a term of interaction between all possible pairs of its points via an interaction function $\Psi(t)$. The three terms being weighed by means of three scalar parameters. The main advantage of HOACs over traditional segmentation methods is that they offer an unprecedented means for including shape prior about an object via the interaction function $\Psi(t)$.

The HOAC set up has been applied successfully to various object extraction problems such as the extraction of networks [40], circular shapes [37], etc, using a specific family of the $\Psi(t)$’s. Our main task in the framework of this project is to extend their work to more general shapes.

Our contributions so far have been in the numerical computation of the optimal HOAC parameters for a given shape. We have shown indeed that the HOAC energy can be fully made linear with respect to $\Psi(t)$, which then makes it easy to solve for $\Psi(t)$ numerically (cf. Fig. 6). This is achieved by first choosing a linear basis to represent $\Psi(t)$ and using K-K-T (Karush-Kuhn-Tucker) optimality criteria to express the fact that a target contour is a local minimum of the energy. Consequently, looking for the optimal values of the coefficients for a given shape amounts first to solve an eigen value problem, and second, to find the linear combinations of the found eigen vectors that satisfy both K-K-T minimality criteria. The computation being carried out in the Fourier domain for sake of computational efficiency. We are currently testing the proposed model on simple shapes such as the butterfly one shown in figure 6.

6.1.2. Optimization of the compression-restoration chain for satellite images

Participants: Mikael Carlavan, Laure Blanc-Féraud [contact].

This project involves the French Space Agency (CNES) and the CESBIO, on collaboration with TAS and I3S (Marc Antonini).

This work concerns the study of the optimal imaging chain in the context of satellite imaging. The main goal of this study is to propose a new method to address the problem of decoding-deconvolution-denoising and consists in a characterization and optimization of the compression/restoration processes considering the instrumental characteristics (FTM, noise, sampling). A theoretical study first showed that current processes of compression and restoration are better fitted if the restoration is performed on-board before the compression. Indeed, current restoration algorithm is designed to remove the blur and the instrumental noise but does not take into account the coding noise, and it is well-known that compression algorithms do not perform
Figure 6. An example of estimation of the $\Psi$ function for a given shape. (a) An example of a butterfly-like shape; (b) The numerical estimation of its optimal $\Psi$ function; (c) The result of a gradient descent algorithm on the HOAC energy with $\Psi$ of figure (b) until convergence (in black the original shape, and in red the shape found by the gradient descent algorithm). One can see that the estimated $\Psi$ makes the HOAC energy achieve a local minimum at a shape which is very close to the target shape of figure (a).
properly on noisy data. More generally, we concluded that the image should be the closest possible to the real image before the step of coding, encouraging, thus, to move the restoration step on-board before the compression. Figure 7 shows the global distortion w.r.t. the coding rate if the restoration step is done either before (on-board) or after (on-ground) the compression. We see that using an on-board restoration leads to a quality improvement of the final image regardless the coding rate. For example, at the usual coding rate of 2.5 bits/pixel, using an on-board restoration improves the quality of the final image about 0.5 dB.

![Figure 7. Global distortion w.r.t coding rate. The solid line is the global distortion with the restoration step done on-ground after the compression and the dashed line is the global distortion if the restoration is performed on-board before the compression.](image)

### 6.1.3. Aerial Image Restoration

**Participants:** Daniele Graziani, Laure Blanc-Féraud [contact].

*This project involves as partners: ATE, Coreti, and Gilles Aubert from the J.A.D. Laboratory at the University of Nice Sophia Antipolis.*

The goal of the project is to built an airborne camera system, and our part is to process aerial images provided by ATE: restoration, microscanning, video, color images. We investigate a convex variational framework to compute high resolution images from a low resolution video. We analyze the image formation process to provide a well designed model for warping, blurring, downsampling and restoration. The microscanning is modeled as a convex minimization problem, which is solved with a domain decomposition technique based on the recent work of M. Fornasier, A. Langer and C. Schonlieb. ("A convergent overlapping decomposition method for total variation Minimization", Numeriske Math. to appear), which allows parallel computing and a realization of a real time algorithm.

### 6.1.4. Contribution of object recognition on forest canopy images to the building of an allometric theory for trees and natural, heterogeneous forests

**Participants:** Jia Zhou, Xavier Descombes, Josiane Zerubia [contact].

*This work is done in collaboration with Dr. Pierre Couteron and Christophe Proisy at IRD, UMR AMAP, Montpellier.*
Individual tree detection methods are more and more present, and improve, in forestry and silviculture domains with the increasing availability of satellite metric imagery. Automatic detection on these very high spatial resolution images aims to determine the tree positions and crown sizes. The mathematical model based on marked point processes has showed advantages w.r.t. several individual tree detection algorithms for plantations. We used this detection method to analyze natural mangrove forests in French Guiana, eucalyptus plantations in Brazil, and other types of tropical forests. The simulated optical images were also used to improve the method and calibrate the detection parameters. To analyze a eucalyptus plantation in Brazil [23], we used 2 optical images acquired by the WorldView-2 satellite. A tentative detection simultaneously with 2 images of different dates (multi-date) was tested for the first time, which estimates individual tree crown variation during these dates. In this work, we tried to find the trees localizations and crown sizes in order to provide a plantation map, and estimate the tree crown growth during the period between 2 images, and compared these results with the field measurements and expected dynamics of corresponding populations. An example of multi-date detection result is showed in figure 9.

The detection method was also applied on simulated optical DART (Discrete Anisotropic Radiative Transfer) images, where exact field inventory could be provided on large surfaces. We assessed the detection results with these “ground-truth” maps.

6.2. SAR imagery for remote sensing

6.2.1. Stochastic modeling for very high resolution SAR image processing

Participants: Aurélie Voisin, Vladimir Krylov, Josiane Zerubia [contact].

This work is done in collaboration with DIBE, University of Genoa, with Dr Gabriele Moser and Prof. Sebastiano B. Serpico [http://spt.dibe.unige.it/] with partial financial support of the French Defense Agency, DGA [http://www.defense.gouv.fr/dga/] with partial financial support of the French Defense Agency, DGA [http://www.defense.gouv.fr/dga/]. The data are provided by the Italian Space Agency, ISA [http://www.asi.it/en].
We deal with the environmental risk assessment by addressing the problem of classifying SAR images of urban areas. Several difficulties need to be considered to address the SAR classification problem. The first one is related to the inherent multiplicative noise known as speckle, which degrades appreciably the registered imagery. Another difficulty is the heterogeneity of urban areas on very high resolution (VHR) images that leads to heterogeneous statistical modeling, reflecting the different ground materials such as asphalt, concrete, metal, etc. We propose a hierarchical statistical Bayesian supervised classification approach that consists of two steps. The first step deals with the SAR amplitude statistical modeling for each target class (e.g. vegetation, urban, etc.) by using a finite mixture model, estimated by resorting to a dictionary-based stochastic expectation maximization (DSEM) algorithm. More specifically, the SAR amplitude probability density functions (PDFs) are assumed to be mixtures of $K$ PDFs automatically chosen inside a predefined dictionary of SAR-specific distribution families. Such mixtures are intended to take into account the above mentioned VHR SAR statistics heterogeneity. We further consider an additional source of information obtained by extracting a textural feature map from the original SAR image in order to optimize the detection of urban areas. Typically, the textural feature is generated by using a Grey Level Co-occurrence Matrix (GLCM)-based method. The marginal PDFs of the original SAR image and the textural feature are combined via copulas, leading to a joint PDF for each class. On the second step the classification map is generated, using the joint copula-based statistics. To improve the robustness with respect to speckle noise, we consider a contextual model based on Markov random fields (MRFs), and, more specifically, a hierarchical MRF, which offers the possibility to take into account the multi-scale information and to deal with multi-resolution imagery [28]. A variety of algorithms were proposed to estimate the labels on hierarchical graphs. The consideration of a specific graph, here a quad-tree, allows to benefit from its good properties (e.g. causality) and to apply non iterative algorithms. Among the different algorithms employed in the literature, we chose to take into account an exact estimator of the marginal posterior mode (MPM). The cost function associated to this estimator offers the possibility to penalize the errors according to their number and the scale at which they occur: an error at the coarsest scale is stronger penalized than an error at the finest scale. Moreover, we introduce a prior estimation update that experimentally leads to improved results and is less affected by speckle noise when compared to a predefined prior [35]. The challenge of the problem considered here is that our given input is a single-polarized SAR image at a single resolution. To improve the classification, we extract an extra information in the form of a multi-scale wavelet decomposition from the initial image. Then, at each level, the textural feature map is obtained from each image in the decomposition stack. Finally, at each level, the wavelet image is combined with the textural image by using copulas, as described previously in Ariana activity reports. The hierarchical method was tested on real COSMO-SkyMed images. We illustrate the obtained results with an example of a SAR acquisition of the Port-au-Prince quay (Haiti). Spatially disjoint training and test areas were manually annotated. The classification is done following 3 classes: urban areas, natural landscape and wet areas. The
Vision, Perception and Multimedia Understanding - New Results - Project-Team ARIANA

results are shown qualitatively in figure 10. The computation of numerical results gives an average accuracy of 95.65 percent for the considered test areas.

![Image of initial SAR image of Port-au-Prince (Haiti) and classification map obtained with the hierarchical method for the 3 classes: Blue: water; Green: vegetation; Red: urban area.]

Figure 10. Left: Initial SAR image of Port-au-Prince (Haiti) (©ISA, 2009). Right: Classification map obtained with the hierarchical method for the 3 classes (Blue: water; Green: vegetation; Red: urban area).

Our previous work was based on single-scale MRF, thus the hierarchical approach is a direct extension. Part of our work was dedicated to the comparison of these two methods [35], [28], and we also compared the MRF-based model to a novel products of experts approach [18].

6.2.2. Parameter estimation procedures for HR SAR image classification

Participants: Vladimir Krylov, Josiane Zerubia [contact].

This work is conducted in collaboration with DIBE, University of Genoa with Dr. Gabriele Moser and Prof. Sebastiano Serpico [http://spt.dibe.unige.it/] with the support of the Italian Space Agency, ASI [http://www.asi.it/en].

Parameter estimation of probability density functions is one of the major steps in the mainframe of statistical image and signal processing. We have explored the properties and limitations of the recently proposed method of logarithmic cumulants (MoLC) parameter estimation approach which is an alternative to the classical maximum likelihood (ML) and method of moments (MoM) approaches. We have derived the general sufficient condition of strong consistency of MoLC estimates which represents an important asymptotic property of any statistical estimator [33]. We have demonstrated the strong consistency of MoLC estimates for a selection of widely used distribution families originating (but not restricted to) synthetic aperture radar (SAR) image processing. We have then derived the analytical conditions of applicability of MoLC to samples generated from several distribution families in our selection. We have conducted various synthetic and real data experiments to assess the comparative properties, applicability and small sample performance of MoLC notably for the generalized gamma and $K$ family of distributions. The synthetic-data experiments have demonstrated a competitive accuracy of MoLC estimates and a reliable behavior of this estimator for small samples which is a critical issue in applications. We have performed real-data image processing experiments to the problem of supervised classification applied to high resolution satellite SAR imagery. These experiments confirmed the stability of MoLC estimator with respect to sample size and at the same time illuminated the critical side of MoLC given by applicability restrictions. The experiments suggested the efficiency of use of the MoLC estimator for finite-mixture estimation problems [3], [19], and ML-based classification approaches [4], [28], [35].
6.2.3. Unsupervised amplitude and texture based classification of SAR images with multinomial latent model

Participants: Koray Kayabol, Aurélie Voisin, Vladimir Krylov, Josiane Zerubia [contact].

The participants would like to thank the Italian Space Agency (ASI) for providing the COSMO-SkyMed images. The TerraSAR-X images are provided from http://www.infoterra.de/.

We combine both amplitude and texture statistics of the Synthetic Aperture Radar (SAR) images using Products of Experts (PoE) approach for classification purpose. We use Nakagami density to model the class amplitudes and a non-Gaussian Markov Random Field (MRF) texture model with t-distributed regression error to model the textures of the classes. A non-stationary Multinomial Logistic (MnL) latent class label model is used as a mixture density to obtain spatially smooth class segments. The Classification Expectation-Maximization (CEM) algorithm is performed to estimate the class parameters and to classify the pixels [18]. Determining the necessary number of classes to represent the data and initialization are some drawbacks of the EM type algorithms. In [17] and [32], we combine hierarchical agglomeration, CEM and Integrated Classification Likelihood (ICL) criterion to get rid of the drawbacks of EM. We obtained some classification results of water, land and urban areas in both supervised and unsupervised cases on TerraSAR-X, as well as COSMO-SkyMed data [18], [17], [32]. The proposed unsupervised ATML-CEM (Amplitude and Texture density mixtures of MnL with CEM) method provides significantly better results, see Fig. 12, compared to the corresponding results obtained with K-MnL and its performance is close to supervised ATML-CEM.

6.3. 3D-modelling of urban scenes

6.3.1. Building reconstruction from aerial LiDAR data

Participants: Yannick Verdie, Florent Lafarge [contact], Josiane Zerubia.

The generation of 3D representations of urban environments from aerial and satellite data is a topic of growing interest in image processing and computer vision. Such environments are helpful in many fields including urban planning, wireless communications, disaster recovery, navigation aids, and computer games. Laser scans have become more popular than multiview aerial/satellite images thanks to the accuracy of their measurements and the decrease in the cost of their acquisition. In particular, full-waveform topographic LIDAR constitutes a new kind of laser technology providing interesting information for urban scene analysis. We study new
Figure 12. (a) The original SAR image (COSMO-SkyMed, ©ASI), (b), (c) and (d) classification maps obtained by K-MnL, supervised and unsupervised ATML-CEM methods. Blue, red and green colors represent water, urban and land areas, respectively.
stochastic models for analysing urban areas from LIDAR data. We aim to construct concrete solutions to both urban object classification (i.e. detecting buildings, vegetation, etc.) and the 3D reconstruction of these objects. Probabilistic tools are well adapted to handling such urban objects, which may differ significantly in terms of complexity, diversity, and density within the same scene. In particular, jump-diffusion based samplers offer interesting perspectives for modelling complex interactions between the various urban objects. We investigated a first approach aiming at producing accurate, watertight and compact meshes from planar patches under planar constraint especially designed for urban scenes. The LiDAR point cloud is classified through a non-convex energy minimization problem. The planar structures are extracted and connected to generate a compact, and watertight mesh of the building. Experiments highlight the potential of our method in term of performance, compactness, and accuracy. This work has been published in [22]. We illustrated our results in figure 13.

Figure 13. Results of 3D urban reconstruction by the framework described in [22].

6.3.2. Modeling large urban environments from unstructured point clouds

Participant: Florent Lafarge [contact].

We present a robust method for modeling cities from unstructured point data. Our algorithm provides a more complete description than existing approaches by reconstructing simultaneously buildings, trees and topologically complex grounds. Buildings are modeled by an original approach which guarantees a high generalization level while having semantized and compact representations. Geometric 3D-primitives such as planes, cylinders, spheres or cones describe regular roof sections, and are combined with mesh-patches that represent irregular roof components. The various urban components interact through a non-convex energy minimization problem in which they are propagated under arrangement constraints over a planimetric map. We experimentally validate the approach on complex urban structures and large urban scenes of millions of points as illustrated on Figure 14.

6.3.3. Parallel Monte Carlo sampler for point processes

Participants: Yannick Verdie, Florent Lafarge [contact], Ioan Dragan.

We designed a new parallel scheme for Markov point processes. These probabilistic models exploit random variables whose realizations are configurations of parametric objects, each object being assigned to a point positioned in the scene. The number of objects is itself a random variable. Another strength of Markov point processes is their ability to take into account complex spatial interactions between the objects and to impose global regularization constraints. Moreover, we proposed to use space-partitioning tree such as quadtree (for 2D data) or octree (for 3D data) for non-homogeneous measure adapted to the problem. We illustrate the results in figure 15.
Figure 14. Reconstruction of the cities of Marseille and Amiens, France, from Lidar point cloud.

Figure 15. Results of the new Marked Point process on (left) 2D data, and (right) LiDAR data.
6.4. Biological imagery

6.4.1. Regularizing parameter estimation with Poisson noise

Participants: Mikael Carlavan, Laure Blanc-Féraud [contact].

The problem is to automatically estimate the regularizing parameter in Poisson noisy image deconvolution using the $L_1$-norm regularization as a total variation or frame coefficients. This problem is addressed using the discrepancy principle. The standard weighted criterion composed of a data term and a regularization term is rewritten as a constrained minimization problem. The constraint is designed on the data term using the discrepancy principle and a new estimation of the bound is proposed as well as an efficient algorithm to solve this constrained minimization problem. This work is published in [10], [24], [9].

6.4.2. Brain vascular network segmentation

Participant: Xavier Descombes [contact].

This work was conducted in collaboration with Franck Plouraboué and Abdelhakim El Boustani from IMFT Toulouse and Caroline Fonta from CerCo Toulouse. It has been partially supported by a PEP II project from CNRS.

Micro-tomography produces high resolution images of biological structures such as vascular networks. We have proposed a new approach for segmenting vascular network into pathological and normal regions from considering their micro-vessel 3D structure only. We consider a partition of the volume obtained by a watershed algorithm based on the distance from the nearest vessel. Each region is characterized by its volume and the local vascular density. The volume and density maps are first regularized by minimizing the total variation. Then, a new approach has been proposed to segment the volume from the two previous restored images based on hypothesis testing. Results are presented on 3D micro-tomographic images of the brain micro-vascular network (see Fig. 16).

![Figure 16. Brain micro-vascular network segmentation](image-url)

6.4.3. Blind restoration of 3D biological image

Participants: Saima Ben Hadj, Laure Blanc-Féraud [contact].
Fluorescence microscopy is a powerful imaging technique providing three-dimensional images of biological living specimen. However, these images are degraded by a depth variant blur due to light diffraction phenomenon as well as refractive index mismatch between the different mediums composing the system and the biological sample. They are also distorted with noise from non-ideal imaging conditions. In order to provide biologist with more suitable images for quantitative studies, many restoration methods were developed. In most of them, the blur function, called Point Spread Function (PSF) is assumed to be piecewise constant in order to avoid the intensive computing time when using a pointwise varying PSF. However, this usually leads to blocking effect in the restored image. In our work, we extend the Space varying (SV) blur model previously proposed in [39] for 2D astronomical images to 3D microscopy images. In that model, the degraded image is a convex combination of convolutions with a space-invariant (SI) PSF. Furthermore, we fit to that model two restoration procedures which are basically developed for a SI PSF. On the one hand, we use the Richardson-Lucy method with Total Variation regularization which is carried out under Poisson noise assumption in order to restore confocal microscopy images. We employ another method with total variation regularization adapted to images with an additive Gaussian noise in order to restore Wide Field Microscopy images. For that, we rely on a fast optimization method based on a domain-decomposition technique [36]. In particular, we study its convergence properties when using the SV blur model [31], [27]. To illustrate the interest of the proposed method, we show in Fig 17 some results obtained on a simulated bead image of Wide Field Microscopy.

Figure 17. (X,Z) slices of the (a) original image, (b) degraded image (c) restored image using a SV blur model, and (d) restored image with a space-invariant PSF.

6.4.4. Axon imaging

Participant: Florence Besse [contact].

During brain development, neurons extend cellular processes (dendrites and axons) to connect to specific targets and establish functional networks. Understanding how axonal processes migrate to reach their targets and how they form new branches to build up a complex axonal tree is thus key. To characterize the properties of axonal trees, we have generated a collection of confocal 3D pictures of normal and mutant single axons labeled using a fluorescent protein. To analyze the formation of axonal trees in real-time, we have developed a protocol to dynamically image growing axons within intact Drosophila brains. In this protocol, entire brains are cultured in conditions where they can undergo cell differentiation and maturation. Axons are labeled by the fluorescent molecule GFP and are imaged over 12h, with very low-photobleaching and no associated phototoxicity, using an ultra-sensitive 2-photon microscope. Several image sequences corresponding to the growth of axons in normal conditions have been acquired (see Fig. 18).

6.4.5. Detection of Axons in Neuronal Images

Participants: Alejandro Mottini, Florence Besse, Xavier Descombes [contact].
Figure 18. Axon growth imaging

Imaging techniques such as confocal and two-photon microscopy provide an efficient way of analyzing supra-cellular structures. It is known that the precise shape of these objects provides information on their functioning and allows the characterization of pathological states. Therefore, the analysis of the morphological differences between normal and pathological structures is of paramount importance. In particular, the analysis of neuronal axon topologies allows biologists to study the causes of neurological diseases such as Fragile X Syndrome and Alzheimer’s disease. For this purpose biologists have acquired static 3D images of mature neuron axons using fluorescent confocal microscopy.

Figure 19. Static axon images (2D maximum intensity projection).

Due to the high volume of generated image data and the tortuous nature of the axons, manual processing is infeasible. Therefore, it is necessary to develop techniques for the automatic extraction and analysis of the neuronal structures. However, since both types of images present different characteristics, two different methodologies need to be developed. The main objective of the static case is the study of the length and number of bifurcation points of the two populations of neurons. On the other hand, the focus in the dynamic
case is put on the tracking of the axonal tips. The automatic extraction of axons from confocal microscope images is a key problem in the field of neuron axon analysis. In this work we propose a solution which combines algorithms for the denoising, binarization, skeletonization, gap filling, point detection and statistical analysis in a pipeline capable of extracting the axons. Furthermore, these algorithms were incorporated in a software, developed in Matlab, which includes an easy to use GUI along with functionalities to facilitate high data throughput analysis. The software was evaluated on several 3D confocal microscope images of normal and mutated axons. Our results support the potential use of the software in helping biologist perform automatic morphological analysis of axons in fluorescent confocal microscopy images.

Figure 20. Original (left) and extracted (right) normal axon image (2D maximum intensity projection).

6.4.6. Axons tracking

Participants: Alejandro Mottini, Huei-Fang Yang, Florence Besse, Xavier Descombes [contact].

This work was funded by project ARC-DADA (INRIA/CNRS/UNSA) [http://www-sop.inria.fr/members/Xavier.Descombes/DADA/home.html]. It is done in collaboration with Serpico team at INRIA Bretagne (C. Kervrann, P. Houllier)

To study axon growth process biologists have acquired dynamic 3D+t image sequences of developing neurons using fluorescent two-photon microscopy.

Live cell two-photon microscopy is an effective tool for the analysis of dynamical processes occurring in living samples that, when combined with fluorescence, allows the detection of objects of interest in 3D space and time. These labeled objects appear as bright spots which need to be detected. The low resolution and signal-to-noise-ratio (SNR) make this analysis difficult even for experienced biologists. As a consequence, automatic detection techniques need to be robust and flexible. To this end, the Marked Point Process (MPP) detection framework was selected. Since axonal extremities appear either as circular objects or as ending points of thin filaments, we proposed two different MPP models. These models were formulated using a Gibbs energy function and optimized with Multiple Births and Deaths, a newly proposed algorithm which guarantees a fast convergence to the global minimum. The first model is designed to detect spheres or disks and the second filaments (both in 3D). Both models were tested on several 3D static images. To evaluate the performance of the detection, results were compared against images labeled by an experienced biologist.

Figure 22 shows the results obtained on one frame of a video sequence (the shown image is a slice of the 3D frame) for the disk/sphere model. One can appreciate that the axonal tip marked by the biologist (in white, left) was detected. However, a false positive is also present.
Figure 21. Dynamic images (2D maximum intensity projection).

Figure 22. Original (left) and resulting (right) images (detected disks in pink). The true axonal tip is circled in white (original image) and red (result). In the original image, axons are labeled in white.
Figure 23 shows the result obtained on one of the slices (for the same video frame) for the filament model. Although the good filament was detected, many false positives are also present in the image. Similar results were obtained for other images.

Figure 24 shows the preliminary results obtained by applying the particle filtering technique on the coronal KESM (Knife-Edge Scanning Microscopy) sections of the mouse cerebellum, where the green rectangles indicate the tracked objects.

Once the detection of the extremities in each frame of the video has been solved, the solution should be embedded into a tracking algorithm to obtain an estimation of the axon’s trajectory during the growing stage. To this end, the particle filtering technique was considered. This technique consists in estimating the posterior distribution of the current state $x_t$ of the target of interest at time $t$ based on the measurements $z_1^{t}$:

$$p(x_t | z_1^{t}) \propto p(z_t | x_t) \int p(x_t | x_{t-1}) p(x_{t-1} | z_1^{t-1}) \, dx_{t-1},$$

(1)

where $p(x_t | x_{t-1})$ is the transition distribution, and $p(z_t | x_t)$ is the likelihood. In the current implementation, a simple kinematic model is used for the transition distribution, and a color histogram is applied to the computation of the likelihood. Figure 24 shows the preliminary results obtained by applying the particle filtering technique on the coronal KESM (Knife-Edge Scanning Microscopy) sections of the mouse cerebellum, where the green rectangles indicate the tracked objects.
The next step will be to design a more sophisticated transition distribution and likelihood model that are suitable for tracking the trajectories of axons during their growth.

6.5. Dermatology

6.5.1. Statistical analysis of skin pigmentation under treatment

Participants: Sylvain Prigent, Xavier Descombes, Josiane Zerubia [contact].

This work was partially funded by a contract with Galderma R&D [http://www.galderma.com/RampD.aspx].

One of the steps to evaluate the efficiency of a therapeutic solution is to perform measurements on a series of patients who received the studied treatment. In parallel another treatment is tested on another group of people or on the same group of patients on another skin area. This second treatment is the reference one for the studied pathology or a placebo. We will call it ‘vehicle’.

For facial hyper-pigmentation, for each studied treatment, a group of $N_e$ patients receives the treatment on one cheek and the vehicle on the other. To this end, patients are selected to have the same hyper-pigmentation severity on the two cheeks. Then multi-spectral images are taken at different times $t$ along the treatment period.

We propose a methodology to estimate the efficiency of a treatment by calculating a spectral criteria that maximizes the visibility of the disease comparatively to a healthy reference area. To design such a criterion, we compare three approaches. The first one gives weights to spectral bands in order to get an equivalent of the luminance from the $CIEL^*a^*b^*$ decomposition. This is the standard measure in dermatology. We compare this spectral measurement to a spectral signature obtained by ICA (Independent Components Analysis) in a whole study, and a criterion that searches for the highest contrasted band.

Once a criterion is designed, we compute the hyper-pigmentation severity of a patient by normalizing the pathological area with the healthy area and the active treatment measurement by the vehicle. That gives severity measurement distributions from patients at different times. A Student paired test allows to determine if an active treatment has an effect between two measurement times. The experiments done on 3 treatments and their associated vehicles brought to the following conclusion: The severity measure based on the selection of an optimal band allows to detect a treatment effect sooner than the two other approaches. The severity measurements obtained by the optimal band method is illustrated in figure 25.

![Figure 25. Evolution of skin pigmentation in the time and its quantification.](image-url)
6. New Results

6.1. Feature space modelling

6.1.1. A novel shape boundary based description for leaf identification

Participants: Itheri Yahiaoui, Olfa Mzoughi, Nozha Boujemaa.

The problem of automatic leaf identification is particularly difficult for two main reasons: (i) the first is the enormous number of leaf species and (ii) the second, which is relevant for some special species but more complex, is the high inter-species and the low intra-species similarity.

Our research has focused on analysing leaf morphology in order to determine a numeric key description for leaf species robust to all the above mentioned constraints. The approach that we propose is a shape boundary description that combines two complementary information: (i) the first one outlines local variations of the leaf margin. This is performed using the Directional Fragment Histogram (DFH), introduced in [25], which encodes the relative frequency distribution of groups of contour points with uniform orientation, (ii) the second property emphasizes the spatial distribution of contour points (in terms of distances). This is done by comparing the shape to standard geometric ones (such as circle, rectangle, ellipse, convex hull, etc.).

This descriptor was evaluated within the framework of ImageCLEF 2011 plant task where a crowd-sourced database, called Pl@ntLeaves [13], was used and a high number of image retrieval techniques was tested (a total of 8 groups from all around the world that have submitted 20 runs [19]). Our descriptor brought the best rate for scan-like pictures and was close to the best rate for scan pictures. Besides to the accuracy, this descriptor requires very low computational time, which accomplishes a basic condition for real world application.

6.1.2. Visual-based plant species identification from crowdsourced data


Inspired by citizen sciences, the main goal of this work is to speed up the collection and integration of raw botanical observation data, while providing to potential users an easy and efficient access to this botanical knowledge. We therefore designed and developed an original crowdsourcing web application dedicated to the access of botanical knowledge through automated identification of plant species by visual content.

Technically, the first side of the application deals with content-based identification of plant leaves. Whereas state-of-the-art methods addressing this objective are mostly based on leaf segmentation and boundary shape features, we developed a new approach based on local features and large-scale matching. This approach obtained the best results within one sub-task of ImageCLEF 2011 plant identification benchmark [19]. The second side of the application deals with interactive tagging and allow any user to validate or correct the automatic determinations returned by the system.

Overall, this collaborative system allows to enrich automatically and continuously the visual botanical knowledge and therefore to increase progressively the accuracy of the automated identification. A demo of the developed application was presented at ACM Multimedia conference [13]. This work was done in collaboration with INRIA team ZENITH and with the botanists of the AMAP UMR team (CIRAD). It is also closely related to a citizen science project around plant’s identification that we developed with the support of the TelaBotanica social network inside the Pl@ntNet project.

6.1.3. Spatial relations between salient points on a leaf

Participants: Sofiène Mouine, Itheri Yahiaoui, Anne Verroust-Blondet.
Figure 1. Normalized classification scores for scan images (up) and scan-like images (down). Our approach bin is in red, other ImageCLEF methods in blue.
In the scope of the Pl@ntNet project, our recent work has consisted in finding spatial relationships between salient points on a leaf. As a first step, classic detectors were used to find significant points in the leaf area and then the Shape context descriptor, originally applied on contour points, was introduced to measure a spatial relation between interest points. We have tested different configurations by varying the set of voting points. First results confirm that including spatial relations enriches the local description of each point. We are currently improving a veins and landmark extraction approach [12] in order to include also veins points in the voting set.

6.1.4. 3D mesh segmentation by example

Participants: Esma Elghoul, Anne Verroust-Blondet.

In recent years, there has been an increasing interest for automatic 3D segmentation. Indeed, segmentation of 3D objects is an important step in many applications such as part indexing of 3D objects, pattern recognition, compression, morphing, texture mapping and simplification. It refers to the process of partitioning 3D shapes into multiple parts, based on semantic criteria and/or geometric criteria.

Our work consists in introducing an approach to segment a 3D object class referring to a given segmented object from this class (we called it segmentation by example). The considered segmentation method is not automatic: we want to use interactive tools that proved advantageous to segment a 3D shape into relevant parts.

As a first task, we reviewed the state of the art in 3D segmentation techniques recently proposed in the literature. The different techniques were evaluated and classified for the purpose of choosing the more appropriate one for our work. We opted for extending the technique of random walks [24] to build an interactive tool of 3D segmentation.

For the second task, we had to solve a basic problem: that of similarly direct objects belonging to the same class. Indeed, each 3D model is provided in a random orientation in the space. In order to align objects of a same class, we used the alignment approach developed in [22] which computes 3 alignment axes. To properly orient our objects between them, we had to develop an additional process to the last one. It combines a 2D ICP and a 3D ICP approaches to give the best orientation among 48 possibilities and pair each two objects meshes.

So having a user-supplied already segmented model (model (1)) and a second model (model (2)) belonging to the same class (not segmented but similarly oriented), we developed a method to put into correspondence segmented parts of model (1) with faces of model (2). Then we computed a segmentation of model (2) using a derivative approach of the random walks. We applied this technique as well to segment all the objects that belong to the class of model (1). Our approach provides good results for manufactured object classes such as chairs and tables.

6.2. Feature space structuring

6.2.1. Random Maximum Margin Hashing

Participant: Alexis Joly.

Following the success of hashing methods for multidimensional indexing, more and more works are interested in embedding visual feature space in compact hash codes. Such approaches are not an alternative to using index structures but a complementary way to reduce both the memory usage and the distance computation cost. Several data dependent hash functions have notably been proposed to closely fit data distribution and provide better selectivity than usual random projections such as LSH. However, improvements occur only for relatively small hash code sizes up to 64 or 128 bits due to the lack of independence between the produced hash functions. In this work, we introduced a new hash function family that attempts to solve this issue in any kernel space. Rather than boosting the collision probability of close points, this method focus on data scattering. By training purely random splits of the data, regardless the closeness of the training samples, it is indeed possible to generate consistently more independent hash functions. On the other side, the use of large margin classifiers allows to maintain good generalization performances. Experiments did show that our
new Random Maximum Margin Hashing scheme (RMMH) outperforms four state-of-the-art hashing methods, notably in kernel spaces. Overall, this new concept of randomly trained classifiers opens the door to many other problems including large-scale learning, visual vocabulary construction or distributed content-based retrieval methods. A paper describing RMMH was published in the proceedings of CVPR 2011 [14].

6.2.2. Scalable information retrieval in distributed architectures

Participants: Mohamed Riadh Trad, Alexis Joly, Nozha Boujemaa.

Organizing media according to the occurrence of real-life events is attracting increasing interest in the multimedia community. However, whereas text based methods are now mature enough to deal with huge datasets, there are still some challenging issues managing multimedia contents. This becomes even more challenging in the context of User Generated Contents. Low-level visual metadata are indeed not simple textual or scalar values, their management requires efficient similarity search in high dimensional spaces.

Similarity search in high dimensional spaces has been the focus of many works in the database community in the recent years. State-of-the-art methods focus mainly on space partitioning techniques and more recently on hash-based probabilistic algorithms.

Although, hash-based approaches proved to be scalable, the computational cost is still too high for some real world applications and K-Nearest Neighbours Graph constructed can be more desirable than the costly online K-NN search. In fact, the basic LSH algorithm partitions the space uniformly and thus it does not exploit the clustering property of the data, which may result in slow query response and wasted space with additional hash tables. These limitations were pointed out with our scalable prototype for large scale event matching [18].

Scaling up LSH-based techniques and applications is then closely related to buckets occupations and objects distribution within the index structure. Recent works achieve better data distribution over the buckets with guarantees on occupation. As one result, we easily bound the similarity join size and evaluate bound algorithms complexity.

Based on these works, we designed and implemented a scalable prototype for distributed similarity search and K-NN graph construction. We have made several experiments querying real world large datasets. The prototype proved to be efficient for both search and K-NN graph construction.

Ongoing experiments process a 1.2 million images dataset. Results will be submitted for publication.

6.2.3. Visual similarity sensitive hashing methods for semantic image search in very large collections of images


With the rapid development of information acquisition technology, we have witnessed an explosive growth in the scale of shared data collections. Then, it is now possible to tackle fundamental problems with very large datasets’ context. Especially those addressing challenging tasks in machine learning for developing large scale approaches for multimedia retrieval and mining. Computer Vision is experiencing this paradigm shift, with large annotated image and video datasets becoming available. Indeed, various benchmark datasets for image classification have been released such as image-net and LabelMe. Therefore, a key challenge is taken up through out the Phd aiming to build efficient methods for training and matching efficiently very large collections of images.

We proposed several SVM-based strategies to build new supervised hash function families from large annotated collections of features. We indeed investigated with an approach consisting in benefitting from different embedding approaches in order to build compact codes indexed with efficient similarity search structures. Therefore, we have extended a kernelized hashing method [14] with multi-class SVM to solve a K-class classification problem by choosing the maximum applied to the outputs of K SVMs. We indeed proposed hashing methods based on the multi-class SVM classification strategies: One vs One (OVO) And One vs All (OVA). An important task during this process was to experimentally evaluate the quality lost induced by such representations with respect to the efficiency gains. We then compared multi-class SVM strategies with different underlying kernels.
Inspired by state of the art hashing in kernel space methods we investigated an approach consisting of benefiting from both semantic hashing like techniques and kernel embedding approach in order to build compact category aware codes indexed with efficient similarity. Experiments, are performed on image-net ILSVRC 2010 dataset \[23\]. Results will be submitted for publication.

6.3. Pattern recognition and statistical learning

6.3.1. Machine identification of biological shapes

**Participants:** Asma Rejeb Sfar, Donald Geman, Nozha Boujemaa.

Stored images of biological objects are accumulating at a staggering rate due to new sensor technologies, expanding use in medical diagnostics, web-based search engines and growing demands for web-based services in traditional sciences such as botany. These developments have been accompanied by an increasing demand for the automated analysis of these data, such as counting cell types, detecting lesions and other abnormalities in medical images, and identifying botanical shapes.

All these tasks have one feature in common: massive diversity among the shapes. Indeed, such shapes display enormous within-class variation and are generally highly deformable. Also, they often exhibit a hierarchical organization resulting from evolutionary processes.

There is currently no existing methodology in image analysis and computer vision which can be applied to a multi-class shape recognition problem of this complexity. Consequently, there is a need for a new, generic methodology for categorizing hierarchically-structured families of deformable shapes, particularly when both the number of categories and the within-category variation are very large.

We proposed a coarse-to-fine (CTF) approach in both shape representation and image parsing. The representation is hierarchical in both class and pose.

We focused on botanical shapes, specifically categorizing simple leaves according to species. So, we determined a suitable representation for the pose of a simple leaf and designed and tested a two-stage pose detector. Then, we constructed classifiers based on the plant taxonomy.

Results will be submitted for publication.

6.4. Interactive search and personalisation

6.4.1. Database denoising and multi visual queries

**Participant:** Sébastien Poullot.

One of IMEDIA’s tasks inside the SCARFACE project is to introduce and develop a character retrieval system. For this purpose, we take as entries the tracking of the persons in video sequences computed by Thalès and construct a database of the profiles. A profile is a 3D frame, a bounding box that changes along the time line. Two original works have been proposed for searching in the profile database.

The first one consists in analysing features of each profile with respect to all the profiles in order to extract relevant features from it, and construct more representative databases.

The second one is to be able to search inside the database with a set of queries (pictures of the same person). An a priori work can be done on this set of queries in order to extract the relevant features (and remove the irrelevant ones). On the other side an a posteriori work can be done on late merging depending on the specificities of each sub query.

**TRECVID Instance Search 2011**

Before starting the developments for SCARFACE, we tested various algorithms in TRECVID 2011 INS (instance search) task. This task is close to the SCARFACE one: from a set of captures of one object, one should find its occurrences in a set of video sequences. This work has been done during the stay of Sébastien Poullot at NII (the Japan National Institute of Informatics) in July and August 2011.
The differences with SCARFACE are:

- a high diversity in the type of the objects (people but also, places, vehicle, animals, etc),
- the location of the object in the database is not given.

Our approach obtains good results (above the median scores of all teams) and works in a very short time (and without indexing system for speeding up the process) [15]. The choice for SCARFACE’s method partially depends on these results. We still continue on the INS task in order to achieve better scores (various descriptors and various post and late fusion between sub queries).

Query generative models

Moreover, in order to enhance visual query results, we want to create some visual query generative models. It is directly linked to SCARFACE (a priori processes) and TRECVID works: given a set of images (considered as queries), we extract what gather them and what separate them in order to construct artificial relevant queries. For now we essentially work on some logo databases.

6.4.2. Object-based Visual Query Suggestion


After our work on the shared neighbours clustering methods in multi-sources case published in [10], we are interested now to the case of a bipartite graph that we apply to object-based visual query suggestion using the visual words mining technique [16]. In fact, state-of-the-art visual search systems allow to retrieve efficiently small rigid objects in very large datasets. They are usually based on the query-by-window paradigm: a user selects any image region containing an object of interest and the system returns a ranked list of images that are likely to contain other instances of the query object. User’s perception of these tools is however affected by the fact that many submitted queries actually return nothing or only junk results (complex non-rigid objects, higher-level visual concepts, etc.). We address the problem of suggesting only the object’s queries that actually contain relevant matches in the dataset. This requires to first discover accurate object’s clusters in the dataset (as an off-line process); and then to select the most relevant objects according to user’s intent (as an online process). We therefore introduce a new object’s instances clustering framework based on two main contributions: efficient object’s seeds discovery with adaptive weighted sampling and bipartite shared-neighbours clustering. Experiments show that this new method outperforms state-of-the-art object mining and retrieval results on OxfordBuilding dataset. We finally describe two object-based visual query suggestion scenarios using the proposed framework and show examples of suggested object queries.

6.4.3. Interpretable Visual Models for Human Perception-based Object Retrieval


Understanding the results returned by automatic visual concept detectors is often a tricky task making users uncomfortable with these technologies. In this work we attempt to build humanly interpretable visual models, allowing the user to visually understand the underlying semantic. We therefore proposed a supervised multiple instance learning algorithm that selects as few as possible discriminant local features for a given object category. The method finds its roots in the lasso theory where a $L_1$-regularization term is introduced in order to constraint the loss function, and subsequently produce sparser solutions. Efficient resolution of the lasso path is achieved through a boosting-like procedure inspired by B-Lasso algorithm. Quantitatively, the method achieved similar performance as current state-of-the-art, and qualitatively, it allows users to construct their own model from the original set of patches learned, thus allowing for more compound semantic queries. This work is part of the PhD of Ahmed Rebai [8] and it was published in ICMR 2011 proceedings [17]. This work was then extended to using geometrically checked feature sets rather than using single local features to describe the content of visual patches. We did show that this allows drastically reducing the number of the selected visual words while improving their interpretability. A publication was submitted to pattern recognition journal [11].
6.4.4. Relevance feedback on local features: Application to plants annotations and identification

Participants: Wajih Ouertani, Michel Crucianu, Nozha Boujemaa.

As biological image databases are increasing rapidly, automated species identification based on digital data is of great interest for accelerating biodiversity assessment, researches and monitoring. In this context, our work falls within an investigation of computer vision techniques or more precisely: object recognition and content based image retrieval techniques to help botanist identifying and organizing his digital images’ collections. Under believe that perception, recognition and decision are parts of human skills, this work focus on an interactive mechanism which tries to extract useful information from the user and gives him help to deal with large data amount. We adopted an explicit relevance feedback (RF) schema and we worked on extending it to deal with local intention through local features (LF) description. This mechanism helps discovering and dynamically defining new concept and interesting plant parts and feed identification ways interactively. Moreover since it relies to the content rather than labels one direct application is to fill the initially sparse annotation space with right annotations and in a reasonable time and with the introduce of one or many expertises. We recently explored and tested images local features matching involving high order features and non-rigid adaptation tentative to structure database with a patterns’ discovery stage. Using those type of methods we expect to introduce a high level appearance information that tends to go beyond classical bag of features and histogram based distances at least from semantic gap and interpretation point of view. We argue our exploration way with the fact that initial search space can be exceedingly rich. By pre-structuring it we can hope to obtain a smaller search space together with more reliable inference. Also learning parts interactively with localized local features may require a lot of interaction since it requires a considerable number of examples. We experienced the design of combined machine learning and prior mining of matches which we are actually improving.

6.5. Software

6.5.1. IKONA/MAESTRO software


This year, IKONA has been extended in the context of Pl@ntNet, Glocal, I-SEARCH and R2I projects. For the Pl@ntNet project, along the continuing improvements in the MAESTRO software, a number of new features were added. Namely the support for the automatic image segmentation and subsequent use of segmented regions; descriptors with the various shape’s geometric parameters; use of multiple orientations for Harris points; run-time additions to the external database and immediate availability of the new images for the search; descriptors to facilitate external data usage; colour SIFT and Affine Covariance descriptors; integration of the thesis work of Ahmed Rebai for objects retrieval; and tools for statistical tests.

In addition, a number of new web services were developed and deployed: the dynamic indexation system of the on-line pad (“carnet-en-ligne” of Tela Botanica)images; the search with multiple views; the update of Pl@ntNet internal demonstration allowing to present features such as visual similarity search, textual search, filtering (pre- and post-filtering), and different methods of research; the implementation of the organ prediction web service and other web services of botanical information statistics; the administration of the indexation system and the experimentation of new research methods (GPS spatial and temporal search).

For the Glocal project, an interface was developed for the demonstrations of a search engine in large scale events database (the queries are event images and the result is a list of the closest events in terms of time alignment and image content), and new web services were developed and updated according to the data exchange format and the middle-ware of the project - among others: fraud detection, import media from the web, associate media with existing event in the repository, and event matching web services. The queries are composed of either a medium link (an external image) or an event link (set of external images).
For the I-SEARCH project, an integration was performed to provide global and local 2D image low level descriptors. For videos, an automated extraction of visual words tool was integrated to show to users image patches which are the most meaningful.

For the R2I project, a detailed technical documentation of the procedure of maestro’s installation, web services and tomcat server were provided to Exalead partner.
6. New Results

6.1. Large-scale image search

6.1.1. Aggregating local image descriptors into compact codes

Participants: Matthijs Douze, Hervé Jégou [INRIA Rennes], Patrick Pérez [Technicolor], Florent Perronnin [Xerox RCE], Jorge Sánchez [Xerox RCE], Cordelia Schmid.

In [5] we consolidate and extend earlier results for large-scale image search. Different ways of aggregating local image descriptors into a vector are compared. The Fisher vector, see Figure 1, is shown to achieve better performance than the reference bag-of-visual words approach for any given vector dimension. Furthermore, we jointly optimize dimensionality reduction and indexing in order to obtain a precise vector comparison as well as a compact representation. The evaluation shows that the image representation can be reduced to a few dozen bytes with good search accuracy. Given such small codes, searching a 100 million image dataset takes about 250 ms on one processor core.

Figure 1. Illustration of the similarity of the Fisher vectors of local image regions despite viewpoint changes.

6.1.2. Searching in one billion vectors: re-rank with source coding

Participants: Laurent Amsaleg [CNRS, IRISA], Matthijs Douze, Hervé Jégou [INRIA Rennes], Romain Tavenard [University Rennes I].

In this work [13] we extend our earlier work [4]. An additional level of processing is added to the product quantizer to refine the estimated distances. It consists in quantizing the difference vector between a point and the corresponding centroid. When combined with an inverted file, this gives three levels of quantization. Experiments performed on SIFT and GIST image descriptors show excellent search accuracy outperforming three state-of-the-art approaches.

6.1.3. Combining attributes and Fisher vectors for efficient image retrieval

Participants: Matthijs Douze, Arnau Ramisa, Cordelia Schmid.

Attributes were recently shown to give excellent results for category recognition. In [9] we demonstrate their performance in the context of image retrieval. We show that combining attributes with Fisher vectors improves performance for retrieval of particular objects as well as categories. Furthermore, we implement an efficient coding technique for compressing the combined descriptor to very small codes. Experimental results show that our approach significantly outperforms the state of the art, even for a very compact representation of 16 bytes per image. We show that attribute features combined with Fisher vectors improve the retrieval of image categories and that those features can supplement text features.

6.1.4. Bag-of-colors for improved image search

Participants: Matthijs Douze, Hervé Jégou [INRIA Rennes], Christian Wengert [Kooaba].
In [19] we investigate the use of color information when used within a state-of-the-art large scale image search system. We introduce a simple color signature generation procedure, used either to produce global or local descriptors. As a global descriptor, it outperforms several state-of-the-art color description methods, in particular the bag-of-words method based on color SIFT. As a local descriptor, our signature is used jointly with SIFT descriptors (no color) to provide complementary information.

6.2. Learning and structuring of visual models

6.2.1. Learning to rank and quadratic assignment

Participants: Thomas Mensink, Jakob Verbeek, Tiberio Caetano [NICTA Canberra].

In [16] we show that the optimization of several ranking-based performance measures, such as precision-at-k and average-precision, is intimately related to the solution of quadratic assignment problems, especially when the score function allows for pairwise label dependencies. Both the task of test-time prediction of the best ranking and the task of constraint generation in estimators based on structured support vector machines can all be seen as special cases of quadratic assignment problems. Although such problems are in general NP-hard, we identify a polynomially-solvable subclass (for both inference and learning) that still enables the modeling of a substantial number of pairwise rank interactions. We show preliminary results on a public benchmark image annotation data set, which indicates that this model can deliver higher performance over ranking models without pairwise rank dependencies. This work was performed during a visit to NICTA Canberra by T. Mensink (March – June, ’11) and J. Verbeek (May ’11).

![Figure 2. An automatically obtained dependency tree over 96 labels, that contains 3 labels per node.](image)

6.2.2. Learning structured prediction models for interactive image labeling

Participants: Thomas Mensink, Jakob Verbeek, Gabriela Csurka [Xerox RCE].

In [25] we propose structured models for image labeling that take into account the dependencies among the image labels explicitly. These models are more expressive than independent label predictors, and lead to more accurate predictions. While the improvement is modest for fully-automatic image labeling, the gain is significant in an interactive scenario where a user provides the value of some of the image labels. Such an interactive scenario offers an interesting trade-off between accuracy and manual labeling effort. The structured models are used to decide which labels should be set by the user, and transfer the user input to more accurate
predictions on other image labels. Experimental results on three publicly available benchmark data sets show that in all scenarios our structured models lead to more accurate predictions, and leverage user input much more effectively than state-of-the-art independent models. See Figure 2.

6.2.3. Modeling spatial layout with Fisher vectors for image categorization

Participants: Frédéric Jurie [University of Caen], Josip Krapac, Jakob Verbeek.

In [15] we introduce an extension of bag-of-words image representations to encode spatial layout. Using the Fisher kernel framework we derive a representation that encodes the spatial mean and the variance of image regions associated with visual words. We extend this representation by using a Gaussian mixture model to encode spatial layout, and show that this model is related to a soft-assign version of the spatial pyramid representation. We also combine our representation of spatial layout with the use of Fisher kernels to encode the appearance of local features. Through an extensive experimental evaluation, we show that our representation yields state-of-the-art image categorization results, while being more compact than spatial pyramid representations. In particular, using Fisher kernels to encode both appearance and spatial layout results in an image representation that is computationally efficient, compact, and yields excellent performance while using linear classifiers.

6.2.4. Unsupervised metric learning for face identification in TV video

Participants: Ramazan Cinbis, Jakob Verbeek, Cordelia Schmid.

The goal of face identification is to decide whether two faces depict the same person or not. In [8] we address the identification problem for face-tracks that are automatically collected from uncontrolled TV video data. Face-track identification is an important component in systems that automatically label characters in TV series or movies based on subtitles and/or scripts: it enables effective transfer of the sparse text-based supervision to other faces. We show that, without manually labeling any examples, metric learning can be effectively used to address this problem. This is possible by using pairs of faces within a track as positive examples, while negative training examples can be generated from pairs of face tracks of different people that appear together in a video frame. In this manner we can learn a cast-specific metric, adapted to the people appearing in a particular video, without using any supervision. Identification performance can be further improved using semi-supervised learning where we also include labels for some of the face tracks. We show that our cast-specific metrics not only improve identification, but also recognition and clustering. See Figure 3.

Figure 3. Projections of face signatures projected to two dimensions, using (a) a metric trained on faces detected in still images, (b) using hand labeled faces detected in videos, (c) a metric trained from face tracking results (no manual labeling). Face signatures of different people are color coded. A good face metric can be learned directly from face tracking results, without using any hand labeled examples.
6.2.5. Large-scale image classification

Participants: Miro Dudik [Yahoo! Research], Zaid Harchaoui, Jerome Malick [INRIA Grenoble, BIPOP Team].

We introduced in [10] a new scalable learning algorithm for large-scale multi-class image classification, based on the multinomial logistic loss and the trace-norm regularization penalty. Reframing the challenging non-smooth optimization problem into a surrogate infinite-dimensional optimization problem with regular $\ell_1$-regularization penalty, we propose a simple and provably efficient coordinate descent algorithm. Furthermore, we showed how to perform efficient matrix computations in the compressed domain for quantized dense visual features, scaling up to 100,000s examples, 1,000s-dimensional features, and 100s of categories. Promising experimental results on the “Fungus”, “Ungulate”, and “Vehicles” subsets of ImageNet were obtained, where our approach performed significantly better than state-of-the-art approaches for Fisher vectors with 16 Gaussians.

6.3. Human action recognition

6.3.1. Action recognition by dense trajectories


Feature trajectories have shown to be efficient for representing videos. Typically, they are extracted using the KLT tracker or matching SIFT descriptors between frames. However, the quality as well as quantity of these trajectories is often not sufficient. Inspired by the recent success of dense sampling in image classification, in [18] we propose an approach to describe videos by dense trajectories. An overview of our framework is shown in Figure 4. We sample dense points from each frame and track them based on dense optical flow. Our trajectories are robust to fast irregular motions as well as shot boundaries. Additionally, dense trajectories cover the motion information in videos well. We also investigate how to design descriptors to encode the trajectory information. We introduce a novel descriptor based on motion boundary histograms, which is robust to camera motion. This descriptor consistently outperforms other state-of-the-art descriptors, in particular in uncontrolled realistic videos. We evaluate our video description in the context of action classification with a bag-of-features approach. Experimental results show a significant improvement over the state of the art on four datasets of varying difficulty, e.g., KTH, YouTube, Hollywood2 and UCF sports.

![Figure 4. Illustration of dense trajectories extraction and description. Left: dense sampling of feature points at multiple scales; middle: tracking feature points with a dense optical flow field; right: descriptors are computed along the trajectory.](image-url)
6.3.2. Weakly supervised learning of interactions between humans and objects

Participants: Vittorio Ferrari [ETH Zürich], Alessandro Prest, Cordelia Schmid.

In [7] we introduced a weakly supervised approach for learning human actions modeled as interactions between humans and objects. Our approach is human-centric: we first localize a human in the image and then determine the object relevant for the action and its spatial relation with the human. The model is learned automatically from a set of still images annotated only with the action label. Our approach relies on a human detector to initialize the model learning. For robustness to various degrees of visibility, we build a detector that learns to combine a set of existing part detectors. Starting from humans detected in a set of images depicting the action, our approach determines the action object and its spatial relation to the human. Its final output is a probabilistic model of the human-object interaction, i.e. the spatial relation between the human and the object. We present an extensive experimental evaluation on the sports action dataset from Gupta et al., the PASCAL 2010 action dataset, and a new human-object interaction dataset. In the PASCAL visual object classes challenge 2011 our approach achieved best results on three out of ten action classes and the best result on average over all classes.

6.3.3. Explicit modeling of human-object interactions in realistic videos

Participants: Vittorio Ferrari [ETH Zürich], Alessandro Prest, Cordelia Schmid.

In [26] we introduced an approach for learning human actions as interactions between persons and objects in realistic videos. Previous work typically represents actions with low-level features such as image gradients or optical flow. In contrast, we explicitly localize in space and track over time both the object and the person, and represent an action as the trajectory of the object wrt to the person position. Our approach relies on state-of-the-art approaches for human and object detection as well as tracking. We show that this results in human and object tracks of sufficient quality to model and localize human-object interactions in realistic videos. Our human-object interaction features capture relative trajectory of the object wrt the human. Experimental results on the Coffee & Cigarettes dataset show that (i) our explicit human-object model is an informative cue for action recognition; (ii) it is complementary to traditional low-level descriptors such as 3D-HOG extracted over human tracks. When combining our human-object interaction features with 3D-HOG features, we show to improve over their separate performance as well as over the state of the art. See Figure 5.

6.3.4. Actom sequence models for efficient action detection


In [12] we address the problem of detecting actions, such as drinking or opening a door, in hours of challenging video data. We propose a model based on a sequence of atomic action units, termed “actoms”, that are characteristic for the action. Our model represents the temporal structure of actions as a sequence of histograms of actom-anchored visual features. Our representation, which can be seen as a temporally structured extension of the bag-of-features, is flexible, sparse and discriminative. We refer to our model as Actom Sequence Model (ASM). Training requires the annotation of actoms for action clips. At test time, actoms are detected automatically, based on a non-parametric model of the distribution of actoms, which also acts as a prior on an action’s temporal structure. We present experimental results on two recent benchmarks for temporal action detection. We show that our ASM method outperforms the current state of the art in temporal action detection.

6.3.5. A time series kernel for action recognition


In [11] we address the problem of action recognition by describing actions as time series of frames and introduce a new kernel to compare their dynamic aspects. Action recognition in realistic videos has been successfully addressed using kernel methods like SVMs. Most existing approaches average local features over video volumes and compare the resulting vectors using kernels on bags of features. In contrast, we model actions as time series of per-frame representations and propose a kernel specifically tailored for the purpose of action recognition. Our main contributions are the following: (i) we provide a new principled way to compare
the dynamics and temporal structure of actions by computing the distance between their auto-correlations, (ii) we derive a practical formulation to compute this distance in any feature space deriving from a base kernel between frames, and (iii) we report experimental results on recent action recognition datasets showing that it provides useful complementary information to the average distribution of frames, as used in state-of-the-art models based on bag-of-features.
6. New Results

6.1. Scene and camera reconstruction

Participants: Marie-Odile Berger, Srikrishna Bhat, Nicolas Noury, Gilles Simon, Frédéric Sur.

6.1.1. Image point correspondences and repeated patterns

Matching or tracking interest points between several views is one of the keystones of many computer vision applications, especially when considering structure and motion estimation. The procedure generally consists in several independent steps: interest point extraction, then interest point matching by keeping only the “best correspondences” with respect to the similarity between some local descriptors, and final correspondence pruning to keep those that are consistent with a realistic camera motion (here, consistent with epipolar constraints or homography transformation.) Each step in itself is a delicate task which may endanger the whole process. In particular, repeated patterns give rise to lots of false correspondences in descriptor-based matching. Actual correspondences are thus hardly, if ever, recovered by the final pruning step. Dealing with repeated patterns is of crucial importance in man-made environments. Starting from a statistical model by Moisan and Stival [25], we have proposed a one-stage approach for matching interest points based on simultaneous descriptor similarity and geometric constraint. The resulting algorithm has adaptive matching thresholds and is able to pick up point correspondences beyond the nearest neighbour. We have also shown how to improve A_SIFT [26], an effective point matching algorithm to make it more robust to the presence of repeated patterns [5], [23], [8].

6.1.2. Visual words for pose computation

Visual vocabularies are standard tools in the object/image classification literature, and are emerging as a new tool for building point correspondences for pose estimation. Within S. Bhat’s PhD thesis, we have proposed several methods for visual word construction dedicated to point matching, with structure from motion and pose estimation applications in view. The three dimensional geometry of a scene is first extracted with bundle adjustment techniques based on keypoint correspondences. These correspondences are obtained by grouping the set of all SIFT descriptors from the training images into visual words using transitive closure (TC) techniques. We obtain a more accurate 3D geometry than with classical image-to-image point matching. In a second on-line step, these visual words serve as 3D point descriptors that are robust to viewpoint change, and are used for building 2D-3D correspondences on-line during application, yielding the pose of the camera by solving the PnP problem. Several visual word formation techniques have been compared with respect to robustness to viewpoint change between the learning and the test images. Our experiments showed that the adaptive TC visual words are better in many ways when compared to other classical techniques such as K-means [12].

6.1.3. Tracking by synthesis using point features and pyramidal blurring

Tracking-by-synthesis is a promising method for markerless vision-based camera tracking, particularly suitable for Augmented Reality applications. In particular, it is drift-free, viewpoint invariant and easy-to-combine with physical sensors such as GPS and inertial sensors. While edge features have been used successfully within the tracking-by-synthesis framework, point features have, to our knowledge, still never been used. This is probably due to the fact that real-time corner detectors are weakly repeatable between a camera image and a rendered texture.
We compared the repeatability of commonly used FAST, Harris and SURF interest point detectors across view synthesis [17]. We showed that adding depth blur to the rendered texture can drastically improve the repeatability of FAST and Harris corner detectors (up to 100% in our experiments), which can be very helpful, e.g., to make tracking-by-synthesis running on mobile phones. We proposed a method for simulating depth blur on the rendered images using a pre-calibrated depth response curve. In order to fulfil the performance requirements, a pyramidal approach was used based on the well-known MIP mapping technique. We also proposed an original method for calibrating the depth response curve, which is suitable for any kind of focus lenses and comes for free in terms of programming effort, once the tracking-by-synthesis algorithm has been implemented.

6.1.4. Acquisition of 3D calibrated data

Christel Leonet joined the team in October 2010 as an INRIA assistant engineer with the aim to build an integrated 3D acquisition system. More specifically, the objective of her work is to combine an IMU (Inertial Measurement Unit), a GPS receiver, a laser rangefinder and a video camera for ground truth data acquisitions of camera movements and scene structures. These data will be useful to validate several algorithms developed in our team. This year she dealt with the hand-eye coordination between the different devices. Moreover, a 3D laser pointer has being built, which allows to acquire textured 3D polygons by pointing them with the laser attached to the camera and the IMU put on a tripod.

6.2. Medical imaging


6.2.1. Vessel reconstruction with implicit surfaces

Our research activity is led in collaboration with Shacra project-team from INRIA Lille-Nord Europe and the Department of Interventional Neuroradiology from Nancy University Hospital. It was pursued this year in the context of the SOFA-InterMedS INRIA Large-Scale Initiative.

Our objective is the implicit modeling of blood vessels from 3DRA data, with the aim to use these models for real time simulation of interventional procedures. Within A. Yureidini’s PhD thesis, a new model was developed consisting of a tree of local implicit blobby models. This model was implemented in Sofa simulation platform, enabling interactive simulation time (60 fps) and thereby showing an impressive realism during tool navigation [20]. We focused this year on the extensive validation of our RANSAC-based vessel tracking algorithm, by comparison with state of the art Multiple Hypothesis Testing [24] on 10 patient data [18]. Our initial mechanism to fit the implicit model to patient data relies on the minimization of a multi-termed energy. This energy was put under scrutiny, assessing the contribution of each energy term [19]. Our current goal is to reintroduce the raw image data for a more accurate energy computation, with the aim to design a blobby deformable model.

6.2.2. A variational framework for automatic modeling of the vocal tract

Segmenting the vocal tract in MRI is difficult especially because the tongue may move near other edges in the oral cavity, such as the palate or the teeth, which may disturb the segmentation process. The idea explored in our past work was to guide the segmentation with shape priors learnt on a reference speaker within a shape-based variational framework.

Shape priors were incorporated into segmentation via a PCA model with a relatively large number of components to enable the adaptation of the model to strong morphological differences. During this year, this work was continued with the aim to detect tongue contours in physical correspondences, thus allowing us to build a model of the vocal tract. An automatic method for the identification of the end points as well as an improved variational framework to obtain curves in physical correspondences was described in [15]. Second, we extensively assessed the segmentation process. We experimentally showed that the reference model is able to cope with strong morphological differences between speakers with a limited numbers of modes.
6.2.3. Medical simulators based on task analysis

We present here two works done within a collaboration with Imperial College of London.

In order to validate a virtual reality ultrasound-guided targeted liver biopsy procedure simulators previously designed [22], we have worked on task analysis to deconstruct individual procedural tasks followed by metric definition and critical performance indicator identification. Consultant and trainee scores on the performance metrics were compared. Independent t-tests revealed significant differences between trainees and consultants on 3 performance metrics: targeting, probe usage time and mean needle length in beam. ANOVA reported significant differences across years of experience on seven performance metrics: no-go area touched, targeting, length of session, probe usage time, total needle distance moved, number of skin contacts, total time in no-go area. More experienced participants consistently received better performance scores on all 19 performance metrics [9].

We used the same task analysis technique to design an inguinal hernia repair simulator [16]. The task analysis allowed to break down the complex operation into sub-tasks and it also provided the foundation for useful and productive discussions between clinical staff and developers. We deployed our system as an e-learning application, allowing surgeons to easily access the application.

6.3. National Initiatives

  The SOFA-InterMedS large-scale INRIA initiative is a research-oriented collaboration across several INRIA project-teams, international research groups and clinical partners. Its main objective is to leverage specific competences available in each team to further develop the multidisciplinary field of Medical Simulation research. Our action within the initiative takes place in close collaboration with both Shacra INRIA project-team in Lille and the Department of diagnostic and therapeutic interventional neuroradiology of Nancy University Hospital. We aim at providing in-vivo models of the patient’s organs, and in particular a precise geometric model of the arterial wall. Such a model is used by Shacra team to simulate the coil deployment within an intracranial aneurysm. The associated medical team in Nancy, and in particular our external collaborator René Anxionnat, is in charge of validating our results.

- **ANR ARTIS (2009-2012)**
  Participants: M.O. Berger, A. Eryildirim, E. Kerrien.
  The main objective of this fundamental research project is to develop inversion tools and to design and implement methods that allow for the production of augmented speech from the speech sound signal alone or with video images of the speaker’s face. The Magrit team is especially concerned with the development of procedures allowing for the automatic construction of a speaker’s model from various imaging modalities.

- **ANR Visac (2009-2012)**
  Participants: M.O. Berger, B. Wrobel-Dautcourt.
  The ANR Visac is about acoustic-visual speech synthesis by bimodal concatenation. The major challenge of this project is to perform speech synthesis with its acoustic and visible components simultaneously. Within this project, the role of the Magrit team is twofold. One of them is to build a stereovision system able to record synchronized audio-visual sequences at a high frame rate. Second, a highly realistic dense animation of the head must be produced.

6.4. European Initiatives

6.4.1. Major European Organizations with which you have followed Collaborations

Partner 1: Imperial College, London.
Pierre-Frédéric Villard has a Honorary Research Fellow contract with Imperial College. The research focusing on medical simulators based on task analysis have been done within this link. The collaboration has involved 2 research visits in London to mainly incorporate work done in Lorraine both at the LORIA and with Nancy University intern students. There was also a participation as an activity leader in a one-week summer school on Haptic Technology (to give the basics of computer haptics, including visual and haptics rendering, force feedback, haptic interfaces, collision detection, collision response and deformation modelling).
6. New Results

6.1. Mesh repair with topology control
Participant: Franck Hétroy.

This work is done in collaboration with Carlos Andújar, Pere Brunet and Álvar Vinacua from Universitat Politecnica de Barcelona, Spain, and has been published in the CAD journal [7]. The purpose is to propose an efficient method to create 2-manifold meshes from real data, obtained as soups of polygons with combinatorial, geometrical and topological noise (see Figure 3). We propose to use a voxel structure called a discrete membrane and morphological operators to compute possible topologies, between which the user chooses.

![Figure 3. Two topologically different 2-manifold mesh repairs, from the same polygon soup.](image)

6.2. Topology computation on simplicial shapes
Participants: Dobrina Boltcheva, Franck Hétroy.

This work is a part of the BQR project IDEAL (see Section 8.1.1) which is performed in collaboration with Leila de Floriani from the University of Genova in Italy. The main goal of this project is to study non-manifold geometrical models and to find out features allowing to classify these models and criteria for determining their shape. We are interested in non-manifold models such as idealized industrial CAD models, since they are still ill-understood even if they are frequently used in computer graphics and many engineering applications.

We have developed an efficient method to compute the homology of a large (non-manifold) simplicial complex, from the homologies of its sub-complexes. Computed topological invariants play a crucial role in the field of shape description and analysis. This work has been published in the CAD journal [5] and presented at the SIAM conference on geometric and physical modeling (GD/SPM’11).

6.3. Scale Space Representations on Manifolds
Participant: Edmond Boyer.
In collaboration with Radu Horaud and Andrei Zaharescu, we developed a novel approach for the scale-space representations of scalar functions defined over Riemannian manifolds. One of the main interest in such representations stems from the task of 3D modelling where 2D surfaces, endowed with various physical properties, are recovered from images. Multi-scale analysis allows to structure the information with respect to its intrinsic scale, hence enabling a wide range of low-level computations, similar to what is usually used for representing images. In contrast to the Euclidean image domain, where scale spaces can be easily obtained through convolutions with Gaussian kernels, surfaces require a more general approach that must handle non-Euclidean spaces. Such a generalized scale-space framework is the main contribution of this work, which builds on the spectral decomposition available with the heat-diffusion framework to derive a computational approach for representing scalar functions on 2D Riemannian manifolds using an intrinsic scale parameter. In addition, we proposed a feature detector and a region descriptor, based on these representations, extending the widely used DOG detector and HOG descriptor to manifolds. Experiments on real datasets with various physical properties, i.e., scalar functions, demonstrated the validity and the interest of this approach[16].

6.4. Topologically-Robust 3D Shape Matching

Participant: Edmond Boyer.

3D Shape matching is an important problem in computer vision. One of the major difficulties in finding dense correspondences between 3D shapes is related to the topological discrepancies that often arise due to complex kinematic motions. In this work done in collaboration with Jan Cech, Radu Horaud and Avinash Sharma a shape matching method is proposed that is robust to such changes in topology. The algorithm starts from a sparse set of seed matches and outputs dense matching. We use a shape descriptor based on properties of the heat-kernel and which provides an intrinsic scale-space representation. This descriptor incorporates (i) heat-flow from already matched points and (ii) self diffusion. At small scales the descriptor behaves locally and hence it is robust to global changes in topology. Therefore, it can be used to build a vertex-to-vertex matching score conditioned by an initial correspondence set. This score is then used to iteratively add new correspondences based on a novel seed-growing method that iteratively propagates the seed correspondences to nearby vertices. The matching is further densified via an EM-like method that explores the congruency between the two shape embeddings. The method is compared with two recently proposed algorithms and we show that we can deal with substantial topological differences between the two shapes[15].

6.5. Motion-based segmentation of mesh sequences

Participants: Romain Arcila, Franck Hétroy.
Mesh animations, or sequences of meshes, represent a huge amount of data, especially when acquired from scans or videos. In collaboration with the university of Lyon (LIRIS lab), we address the problem of partitioning these sequences, in order to both recover motion information and be able to compress them. Following last year’s method, we proposed this year a second and third motion-based segmentation algorithm, which clusters mesh vertices into static or rigidly moving components (see Figure 5). These methods are based on spectral clustering of the vertex transformations and are more robust and general than the previous one. This work has been submitted for publication to a journal, and is part of the PhD thesis of Romain Arcila [1].

Figure 5. Temporally varying segmentation of a mesh sequence, into rigidly moving components. Right hand and arm are merged at some time since they start to move accordingly. They are later split, when they start to follow different motions.

6.6. Surface Flow

Participants: Antoine Letouzey, Benjamin Petit, Jean-Sébastien Franco, Edmond Boyer.

Recovering dense motion information is a fundamental intermediate step in the image processing chain upon which higher level applications can be built, such as tracking or segmentation. For that purpose, pixel observations in the image provide useful motion cues through temporal variations of the intensity function. We have studied the estimation of dense, instantaneous 3D motion fields over non-rigidly moving surface observed by multi-camera systems. The motivation arises from multi-camera applications that require motion information for arbitrary subjects, in order to perform tasks such as surface tracking or segmentation. To this aim, we have proposed a novel framework that allows to efficiently compute dense 3D displacement fields using low level visual cues and geometric constraints. The main contribution is a unified framework that combines flow constraints for small displacements with temporal feature constraints for large displacements and fuses them over the surface using local rigidity constraints. The resulting linear optimization problem allows for variational solutions and fast implementations. Experiments conducted on synthetic and real data demonstrated the respective interests of flow and feature constraints as well as their efficiency to provide robust surface motion cues when combined[14], [18].

As an extension of this work, we also studied the situation where a depth camera and one or more color cameras are available, a common situation with recent composite sensors such as the Kinect. In this case, geometric information from depth maps can be combined with intensity variations in color images in order to estimate smooth and dense 3D motion fields. We propose a unified framework for this purpose, that can handle both arbitrary large motions and sub-pixel displacements. The novelty with respect to existing scene flow approaches is that it takes advantage of the geometric information provided by the depth camera to define a surface domain over which photometric constraints can be consistently integrated in 3D. Experiments on real and synthetic data provide both qualitative and quantitative results that demonstrated the interest of the approach[13].
6.7. Learning Temporally Consistent Rigidities

Participants: Jean-Sébastien Franco, Edmond Boyer.

We present a novel probabilistic framework for rigid tracking and segmentation of shapes observed from multiple cameras. Most existing methods have focused on solving each of these problems individually, segmenting the shape assuming surface registration is solved, or conversely performing surface registration assuming shape segmentation or kinematic structure is known. We assume no prior kinematic or registration knowledge except for an over-estimate k of the number of rigidities in the scene, instead proposing to simultaneously discover, adapt, and track its rigid structure on the fly. We simultaneously segment and infer poses of rigid subcomponents of a single chosen reference mesh acquired in the sequence. We show that this problem can be rigorously cast as a likelihood maximization over rigid component parameters. We solve this problem using an Expectation Maximization algorithm, with latent observation assignments to reference vertices and rigid parts. Our experiments on synthetic and real data show the validity of the method, robustness to noise, and its promising applicability to complex sequences. This work was presented at the CVPR 2011 conference [11].
6.8. Ontology-based mesh segmentation

Participants: Sahar Hassan, Franck Hétroy.

Patient-specific 3D virtual models of anatomical organs are becoming more and more useful in medicine, for instance for diagnosis or follow-up care purposes. These models are usually created from 2D scan or MRI images. However, small or thin geometrical features, such as ligaments, are sometimes not visible on these images. We propose to use an anatomical ontology, called MyCorporisFabrica [http://www.mycorporisfabrica.org/], to add missing parts to reconstructed virtual organs. This ontology describes definitions of and relationships between organs: e.g., femur is part of the leg. The first step towards the full achievement of this process is to segment virtual models, often represented by 2D meshes, into meaningful parts. In our case, “meaningful” means “related to the ontology”: each part should refer to an organ defined in the ontology. An algorithm to decompose a given organ into sub-organs according to the ontology has been proposed in the PhD thesis of Sahar Hassan [2]: first, we approximate organ shapes by geometric primitives, then we segment a given organ mesh by optimizing objective functions which are related to these primitives.

6.9. Detection and quantification of brain aneurysms

Participants: Sahar Hassan, Franck Hétroy.

Aneurysms are excrescences on blood vessels. They can break, letting the blood propagate outside the vessel, which often leads to death. In some cases, the blood clots sufficiently fast so that people survive. However, a neurosurgeon or a neuroradiologist should intervene very quickly in order to reparer the vessel before the aneurysm breaks once more.

The purpose of our research is to help neurosurgeons and neuroradiologists to plan surgery, by giving them quantitative information about the size, shape and geometry position of aneurysms. This work was part of the PhD of Sahar Hassan [2], and has presented at the International Conference on Computer Analysis of Images and Patterns (CAIP) [12]. The method we propose first extracts a centered skeleton from the input voxel set of the vascular tree, then detects aneurysms by studying variations of vessel diameters along the skeleton. The name of an aneurysm-carrying vessel is also given thanks to a partial graph matching technique, and accurate measures to decide the treatment are provided.

6.10. Dimensionality reduction for character animation

Participants: Maxime Tournier, Lionel Reveret.

This work investigates and proposes a mathematical framework to perform statistical analysis and dimensionality reduction on rotational trajectories derived from motion capture data. Motion capture data consists in a set of trajectories in the space of 3D rotations (SO(3)) and as such do not present properties of an Euclidian space. Consequently there is no easy to way to apply standard dimensionality reduction techniques on these data. Using the formalism of exponential maps and Principle Geodesics Analysis (PGA), it has been shown that it is possible to rigorously derive a dimensionality reduction analysis on such data. This reduction can be typically applied for compression of motion capture data and probabilistic implementation of the Inverse Kinematics problem. This approach has shown good properties in the context of physically-based animation with a Lagrangian formulation of rigid body dynamics coupled with geometric integrators. These integrators allow a good preservation of momentum using only first order equations, achieving both real-time and high level of realism. These works were developed through the PhD thesis of Maxime Tournier [4]. Early development of PGA on motion capture data had been published at Eurographics in 2009. Its integration into a GPLVM framework has been published this year in the IEEE CG&A journal [9]. Its extension into the context of physically-based animation is currently under preparation for publication.

6.11. Animation of quadrupeds locomotion

Participant: Lionel Reveret.
Following a study on locomotion of quadrupeds by a team in the National Museum of Natural History (MNHN), a new theory on motion planning has been proposed. This theory, the Antero-Posterior Sequences (APS), allows a characterization of the sequence of foot placement for quadrupeds for all regular gaits with very few parameters, as well as transition between gaits, starting from stop to full gallop. In collaboration with the MNHN and the robotics department of the University of Versailles-Saint Quentin en Yvelines (UVSQ), a rigorous software implementation has been specified and developed. This software allows automatically generating foot planning of quadrupeds locomotion according to a desired speed transition. Co-workers for this project were Ludovic Maes and Anick Abourachid at the MNHN and Vincent Hugel at the UVSQ. A patent has been written and finalized for this project.

In parallel, collaboration on physical simulation of quadrupeds locomotion has been carried on with Stelian Coros (previously at University of British Columbia (UBC), now at Disney Research) and Michiel van de Panne (UBC). Automatic video analysis of dog walking, trotting and running has been used to optimize parameters of physical controllers. This work has been published at SIGGRAPH 2011 [6].

Figure 8. Inverse Kinematics and motion synthesis using PGA of motion capture data.

Figure 9. Physical animation of dog locomotion
6. New Results

6.1. Calibration of a mixed camera system

An approximately Euclidean representation of the visible scene can be obtained directly from a range, or
time-of-flight, camera. An uncalibrated binocular system, in contrast, gives only a projective reconstruction of
the scene. This paper analyzes the geometric mapping between the two representations, without requiring an
intermediate calibration of the binocular system. The mapping can be found by either of two new methods, one
of which requires point correspondences between the range and colour cameras, and one of which does not. It
is shown that these methods can be used to reproject the range data into the binocular images, which makes it
possible to associate high resolution colour and texture with each point in the Euclidean representation.

6.2. Computation of scene flow

A simple seed growing algorithm for estimating scene flow in a stereo setup is presented. Two calibrated and
synchronized cameras observe a scene and output a sequence of image pairs. The algorithm simultaneously
computes a disparity map between the image pairs and optical flow maps between consecutive images. This,
together with calibration data, is an equivalent representation of the 3D scene flow, i.e. a 3D velocity vector
is associated with each reconstructed point. The proposed method starts from correspondence seeds and
propagates these correspondences to their neighborhood. It is accurate for complex scenes with large motions
and produces temporally-coherent stereo disparity and optical flow results. The algorithm is fast due to inherent
search space reduction. An explicit comparison with recent methods of spatiotemporal stereo and variational
optical and scene flow is provided.

6.3. 3D shape analysis and registration

We address the problem of 3D shape registration and we propose a novel technique based on spectral graph
theory and probabilistic matching. Recent advancement in shape acquisition technology has led to the capture
of large amounts of 3D data. Existing real-time multi-camera 3D acquisition methods provide a frame-
wise reliable visual-hull or mesh representations for real 3D animation sequences. The task of 3D shape
analysis involves tracking, recognition, registration, etc. Analyzing 3D data in a single framework is still a
challenging task considering the large variability of the data gathered with different acquisition devices. 3D
shape registration is one such challenging shape analysis task. The main contribution of this chapter is to
extend the spectral graph matching methods to very large graphs by combining spectral graph matching with
Laplacian embedding. Since the embedded representation of a graph is obtained by dimensionality reduction
we claim that the existing spectral-based methods are not easily applicable. We discuss solutions for the
exact and inexact graph isomorphism problems and recall the main spectral properties of the combinatorial
graph Laplacian; We provide a novel analysis of the commute-time embedding that allows us to interpret the
latter in terms of the PCA of a graph, and to select the appropriate dimension of the associated embedded
metric space; We derive a unit hyper-sphere normalization for the commute-time embedding that allows us to
register two shapes with different samplings; We propose a novel method to find the eigenvalue-eigenvector
ordering and the eigenvector sign using the eigensignature (histogram) which is invariant to the isometric
shape deformations and fits well in the spectral graph matching framework, and we present a probabilistic
shape matching formulation using an expectation maximization point registration algorithm which alternates
between aligning the eigenbases and finding a vertex-to-vertex assignment.
6.4. A differential model for the complex cell

The receptive fields of simple cells in the visual cortex can be understood as linear filters. These filters can be modelled by Gabor functions, or by Gaussian derivatives. Gabor functions can also be combined in an energy model of the complex cell response. This work proposes an alternative model of the complex cell, based on Gaussian derivatives. It is most important to account for the insensitivity of the complex response to small shifts of the image. The new model uses a linear combination of the first few derivative filters, at a single position, to approximate the first derivative filter, at a series of adjacent positions. The maximum response, over all positions, gives a signal that is insensitive to small shifts of the image. This model, unlike previous approaches, is based on the scale-space theory of visual processing. In particular, the complex cell is built from filters that respond to the 2-D differential structure of the image. The computational aspects of the new model are studied in one and two dimensions, using the steerability of the Gaussian derivatives. The response of the model to basic images, such as edges and gratings, is derived formally. The response to natural images is also evaluated, using statistical measures of shift insensitivity. The relevance of the new model to the cortical image-representation is discussed.

6.5. Audiovisual fusion based on a mixture model

The problem of multimodal clustering arises whenever the data are gathered with several physically different sensors. Observations from different modalities are not necessarily aligned in the sense there there is no obvious way to associate or to compare them in some common space. A solution may consist in considering multiple clustering tasks independently for each modality. The main difficulty with such an approach is to guarantee that the unimodal clusterings are mutually consistent. In this paper we show that multimodal clustering can be addressed within a novel framework, namely conjugate mixture models. These models exploit the explicit transformations that are often available between an unobserved parameter space (objects) and each one of the observation spaces (sensors). We formulate the problem as a likelihood maximization task and we derive the associated conjugate expectation-maximization algorithm. The convergence properties of the proposed algorithm are thoroughly investigated. Several local/global optimization techniques are proposed in order to increase its convergence speed. Two initialization strategies are proposed and compared. A consistent model-selection criterion is proposed. The algorithm and its variants are tested and evaluated within the task of 3D localization of several speakers using both auditory and visual data.
PRIMA Project-Team

6. New Results

6.1. A Lightweight Augmented Virtuality System for Providing a Faithful and Spatially Manipulable Visual Hand Representation

**Participants:** Sabine Coquillart, Olivier Martin, Andreas Pusch.

We introduced the technical foundations of a system designed to embed a lightweight, faithful and spatially manipulable representation of the user’s hand into an otherwise virtual world - Augmented Virtuality (AV). A highly intuitive control during pointing-like near space interaction can be provided to the user, as well as a very flexible means to experimenters, in a variety of contexts. Our approach essentially relies on stereoscopic video see-through Augmented Reality (AR) technology and a generic, extendible framework for managing 3-D visual hand displacements. Research from human-computer interaction, perception and motor control has contributed to the elaboration of our proposal which combines a) acting in co-location, b) avoiding occlusion violations by assuring a correct scene depth ordering and c) providing a convincing visual feedback of the user’s hand. This system has already successfully been used in one case and further promising applications are studied [17], [18].

6.2. Effects of Hand Feedback Fidelity on Near Space Pointing Performance and User Acceptance

**Participants:** Sabine Coquillart, Olivier Martin, Andreas Pusch.

We conducted an experiment to test the effects of different hand representations on near space pointing performance and user preference. Subjects were presented with varying levels of hand realism, including real hand video, a high and a low level 3D hand model and an ordinary 3D pointer arrow. Behavioural data revealed that an abstract hand substitute like a 3D pointer arrow leads to significantly larger position estimation errors in terms of lateral target overshooting when touching virtual surfaces with only visual hand movement constraints. Further, questionnaire results show that a higher fidelity hand is preferred over lower fidelity representations for different aspects of the task [18].
6. New Results

6.1. Introduction

This year Pulsar has tackled several issues related to its two main research axes: scene understanding for activity recognition and software engineering for activity recognition.

6.1.1. Scene Understanding for Activity Recognition


This year Pulsar has proposed new algorithms in computer vision (people head and face detection and people re-identification), in reasoning (activity recognition and uncertainty handling). More precisely, the new results for this research axis concern:

- People detection in monocular video sequences (6.2)
- Online Parameter Tuning for Object Tracking Algorithms (6.3)
- Fiber Based Video Segmentation (6.4)
- Multiple Birth and Cut Algorithm for Multiple Object Detection (6.5)
- Exhaustive Family of Energies Minimizable Exactly by a Graph Cut (6.6)
- Steepest Descent in Banach Spaces with Application to Piecewise-Rigid Evolution of Curves (6.7)
- Object Tracking Using a Particle Filter based on SIFT Features (6.8)
- Human Re-identification using Riemannian Manifolds (6.9)
- Global Tracking of Multiples Actors (6.10)
- Crowd Data Collection from Video Recordings (6.11)
- Events Recognition and Performance Evaluation (6.12)
- Group interaction and group tracking for video-surveillance in underground railway stations (6.13)
- Action Recognition in Videos (6.14)
- Activity Recognition Applied on Health Care Application (6.15)
- A Cognitive Vision System for Nuclear Fusion Device Monitoring (6.16)
- Scenario Recognition with depth camera (6.17)
- Trajectory Clustering for Activity Learning (6.18)

6.1.2. Software Engineering for Activity Recognition

This year Pulsar has improved the SUP platform. This latter is the backbone of the team experiments to implement the new algorithms proposed by the team in perception, understanding and learning. We improve our meta-modeling approach to support the development of video surveillance applications based on SUP. We continue the development of a scenario recognition module relying on formal methods to support activity recognition in SUP platform. We also continue to study the definition of multiple services for device adaptive platform for scenario recognition. Finally, we are implementing the new theoretical results obtained last year to improve the Clem toolkit.

The new results related to this research axis concern:
- SUP Software Platform (6.19)
- Model-Driven Engineering and Video-surveillance (6.20)
- Scenario Analysis Module (6.22)
- Multiple Services for Device Adaptive Platform for Scenario Recognition (6.23)
- The Clem Toolkit (6.24)

6.2. People detection in monocular video sequences

Participants: Etienne Corvee, François Brémond, Silviu-Tudor Serban, Vasanth Bathrinaryanan.

A video understanding system analyzes human activity by detecting people in video sequences and tracking their displacement and movement throughout the sequences. The better the detection quality, the higher the semantic level of the information is. People activity can differ greatly from one application to another e.g. the presence of a person in one zone can simply be detected from a moving pixel region in a manually specified zone whereas detecting people fighting in a subway requires more complex information. For people activity to be recognized, one needs to detect people accurately in videos and at real time frame rate. Current state of the art algorithms provide generic people detection solutions but with limited accuracy. In the people monitoring domain, although cameras remain mostly fixed, many issues occur in images. For example, outdoor scenes display strong varying lighting conditions (e.g. sunny/cloudy illumination, important shadows), public spaces can be often crowded (e.g. subways, malls) and images can be obtained with a low resolution and can be highly compressed. Hence, detecting and tracking objects in such complex environment remains a delicate task to perform. In addition, detecting people has to face one major difficulty which is caused by occlusion where important information is hidden. When people overlap onto the image plane, their foreground pixels cannot be separated using a standard thresholding operation from a background reference frame. Therefore vision algorithms need to use information held by the underlying pixels and located at specific locations such as body parts.

We have extended our work by implementing and testing a novel people, head and face detection algorithm using Local Binary Pattern based features and Haar like features. The traditional and efficient Adaboost training scheme is adopted to train object features from publicly available databases. This work has been published in the ICVS Workshop [36].

The work has been tested for group tracking in Vanahem videos (see section 8.2.1.2) and for people tracking in Videold videos. The Videold project aims to re-identify people across a network of non overlapping cameras using iris, face and human appearance recognition. An example of tracked people, head and faces in a testing database is shown in figure 7. An example of re-identified face is shown in figure 8 by the Videold interface in a Paris underground video.

We have evaluated our people detection algorithm on the test human dataset provided by INRIA against state of the art algorithms which we refer as HOG [59] and LBP-HOG [77]. The INRIA human dataset is composed of 1132 human images and 453 images of background scenes containing no human. The results are displayed in figure 9 which shows that we obtain slightly better performances than the HOG-LBP technique in terms of missed detection rate vs. FPPI i.e. False Positive Per Image. In this figure, two extreme functioning modes could be chosen: approximately 2 noisy detections are obtained every 1000 background images for 50% true positive detections or 1 noisy detection every 2 frames for a detection rate of approximately 88%.
Figure 7. Example of tracked people, head and face

Figure 8. Face recognition in Paris underground
Figure 9. People detection evaluation: False Positive Per Image vs. miss detection rate for the INRIA test database

The same evaluation scheme of people detection above is used for face detection evaluation. The FPPI rates are obtained on 997 NICTA [66] background images of 720x576 pixels. 180 faces provided by a CMU test face image database are used to evaluate true positive rates. We have compared our results with the 2 versions of Haar feature provided by the OpenCv library i.e. the standard ‘default’ and alternative ‘alt’ training parameters. The results in table 1 show that the Haar ‘alt’ technique performs better than the traditional Haar one. And our haar based technique called CCR provides similar face detection rates while giving a less false alarm rate. The proposed approach is approximately 1% less successful in detecting faces than the Haar technique while this latter is 32% more noisier than our CCR technique.

<table>
<thead>
<tr>
<th>technique</th>
<th>TP(%)</th>
<th>FPPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haar (default)</td>
<td>91.57</td>
<td>4.132</td>
</tr>
<tr>
<td>Haar (alt)</td>
<td>92.13</td>
<td>1.685</td>
</tr>
<tr>
<td>CCR</td>
<td>91.01</td>
<td>1.274</td>
</tr>
</tbody>
</table>

Table 1. Face detection evaluation

6.3. Online Parameter Tuning for Object Tracking Algorithms

Participants: Duc Phu Chau, Monique Thonnat, François Brémond.

Many approaches have been proposed to track mobile objects in a scene. However the quality of tracking algorithms always depends on scene properties such as: mobile object density, contrast intensity, scene depth and object size. The selection of a tracking algorithm for an unknown scene becomes a hard task. Even when the tracker has appropriately selected, it is difficult to tune online its parameters to get the best performance.

Therefore we propose a new control approach for mobile object tracking. More precisely in order to cope with the tracking context variations, this approach learns how to tune the parameters of object appearance-based tracking algorithms. The tracking context of a video sequence is defined as a set of features: density of mobile objects, their occlusion level, their contrasts with regard to the background and their 2D areas. In an offline supervised learning phase, satisfactory tracking parameters are searched for each training video sequence. Then these video sequences are classified by clustering their contextual features. Each context cluster is associated with the learned tracking parameters. In the online control phase, two approaches are
proposed. In the first one, once a context change is detected, the tracking parameters are tuned for the new context using the learned values. In the second approach, the parameter tuning is performed when the context changes and the tracking quality (computed by an online performance evaluation algorithm [56]) is not good enough. An online learning process enables to update the context/parameter relations.

We have also proposed two new tracking algorithms to experiment the proposed control method. The first tracker relies on a Kalman filter and a global tracking which aims at fusing trajectories belonging to the same mobile object. This work has been published in [35]. The second tracker relies on the similarities of eight object descriptors (2D and 3D positions, area, shape ratio, HOG, color histogram, color covariance and dominant color) to build object trajectories. This work has been published in [34].

(a) CARETEKER   
(b) CAVIAR

Figure 10. (a) CARETEKER: Illustration of the Caretaker video; (b) CAVIAR: Illustration of the Caviar video

The proposed controller has been experimented on a long, complex video belonging to the Caretaker European project 1 (see figure 10 (a)) and 26 videos of Caviar dataset 2 (see figure 10 (b)). For the Caretaker video, when the controller is used, the tracking quality increases from 52% to 78%. For the Caviar dataset, the experimental results show that the tracking performance increases from 78.3% to 84.4% when using the controller. The tracking results on Caviar videos with the proposed controller are as good as the ones obtained with manual parameter tuning.

6.4. Fiber Based Video Segmentation

Participants: Ratnesh Kumar, Guillaume Charpiat, Monique Thonnat.

The aim of this work is to segment objects in videos by considering videos as 3D volumetric data \((\text{space} \times \text{time})\). Figure 11 shows 2D slices of a video volume. Bottom right corner of each figure shows the current temporal depth in the volume, while top right shows the X-time slice and bottom left shows Y-time slice. In this 3D representation of videos, points of static background form straight lines of homogeneous intensity over time, while points of moving objects form curved lines. Analogous to the fibers in MRI images of human brains, we name fibers, these straight and curved lines of homogeneous intensity. So, in our case, to segment the whole video volume data, we are interested in a dense estimation of fibers involving all pixels.

For the detection of these fibers, we use motion flow vectors and intensity correlation of 2D patches over time. As these techniques are not reliable everywhere in the image domain, we sort the fibers based on the reliability of the detections from these techniques. The subsequent goal is then to pick high ranked fibers to propagate motion information and boundary fronts to other regions of the 3D volume.

1 http://cordis.europa.eu/ist/kct/caretaker_synopsis.htm
2 http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/
To reliably propagate information from a fiber, we express the reliability of detection of a fiber and the cost of propagation of information from it. The later can be based on a distance measure of a pixel from a fiber, while reliability of a fiber involves motion coherency, color homogeneity, duration along time axis etc.

Our work closely relates to [72]. A video is represented by a set of particles (trajectory of an image point sample). The algorithm then extends and truncates particle trajectories to model motion near occlusion boundaries.

Figure 12 shows some straight fibers found in a video volume. The reliability of these fibers is based on temporal length. Fibers which have temporal span same as that of the video are colored in green, while fibers which have temporal span of less than 10% of the video are colored blue. Red colored fibers have temporal length in between green and blue colored fibers.

Figure 12. Video Volume and Colored Straight Fibers: Green - Most reliable, Blue - Least reliable

6.5. Multiple Birth and Cut Algorithm for Multiple Object Detection

Participant: Guillaume Charpiat.

In collaboration with the Ariana team (Ahmed Gamal-Eldin, Xavier Descombes and Josiane Zerubia), we developed a new optimization method which we call Multiple Birth and Cut (MBC). It combines the recently proposed Multiple Birth and Death (MBD) algorithm and the Graph-Cut algorithm. MBD and MBC optimization methods are applied to energy minimization of an object based model, the marked point process. The most important advantage of the MBC over MBD is the reduction of number of parameters. By proposing good candidates throughout the selection phase in the birth step, the speed of convergence is increased. In
this selection phase, the best candidates are chosen from object sets by a belief propagation algorithm. The algorithm is applied on the flamingo counting problem in a colony [37], [26].

6.6. Exhaustive Family of Energies Minimizable Exactly by a Graph Cut

Participant: Guillaume Charpiat.

Graph cuts are widely used in many fields of computer vision in order to minimize in small polynomial time complexity certain classes of energies. These specific classes depend on the way chosen to build the graphs representing the problems to solve. We study here all possible ways of building graphs and the associated energies minimized, leading to the exhaustive family of energies minimizable exactly by a graph cut. To do this, we consider the issue of coding pixel labels as states of the graph, i.e. the choice of state interpretations. The family obtained comprises many new classes, in particular energies that do not satisfy the submodularity condition, including energies that are even not permuted-submodular.

We studied in details a generating subfamily, in particular we proposed a canonical form to represent Markov random fields, which proves useful to recognize energies in this subfamily in linear complexity almost surely, and then to build the associated graph in quasilinear time. We performed a few experiments to illustrate the new possibilities offered [33]. We have also started to use this technique to minimize exactly approximations of Markov random field energies instead of minimizing approximately the exact energies, by projecting energies on the family we know to solve globally efficiently.

6.7. Steepest Descent in Banach Spaces with Application to Piecewise-Rigid Evolution of Curves

Participant: Guillaume Charpiat.

This is joint work with Gabriel Peyré (CNRS, Ceremade, Université Paris-Dauphine). We intend to favor piecewise-rigid motions, i.e. articulated movements, during shape evolutions, especially when computing morphings or image segmentation with shape prior. To do this, we first need a dissimilarity measure between shapes, whose gradient is meaningful. We formulate one using kernels and bistochastization.

The parameters of these kernels are automatically estimated in a fixed-point scheme that guarantees physical relevance, and the notion of bistochastization is extended to continuous distributions. Finally, piecewise rigidity is ensured during gradient descents by a change of the norm from which the gradient is derived. This norm is formulated so as to favor sparse second derivatives, which produces articulated movements without knowing by advance the location of the articulations.

The formula of the norm is actually elegantly simple, involving simple geometric quantities, derivatives, and the $L_1$ norm. Note that this norm does not derive from an inner product but defines a gradient in the sense of [5] as the minimizer of an energy. It turns out that in our case the energy defining the gradient is actually convex, and efficient minimization follows.

6.8. Object Tracking Using a Particle Filter based on SIFT Features

Participants: Malik Souded, François Brémond.

The approach consists in detecting SIFT points of interest on the objects to track, calculating their SIFT descriptors, tracking these points with a particle filter, and finally achieving tracking process by linking them along the time with links which are weighted by measures on SIFT descriptors and reliability.

The main contributions in this work are on three points.

The first point consists in techniques of detection and selection of SIFT points, allowing a better distribution of points of interest on the target and allowing better management of partial occlusions, and secondly an optimized computing time thanks to the parallelization of these SIFT computation on modern processors (see figure 13).
The second point concerns the weighting of the particles during tracking. This is done with a combination of two kinds of information: the similarity measure of the SIFT descriptor and the state of motion of pixels corresponding to the particles. This allows more robust tracking of SIFT points regardless of the quality of the background subtraction providing the detected objects.

The last point concerns the selection of temporal links between tracked objects and detected ones. These links are selected according to their weight. The weight of each link is based on the proportion of common SIFT points to both objects (two successive images) potentially linked, and the reliability of each of these SIFT point. This reliability is calculated for each point by measuring the variation of the SIFT descriptor during the tracking time.

The occlusion management is performed using three types of information: SIFT descriptors used for tracking (matching after reappearance) the dominant colors of the object of interest and finally the width, height, and speed (in real world) of the object, which are learned in Gaussian models during the tracking (tracking being used with video cameras which have been calibrated), see figure 14.

The approach was tested on 121 sequences of four different datasets: 80 sequences from CAVIAR, 34 sequences from ETISEO, 3 sequences from PETS2001 and 2 sequences from VS_PETS2003. The obtained results are satisfying. The comparison of these results with the state of the art shows improvements for the benchmarking dataset (ETISEO). The following table compares the proposed approach and state of the art results on ETISEO data base:

<table>
<thead>
<tr>
<th>Metrics</th>
<th>ETI-VS1-BE-18-C4</th>
<th>ETI-VS1-BE-16-C4</th>
<th>ETI-VS1-MO-7-C1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed tracker</td>
<td>M1</td>
<td>0.68</td>
<td>0.54</td>
</tr>
<tr>
<td>Chau et al. (VISAPP 2011)</td>
<td>M1</td>
<td>0.64</td>
<td>0.36</td>
</tr>
<tr>
<td>Best team in ETISEO Project (2005)</td>
<td>M1</td>
<td>0.48</td>
<td>0.44</td>
</tr>
</tbody>
</table>

This work was published in [49].

### 6.9. Human Re-identification using Riemannian Manifolds

**Participants:** Sławomir Bąk, Etienne Corvée, François Brémond, Monique Thonnat.

This work addresses the human re-identification problem, which is defined as a requirement to determine whether a given individual has already appeared over a network of cameras. This problem is particularly hard by significant appearance changes across different camera views. In order to re-identify people, a human signature should handle difference in illumination, pose and camera parameters.
We propose new appearance models based on the *mean riemannian covariance* (MRC) matrices combining the appearance information from multiple images. These mean covariance matrices not only keep information on feature distribution but also carry out essential cues about temporal changes of an appearance. Using MRC-s, we propose two methods for an appearance representation:

- **Learned Covariance Patches (LCP)** [25] - a distinctive representation is extracted by a boosting scheme. The structure of MRC patches (size, position) is learned using boosting algorithm based on confidence-rated predictions (see Fig. 15 (a)). These confidence-rated coefficients are employed to weight appearance characteristics of a specific individual w.r.t. the reference (training) dataset of humans.

- **Mean Riemannian Covariance Grid (MRCG)** [30] - less computationally demanding technique than LCP. We represent a human appearance by a grid of MRC cells (see Fig. 15 (b)). Relevant cells are identified by an efficient discriminant analysis. This analysis takes into account variance of MRC patch in the class of humans (reference dataset). MRC is assumed to be more significant when its variance is larger in the class of humans: (1) the most common patterns belong to the background (the variance is small); (2) the patterns which are far from the rest are at the same time the most discriminative (the variance is large). All operations, such as mean or variance, are performed on covariance manifold specified as Riemannian.

Both methods are evaluated and compared with the state of the art using publicly available datasets. We demonstrate that the proposed approaches outperform state of the art methods. Further, we extract new sets of individuals from i-LIDS data to investigate more carefully advantages of using many images for human re-identification.

The computation complexity is analyzed in the context of distance operator between two signatures. Comparing two human signatures, it is necessary to compute distance between covariance matrices, which requires solving the generalized eigenvalues problem. This operation is computationally heavy. In [31], we propose an implementation for finding generalized eigenvalues and eigenvectors for distance operator, using NVIDIA GPU architecture. We improve significantly the performance, reaching 66 speedup using Tesla S1070.
6.10. Global Tracking of Multiples Actors

Participants: Julien Badie, François Brémond.

We propose a new approach for long term tracking of individuals. Our main objective is to design a tracking algorithm for people reidentification [30] that can track people even if they come back in the scene after leaving it. This algorithm is based on covariance matrix and we have also added some contextual information of the scene (for instance, zones where people can enter the scene) to improve tracking performance. In addition, a basic noise detection system and a tracking correction system are proposed in order to handle short-term tracking errors such as multiplication of IDs corresponding to only one individual. The noise detection system is designed to find and remove objects that are detected in a very small number of consecutive frames (for instance 4) and disappear afterward. The tracking correction system associates IDs recently lost with IDs that have just started to be tracked based on geometrical features and 3D distance criteria.

As a result, the tracking quality is significantly improved on 5 video sequences tested from the ETISEO dataset. The people reidentification algorithm gives encouraging results for future work. The number of IDs associated to one individual is reduced (on average 50% less) and the tracking quality improves due to the IDs stability. This algorithm can detect not only people re-entering the scene but also trajectory interruptions due to occlusions or misdetections. This approach could enable the detection of new kinds of events on video sequences such as long range people tracking on a camera network.

6.11. Crowd Data Collection from Video Recordings

Participants: Jihed Joobeur, François Brémond.

The aim of this work is to analyze crowd behaviors by detecting specific situations: panic, congestion, fighting etc. We validate our work with subway station videos from VANAHEIM project. We use Mixture of Gaussian based segmentation to extract moving point and then detecting moving objects. Subsequently inside these moving objects we detect FAST feature points and compute HOG descriptors for tracking these points. We compute different features based on these points like speed and orientation. To estimate the crowd density we use features based on Grey-Level Co-occurrence Matrix. As these features depend on the distance of people from the camera, we divide the scene into different zones which have each zone same distance from the

---

Figure 15. Computation of the human signature using Riemannian manifold (depicted with the surface of the sphere): (a) LCP - the structure of patches is learned by a boosting algorithm; (b) MRCG - covariances gathered from tracking results are used to compute the MRC-s.
Figure 16. The green lines are the ground-truth and the blue lines are the tracking results using both reidentification and noise removal (left) and with only noise removal (right). The man is tracked throughout the whole video with a single ID (68).
camera. In each area, compiling all the information on speed, direction and learned over a threshold density of the crowd, we can learn and detect different situations. For example, if the density increases and the average speed decreases in a pre-defined zone, that may correspond to a congestion situation.

On figure 17 the FAST feature points are shown in blue points, while the tracking of these points is shown in yellow.

Figure 17. Feature points detection and tracking in different zones.


Participants: Ricardo Cezar Bonfim Rodrigues, François Brémond.

The goal of this work is to evaluate the accuracy and performance of events detection, see workflow in Figure 18. The experiments will be performed using the tools developed in Pulsar team, such as Scene Understand Platform (SUP) 4 a plugin for events detection [50] and ViseVal 5.

The experiments were performed using video sequences of a subway station (VANAHEIN dataset) where the goal was to detect events such as people waiting, entering, buying tickets and so on. Preliminary results showed a very low accuracy and demonstrated that the scenario configuration parameters are very sensitive in this problem. It means many of the expected events were missed or misclassified, specially composite events (when more than one activity recognition is required) see some issues on Figure 19.

Based on the issues, a second experiment was configured using 3 different video sequences. In this new experiment, the scenario was adjusted to give more tolerances to people detection issues, the camera calibration was refined and some events were remodeled. After these changes the results were significantly improved. This last experiments showed that the engine proposed by Pulsar team is able to detect events accurately however events modeling can be very sensitive to the scenario configuration, see the results in Table 3.

Table 3. Results of events detection based on the people detection performed in experiment 2 using an IDIAP algorithm. The global Precision and sensibility of people detection are respectively 0.91 and 0.95.

<table>
<thead>
<tr>
<th></th>
<th>Sequence 1</th>
<th>Sequence 2</th>
<th>Sequence 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision (global)</td>
<td>0.73</td>
<td>1.00</td>
<td>0.88</td>
</tr>
<tr>
<td>Sensitivity (global)</td>
<td>0.82</td>
<td>0.90</td>
<td>0.85</td>
</tr>
</tbody>
</table>

4 http://raweb.inria.fr/rapportsactivite/RA2010/pulsar/uid27.html
5 http://www-sop.inria.fr/teams/pulsar/EvaluationTool/ViSEvAI_Description.html
Figure 18. Events detection workflow

Figure 19. Blue bounding boxes correspond to Annotations (Reference data) and red detected objects. It's possible to observe if there is a detection issue on the left image.
6.13. Group interaction and group tracking for video-surveillance in underground railway stations

Participants: Sofia Zaidenberg, Bernard Boulay, Carolina Garate, Duc-Phu Chau, Etienne Corvéé, François Brémond.

One goal in the European project VANAHEIM is the tracking of groups of people. Based on frame to frame mobile object tracking, we try to detect which mobiles form a group and to follow the group through its lifetime. We define a group of people as two or more people being close to each other and having similar trajectories (speed and direction). The dynamics of a group can be more or less erratic: people may join or split from the group, one or more can disappear temporarily (occlusion or disappearance from the field of view) but reappear and still be part of the group. The motion detector which detects and labels mobile objects may also fail (misdetections or wrong labels). Analyzing trajectories over a temporal window allows handling this instability more robustly. We use the event-description language described in [50] to define events, described using basic group properties such as size, type of trajectory or number and density of people and perform the recognition of events and behaviors such as violence or vandalism ( alarming events) or a queue at the vending machine (non-alarming events). Two approaches to this problem have been implemented. The first approach takes as input the frame-to-frame tracking results of individual mobiles and tries to gather them into groups based on their trajectories through the temporal window. Each group has a coherence coefficient. This coefficient is a weighted sum of three quantities characterizing a group: the group density (average of distances between mobiles), the similarity of mobile’s speed and the similarity of their motion directions. The update of a group consists in re-calculating the group coherence with new mobiles from the current frame. If adding the mobile does not put the coherence under a defined threshold, the mobiles are added to the group. A pre-selection is made by only considering mobiles that are close enough to the center of gravity of the group. After the update step, all mobiles that have not been assigned to a group are analyzed to form new groups if possible.

A first improvement has been done by integrating the use of the LBP-based people detector described in [36]. This makes the algorithm more robust to false detections such as train doors closing. But on the other hand, it also introduces false negatives as, among other things, people are only detected if fully visible in the image. The group tracking algorithm has been tested both with the original, background subtraction-based mobile object detection (noted S hereafter) and the LBP-based people detection (noted LBP hereafter).

For evaluating the detection, we used 3 annotated sequences: Sequence 1 is a short sequence of 128 frames with just one ground truth object (one group), Sequence 2 has 1373 frames and 9 ground truth objects, and Sequence 3 is 17992 frames long and 25 ground truth objects were annotated. Detection and tracking results are shown in table 4.

The whole algorithm chain has been integrated into the common VANAHEIM platform and sent to partners for pre-integration.

We also used videos from the ViCoMo project, recorded in the Eindhoven airport to test our approach. No formal evaluation has been done yet on these sequences due to the lack of ground truth. Nevertheless, these videos contain several acted scenes which could be successfully recognized: groups merging, splitting and entering a forbidden zone.

This work has been published in [50].

In parallel, a new approach is being developed, making use of a long-term tracker described in [35]. This tracker provides more robust individual trajectories to the group tracker, containing less confusions in cases where people cross each other. We apply the Mean Shift clustering algorithm on trajectories of people through a sliding time window (e.g. 10 frames). If the target is lost in one or several frames, we interpolate its positions.

The clustering brings together mobiles having similar trajectories, which is our definition of a group. At each frame, clusters are calculated and then a matching is done to associate clusters to existing groups in the previous frame, and thus track groups. Looking backwards (within a window) on the trajectory of a mobile we might find a mobile on that trajectory that belongs to a group. If such a group is found, it is called the probable
Table 4. Segmentation (S) and Human Detector (HD) Results

<table>
<thead>
<tr>
<th></th>
<th>Sequence 1</th>
<th></th>
<th>Sequence 2</th>
<th></th>
<th>Sequence 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S</td>
<td>HD</td>
<td>S</td>
<td>HD</td>
<td>S</td>
<td>HD</td>
</tr>
<tr>
<td>True Positives (TP)</td>
<td>72</td>
<td>67</td>
<td>1395</td>
<td>1079</td>
<td>5635</td>
<td>3679</td>
</tr>
<tr>
<td>False Positives (FP)</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>111</td>
<td>1213</td>
<td>642</td>
</tr>
<tr>
<td>False Negatives (FN)</td>
<td>6</td>
<td>11</td>
<td>269</td>
<td>585</td>
<td>3686</td>
<td>5642</td>
</tr>
<tr>
<td>Precision (global)</td>
<td>1</td>
<td>1</td>
<td>0.99</td>
<td>0.90</td>
<td>0.82</td>
<td>0.85</td>
</tr>
<tr>
<td>Sensitivity (global)</td>
<td>0.92</td>
<td>0.84</td>
<td>0.83</td>
<td>0.65</td>
<td>0.60</td>
<td>0.40</td>
</tr>
<tr>
<td>Tracking confusion</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.99</td>
<td>0.92</td>
<td>0.96</td>
</tr>
</tbody>
</table>

The group of the current mobile. Each trajectory cluster is associated to the group that is the probable group of most mobiles in the cluster. Several clusters may be associated to the same group. This cluster association makes the algorithm robust to cases where one or several mobiles temporarily separate from the group. If the separation is longer than the time window, the probable group of these mobiles will be empty and a split will be detected.

Additionally, we work on improving the people detection by combining both methods: background subtraction-based and LBP-based. We compare overlapping mobiles from both methods and choose the best one based on their respective confidence values and their sizes. If a target was detected by only one of the two methods, we keep the target given that the confidence is high enough. If a mobile from the background subtraction method is big enough to cover several LBP-detected people (the LBP-based people detection output targets have the size of a human, whereas the background subtraction can detect a bigger mobile with the size of a GROUP_OF_PEOPLE), we attach the LBP-people as sub-mobiles of the group mobile so no information is lost. This method is a work in progress and no evaluation have been done yet.

Figure 20 shows two examples of group and event detection.


Participants: Piotr Bilinski, François Brémond.

The aim of this work is to learn and recognize short human actions in videos. We perform an extensive evaluation of local spatio-temporal descriptors, then we propose a new action recognition approach for RGB camera videos. We also propose a new approach for RGB-D cameras. For all our experiments, we develop an evaluation framework based on the bag-of-words model, SVM and cross-validation technique. We use the bag-of-words model to represent actions in videos and we use non-linear multi-class Support Vector Machines together with leave-one-person-out cross-validation technique to perform action classification.

Local spatio-temporal descriptors have shown to obtain very good performance for action recognition in videos. Over the last years, many different descriptors have been proposed. They are usually evaluated using too specific experimental methods and using different datasets. Moreover, existing evaluations make assumptions that do not allow to fully compare descriptors. In order to explore capabilities of descriptors, we perform an extensive evaluation of local spatio-temporal descriptors for action recognition in videos. Four widely used state-of-the-art descriptors (HOG, HOF, HOG-HOF and HOG3D) and four video datasets (Weizmann, KTH, ADL and KECK) have been selected. In contrast to other evaluations, we test all the computed descriptors, we perform experiments on several codebook sizes and use several datasets, differing in difficulty. Our results show how the recognition rate depends on the codebook size and the dataset. We
observe that usually the HOG descriptor alone performs the worst but outperforms other descriptors when it is combined with the HOF descriptor. Also, we observe that smaller codebook sizes lead to consistently good performance across different datasets. This work has been published in [32].

We also propose a new action recognition method for RGB camera videos based on feature point tracking and a new head estimation algorithm. We track feature points along a video and compute appearance features (HOG-HOF) for each trajectory. Additionally, we estimate a head position for each visible human in the video, using the following chain: segmentation, person, head and face detectors. Finally, we create an action descriptor based on the combination of all these sources of information. Our approach has been evaluated on several datasets, including two benchmarking datasets: KTH and ADL, and our new action recognition dataset. This new dataset has been created in cooperation with the CHU Nice Hospital. It refers to people performing daily living activities like: standing up, sitting down, walking, reading a magazine etc.

We also study the usefulness of low-cost RGB-D camera for action recognition task. We propose a new action recognition method using both RGB and depth information. We track feature points using RGB videos and represent trajectories in a four-dimensional space using additionally depth information. Experiments have been successfully performed on our new RGB-D action recognition dataset, recorded using Microsoft’s Kinect device.

6.15. Activity Recognition Applied on Health Care Application

Participants: Rim Romdhane, Veronique Joumier, François Brémond.

The aim of this work is to propose a constraint-based approach for video event recognition with probabilistic reasoning for handling uncertainty. This work was validated on health care applications.

6.15.1. Event Recognition

We propose an activity recognition framework which is able to recognize composite events with complex temporal relationships. We consider different aspects of the uncertainty of the recognition during the event modeling and the event recognition process to overcome the noise or missing observations which characterize real world applications.
To reach this goal, we manage the uncertainty in the event modeling and event recognition processes by a combination of logical and probabilistic reasoning for handling uncertainty. We improve the event description language developed in Pulsar team and introduce a new probabilistic description based approach to gain in flexibility for event modeling by adding the notion of utility. Utility expresses the importance of sub-events to the recognition of the whole event. We compute the probability of recognition for both primitive (i.e. elementary) events and composite events based on Bayesian theory.

We compute the probability that the event $e$ is recognized given a sequence of observations $O$ as described in [48]. The observations consist of the set of the physical objects $po_e$ moving in the scene. If the probability of an event is over a predefined threshold, the event is recognized.

### 6.15.2. Health Care Application

The proposed event recognition approach is validated using the videos from the health care application SWEETHOME (http://cmrr-nice.fr/sweethome/) and CIUSante (https://extranet.chu-nice.fr/ciu-sante). We have worked in close collaboration with clinicians from Nice hospital to evaluate the behaviours of Alzheimer patients. We have first model 69 event models for health care application using our event modeling formalism. With the help of clinicians we have established a scenario protocol. The scenario is composed of three parts: (1) directed activities (10 min), (2) semi directed activities (20 min), (3) free activities (30 min). Experiments have been performed in a room of Nice hospital equipped with 2 video cameras where 45 elderly volunteers have spent between 15 min to 1 hour. Volunteers include Alzheimer patients, MCI (mild cognitive impairment) and healthy elderly.

The study described in [38] and [27] shows the ability of the proposed automatic video activity recognition system to detect activity changes between elderly subjects with and without dementia during a clinical experimentation. A total of 28 volunteers (11 healthy elderly subjects, 17 Alzheimer’s disease patients (AD)) participate to the experimentation. The proposed study shows that we could differentiate the two profiles of participants based on motor activity parameters, such as the duration of the recognized activities, the strike length and the walking speed, computed from the proposed automatic video activity recognition system. These primary results are promising and validating the interest of automatic analysis of video as an objective evaluation tool providing comparative results between participants and over the time.


Participants: Guillaume Charpiat, Vincent Martin, François Brémond, Monique Thonnat.

In collaboration with Victor Moncada, Jean-Marcel Travere and Thierry Loarer (CEA Cadarache), we propose a cognitive vision-based system for the intelligent monitoring of tokamaks during plasma operation, based on multi-sensor data analysis and symbolic reasoning. The practical purpose is to detect and characterize in real time abnormal events such as hot spots measured through infrared images of the in-vessel components in order to take adequate decisions. Our system is made intelligent by the use of a priori knowledge of both contextual and perceptual information for ontology-driven event modeling and task-oriented event recognition. The system is made original by combining both physics-based and perceptual information during the recognition process. Real time reasoning is achieved thanks to task-level software optimizations. The framework is generic and can be easily adapted to different fusion device environments. The developed system and its achievements on real data of the Tore Supra tokamak imaging system can be found in [39].

### 6.17. Scenario Recognition with depth camera

Participants: Bernard Boulay, Daniel Zullo, Swaminathan Sankaranarayanan, François Brémond.

Thanks to Microsoft and its kinect sensor, RGB-depth camera becomes popular and accessible. The basic idea of depth camera is to combine a visible camera, with an IR camera associated to a laser to determine the depth of each image pixel. This kind of sensor is well adapted for applications which monitor people (e.g. monitoring Alzheimer patient in hospital): because the people are in a predefined area and near the camera.
The depth cameras have two main advantages: first, the output image contains depth information and second, the sensor is independent from the light changes (IR sensor).

In our work, we propose to use the Kinect sensor to acquire 3D images, detect the people and recognize interesting activities (see Figure 21).

Figure 21. People detection and activity recognition in day and night conditions with Kinect sensor.

The nestk library is used to manage the Kinect sensor. This library is based on OpenNI framework (an open source driver) to acquire the images. Moreover, the library is able to compute some treatments (e.g., people detection) and to provide a true 3D map of the scene in the referential of the Kinect.

Basic attributes are computed for each detected person: 3D position, 3D height, ... These attributes are then used to compute more complex information: speed, global posture to recognize interesting activities thanks to the ScreKs framework (scenario recognition based on expert knowledge) of the SUP platform as walking, stopping, standing, sitting, ... (following a protocol delivered by doctors). Then we have a plug and play system able to recognize basic activities associated to a person.

Moreover, if information on the scene, as interesting zones, or equipment are available, complex activities can be recognized as nurse explaining exercise or nurse switching off the light.

The next step is to use the human skeleton detection to recognize precisely the posture of the patient in order to understand more precise activities and infer a behaviour model.

6.18. Trajectory Clustering for Activity Learning

Participants: Jose Luis Patino, Guido Pusiol, Hervé Falciani, Nedra Nefzi, François Brémond, Monique Thonnat.

The discovery, in an unsupervised manner, of significant activities observed from a video sequence, and its activity model learning, are of central importance to build up a reliable activity recognition system. We have deepened our studies on activity extraction employing trajectory information. In previous work we have shown that rich descriptors can be derived from trajectories; they help us to analyze the scene occupancy and its topology and also to identify activities [67], [68], [70], [55]. Our new results show how trajectory information can be more precisely employed, alone or in combination with other features for the extraction of activity patterns. Three application domains are currently being explored: 1) Monitoring of elderly people...
295

Vision, Perception and Multimedia Understanding - New Results - Project-Team PULSAR

at home; 2) Monitoring the ground activities at an airport dock-station (COFRIEND project \(^6\)); 3) Monitoring activities in subway/street surveillance systems.

6.18.1. Monitoring of elderly people at home

We propose a novel framework to understand daily activities in home-care applications; the framework is capable of discovering, modeling and recognizing long-term activities (e.g. “Cooking”, “Eating”) occurring in unstructured scenes (i.e. “an apartment”).

The framework links visual information (i.e., tracked objects) to the discovery and recognition of activities by constructing an intermediate layer of primitive events automatically.

The primitive events characterize the global spatial movements of a person in the scene (“in the kitchen”), and also the local movements of the person body parts (“opening the oven”). The primitive events are built from interesting regions, which are learned at multiple semantic resolutions (e.g. the “oven” is inside the “kitchen”). An example of the regions and possible activities for a single resolution is displayed in Fig. 22.

![Figure 22. Example of some learned regions and possible activities.](http://www-sop.inria.fr/pulsar/personnel/Guido.Pusiol/Home4/index.php)

A probabilistic model is learned to characterize each discovered activity. The modeled activities are automatically recognized in new unseen videos where a pop-up with a semantic description appears when an activity is detected. Examples of semantic labels are illustrated in Fig. 23 (a, b, c).

Recently we introduced 3D (MS. Kinect) information to the system. The preliminary results show an improvement superior to the 30% of the recognition quality. Also, the system can recognize activities in challenging situations as the lack of light. -See Fig.2 (b) and Fig.2 (c) -.

The approach can be used to recognize most of the interesting activities in a home-care application and has been published in [ 43 ]. Other examples and applications are available online in http://www-sop.inria.fr/pulsar/personnel/Guido.Pusiol/Home4/index.php.

6.18.2. Monitoring the ground activities at an airport dock-station

The COFRIEND project aims at creating a system for the recognition and interpretation of human activities and behaviours at an airport dock-station. Our contribution is a novel approach for discovering, in a unsupervised manner, the significant activities from observed videos. Spatial and temporal properties from detected mobile objects are modeled employing soft computing relations, that is, spatio-temporal relations graded with

different strengths. Our system works off-line and is composed of three modules: The trajectory speed analysis module, The trajectory clustering module, and the activity analysis module. The first module is aimed at segmenting the trajectory into segments of fairly similar speed (tracklets). The second aims at obtaining behavioural displacement patterns indicating the origin and destination of mobile objects observed in the scene. We achieve this by clustering the mobile tracklets and also by discovering the topology of the scene. The latter module aims at extracting more complex patterns of activity, which include spatial information (coming from the trajectory analysis) and temporal information related to the interactions of mobiles observed in the scene, either between themselves or with contextual elements of the scene. A clustering algorithm based on the transitive closure calculation of the final relation allows finding spatio-temporal patterns of activity. An example of discovery is given in the figure below. This approach has been applied to a database containing near to 25 hours of recording of dock-station monitoring at the Toulouse airport. The discovered activities are: ‘GPU positioning’, ‘Handler deposits chocks’, ‘Frontal unloading operation’, ‘Frontal loading operation’, ‘Rear loading operation’, ‘Push back vehicle positioning’. An example of discovered activity (Frontal loading) is given on the figure 24. When comparing our results with explicit ground-truth given by a domain expert, we were able to identify the events in general with a temporal overlap of at least 50%. The comparison with a supervised method on the same data indicates that our approach is able to extract the interesting activities signalled in the ground-truth with a higher True Positive Rate (74% TPR for the supervised approach against 80% TPR with our unsupervised method). This work has been published in [42].

6.18.3. Monitoring activities in subway/street surveillance systems

Figure 23. Examples of the recognized activities in 3D and under different light conditions.

Figure 24. Example of an activity cluster obtained. The left panel presents the tracklets of the mobiles participating in the Frontal Loading activity. Filled circles indicate the beginning of a tracklet. Empty circles indicate the end of a tracklet. The right panel presents the start frame of the activity.
In this work we have built a system to extract from video and in an unsupervised manner the main activities that can be observed from the a subway scene. We have setup a processing chain broadly working on three steps: The system starts in a first step by the unsupervised learning of the main activity areas of the scene. In a second step, mobile objects are then characterized in relation to the learned activity areas: either as ‘staying in a given activity zone’ or ‘transferring from an activity zone to another’ or a sequence of the previous two behaviours if the tracking persists long enough. In a third step we employ a high-level relational clustering algorithm to group mobiles according to their behaviours and discover other characteristics from mobile objects which are strongly correlated. We have applied this algorithm to two domains. First, monitoring two hours of activities in the hall entrance of an underground station and showed what are the most active areas of the scene and how rare/abnormal (going to low occupied activity zones) and frequent activities (e.g. buying tickets) are characterized. In the second application, monitoring one hour of a bus street lane, we were able again to learn the topology of the scene and separate normal from abnormal activities. When comparing with the available ground-truth for this application, we obtained a high recall measure (0.93) with an acceptable precision (0.65). This precision value is mostly due to the different levels of abstraction between the discovered activities and the ground-truth. The incremental learning procedure employed in this work is published in [ 52 ] while the full activity extraction approach was published in [ 41 ].

6.19. SUP Software Platform

Participants: Julien Gueytat, Leonardo Rocha, Daniel Zullo, François Brémond.

SUP is a software platform developed by PULSAR team, written in C and C++ for generating activity recognition systems. These systems should be able to perceive, analyze, interpret and understand a 3D dynamic scene observed through a network of sensors.

These activity recognition systems are a combination of algorithms developed by members of Pulsar or state of the art computer vision libraries. The SUP dissemination is targeted for use in real-world applications requiring high-throughput.

SUP is made as a framework allowing several computer vision workflows to be implemented. Currently, the workflow is static for a given application but our goal is to make it dynamic. A given workflow is the composition of several plugins, each of them implementing an algorithmic step in the video processing chain (i.e. the segmentation of images, the classification of objects, etc.). The design of SUP allows to execute at run-time the selected plugins.

During 2011 several tasks have been accomplished:

- A stable packaged release is available
- 3D simulation from a scenario description
- Existing algorithms have been improved in performance and accuracy
- Kinect sensor has been added to the hardware supported

Several plugins are available:

- 2 plugins are wrappers on industrial implementations of video processing algorithms (made available by Keeneo). They allow a quick deployment of a video processing chain encompassing image acquisition, segmentation, blob construction, classification and short-term tracking. These algorithms are robust and efficient algorithms, but with the drawback that some algorithms can lack accuracy.
- Several implementations by the Pulsar team members which cover the following fields:
  1. Image acquisition from different types of image and camera video streaming.
  2. Segmentation removing the shadows.
  3. Two classifiers, one being based on postures and one on people detection.
  4. Four frame-to-frame trackers, using as algorithm:
1. a simple tracking by overlapping,
2. neural networks,
3. tracking of feature points,
4. tracking specialized for the tracking of people in a crowd.
5. Three scenario recognizers, one generic algorithm allowing expression of probabilities on the recognized events, the second one focusing on the recognition of events based on postures and the third one (see section Extendable Event Recognition algorithm: SED in this document) uses the complete ontology of the domain as a parameter (e.g. the definition of objects of interest, scenario models, etc.).
6. 3D animation generation, it generates a virtual 3D animation from information provided by different plugins of the processing chain together with 3D contextual environment.
7. 3D simulation from description, it generates a virtual 3D animation from information provided from a text file with the description of the scenario.

From a software engineering point-of-view, the goal is to obtain a flexible platform being dynamically reconfigurable for the generated scene understanding systems to be autonomous and adaptable for handling changing environment.

SUP relies on DTK, a generic platform developed by the DREAM service at INRIA Research Center Sophia-Antipolis Méditerranée.

The purpose of DTK is to provide a software infrastructure allowing the generation of a new system by the composition of plugins, each plugin being an algorithmic step of the whole processing chain. SUP is oriented to help developers building activity recognition systems and describing their own scenarios dedicated to specific applications. By relying on the DTK software infrastructure, the possibilities are:

- To simplify the exchanges of algorithms between different INRIA teams using the DTK.
- To use the facilities already provided by the DTK allowing to compose quickly existing plugins. Currently a python interface is operational, and we plan to take advantage of the graphical composer to prototype quickly new work-flows, or reconfigure existing ones, for the experimentation conducted by the team.

In order to be confident on the results obtained with the SUP platform, an important effort is done to check:

- The correct behavior of the platform from a software engineering point of view, i.e. that the functionality of the SUP software is correctly provided, or is not broken by modifications.
- A qualitative evaluation tool (see ViSEvAl in this document) for the algorithms, which compares and assesses the results obtained with the algorithms to ground truth for several reference videos.

Both kinds of test are performed on a daily basis and on several hardware/software architectures.

### 6.20. Model-Driven Engineering and Video-surveillance

**Participants:** Sabine Moisan, Jean Paul Rigault, Sagar Sen, François Brémont.

In the framework of our research on model engineering techniques for video-surveillance systems, we have focused this year on the runtime adaptation of such systems.

Video-surveillance systems are complex and exhibit high degrees of variability along several dimensions. At the specification level, the number of possible applications and type of scenarios is large. On the software architecture side, the number of components, their variations due to possible choices among different algorithms, the number of tunable parameters... make the processing chain configuration rather challenging. Moreover, the context of an application may change in real time, requiring dynamic reconfiguration of the chain. This huge variability raises problems at design time (finding the configurations needed by the chain, foreseeing the different possible contexts), at deployment time (selecting the initial configuration), and at run time (switching configurations to react to context changes).
The first step was to formalize in an unified way all the necessary concerns —at the specification as well as at the component level— and their relations. To this end, we rely on Feature Models and (semi) automatic model transformations. Feature Models are widely used to represent systems with many possible variation points. Moreover they are liable to formal analysis (using propositional logic and satisfiability techniques) and thus lead to valid configurations, by construction. We have developed two feature models, one for the specification of the application (type of application, context of execution, expected quality of service, etc.) and one for the implementation representation (components and their assembly). Each model has its own internal constraints. Moreover, the two models are not independent: they are connected by cross model transformation rules that formalize the bridge between application requirements and component assemblies that realize them.

Second, we propose a framework to derive valid possible system configurations and to adapt running configurations to context changes. Users can select features describing their application in the specification model, through a simple graphic interface. The outcome is a sub-model of the specification model. Based on the cross model transformation rules, our framework automatically transforms this sub-model into a sub-model of the component model. The latter represents all possible component configurations of the target video-surveillance system that satisfy the specifications. Both sub-models will be kept throughout the system life. They are used while the system is running to adjust its configuration in response to context changes.

To achieve this dynamic adaptation, our framework sets up three collaborating modules as shown in Figure 25:

- the Run Time Component Manager (RTCM) captures low level events manifesting context changes (e.g., lighting changes); it forwards them to the Configuration Adapter which returns a

---

7Here the “features” correspond to selectable concepts of the systems; they can be at any abstraction level (a feature may correspond to an specification entity such as “Intrusion detection” or to a more concrete element such as “High frame rate”). The features are organized along a tree, with logical selection relations (optional, mandatory features, exclusive choices...) and some constraints that restrict the valid combinations of features (i.e., configurations).
new component configuration; the RTCM is then responsible for applying this configuration, that is to tune, add, remove, or replace components, and possibly to change the workflow itself.

- the **Configuration Adapter** (CA) receives change events from the RTCM, translates them into the feature formalism, and forwards the result to the Model Manager; in return, it obtains a sub-model of component configurations compatible with the change; this sub-model is a compact representation of a set of valid configurations and the CA is responsible to select one and to instruct the RTCM to apply it; this selection uses some heuristics, possibly based on a cost function such as minimizing the number of component changes in the processing chain or maximizing the quality of service (e.g., accuracy, responsiveness).

- the **Model Manager** (MM) manages the representation of the two specialized Feature Models corresponding to the specification and possible component assemblies of the current application together with their constraints; its role is to enforce configuration validity. It is also responsible of the set of rules relating run time events and (de)selection of features in both models. From the CA, the Model Manager receives information about incoming events; it uses the rules to select or deselect the corresponding features; it then applies constraints, rules, and model transformations to infer a component sub-model that represents a subset of valid component configurations and that it returns to the CA.

This year, we have tested our approach on simple applications using well-known libraries (OpenCV) on different scenarios. At the moment, 77 features and $10^5$ configurations are present in the specification model while 51 features and $10^6$ configurations are present in the component model. Once the video surveillance designer has selected the features required by an application, before deployment, the average number of features to consider at runtime in the component model is less than $10^4$. The configuration spaces is reduced by several orders of magnitude and enables the use of the other tools in the end-to-end engineering process, whereas it would not have been possible without. Our experiments show the feasibility of such an approach with a limited performance overhead (if any) compared to traditional run time control where ad hoc adaptation code is hardwired and does not rely on the run time availability of an abstract representation of the application and its context evolution [29]. The next step will be to test our approach on our SUP platform and to study intelligent configuration selection heuristics.

### 6.21. The Girgit Software

**Participants:** Leonardo Rocha, Sabine Moisan, Jean-Paul Rigault, Sagar Sen.

Girgit is a Python based framework to build context-aware self-adaptive software systems.

Girgit is a simple and small [1] framework that allows dynamic reconfiguration of data processing chains and accepts any set of components for the configuration as long as they have the corresponding Python wrapper.

The basic idea of Girgit is to provide a platform to be able to make dynamic adaptive systems. It provides a dynamic adaptive engine that can deal with event/action pairs called rules and also provides an application programming interface to be able to use it.

As the system is designed to be interactive, any user can interact with it, be a human operator or a reasoning engine. Rules can be pre-loaded with the configuration, and components can launch events, this allows pre-configured rules to be loaded at launch time of the Girgit.

Girgit has been used to evaluate the performance on real time video applications and show the architecture. Three publications [47], [45], [46] where based on the framework.

### 6.22. Scenario Analysis Module

**Participants:** Sabine Moisan, Annie Ressouche, Jean Paul Rigault.
To generate activity recognition systems we supply a scenario analysis module (SAM) to express and recognize complex events from primitive events generated by SUP or others sensors. In this framework, this year we focus on recognition algorithm improvement in order to face the problem of large number of scenario instances recognition.

The purpose of this research axis is to offer a generic tool to express and recognize activities. Genericity means that the tool should accommodate any kind of activities and be easily specialized for a particular framework. In practice, we propose a concrete language to specify activities in the form of a set of scenarios with temporal constraints between scenarios. This language allows domain experts to describe their own scenario models. To recognize instances of these models, we consider the activity descriptions as synchronous reactive systems and we apply general modeling methods to express scenario behaviors. This approach facilitates scenario validation and allows us to generate a recognizer for each scenario model. The SAM module thus provides users with (1) a simulation tool to test scenario behaviors; (2) a generator of a recognizer for each scenario model; (3) an exhaustive verification of safety properties relying on model checking techniques our approach allows. The latter offers also the possibility to define safety properties to prove as “observers” expressed in the scenario language.

Last year we have completed SAM in order to address the life cycle of scenario instances. For a given scenario model there may exist several (possibly many) instances at different evolution states. These instances are created and deleted dynamically, according to the input event flow. The challenge is to manage the creation/ destruction of this large set of scenario instances efficiently (in time and space), to dispatch events to expecting instances, and to make them evolve independently. This year, to face this challenge we first replace some operators of the language, by others having a more strict semantics. For instance, we replace the before operator whose semantics allowed that events can meet, by two operators, a strict before and a meet. Hence, the number of events a scenario instance reacts to decreases. Second, we now generate within the recognition engine, the expected events of the next step. This avoids to run the engine automatically with events that are not relevant for the recognition process.

Presently, we still rely on the existing synchronous language (Lustre) to express the equational semantics of scenario models and to generate recognizers because this language offers simulation and verification means. But, to improve efficiency, we plan to build our own compiler and to generate recognizers directly from the Boolean equation systems modeling scenario models. This implies that we must supply our own simulation tool and that we interface with a model checking tool as NuSMV.

Now the challenge is to take into account some uncertainty on the primitive events due to input sensor errors. In the family of synchronous languages, the Lutin language could be able to automate the generation of realistic input sequences of events, taking into account probabilistic distributions over primitive events. In other words, it could generate a set of input events for which a set a constraints can be verified. In complement, it offers also means to compute the real values verifying these constraints. Thus, we think to rely on Lutin to express uncertainty on primitive events and get input events to feed scenario recognition engines.

6.23. Multiple Services for Device Adaptive Platform for Scenario Recognition

Participants: Annie Ressouche, Jean-Yves Tigli.

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 6.19 ) with a service oriented middleware approach dealing with dynamic evolutions of system infrastructure. The Rainbow team (Nice-Sophia Antipolis University) proposes a component-based adaptive middleware (WComp) to dynamically adapt and recompose assemblies of components. These operations must obey the “usage contract” of components. The existing approaches don’t really ensure that this usage contract is not violated during application design. Only a formal analysis of the component behavior models associated with a well sound modeling of composition operation may guarantee the respect of the usage contract.

http://www-verimag.imag.fr/Lutin.html
The approach we adopted introduces in a main assembly, a synchronous component for each sub assembly connected with a critical component. This additional component implements a behavioral model of the critical component and model checking techniques apply to verify safety properties concerning this critical component. Thus, we consider that the critical component is validated.

When a critical component has multiple synchronous monitors corresponding to several concern managements in the application, we want to build an only synchronous model component which agrees with all these primitive synchronous monitors. To specify how output events sent by different synchronous monitors and connected to a critical component, we introduce a sound (with respect to our mathematical formalism) operation of composition under constraints of synchronous models (see figure 26). We proved that this operation preserves already separately verified properties of synchronous components. This operation is an answer to the multiple access to critical components. Actually, we supply a graphical interface to design both critical component behaviors and properties as observers in the synchronous language Lustre [62]. Then the validation of properties and the creation of the validated synchronous component is automatic [44], [53].

Figure 26. Example of multiple access to alarm critical component: (a) SM - the synchronous monitors; (b) CSM - the composition under constraints operation

This year we focus on the main challenge of this approach which is to deal with the possibly very large number of constraints a user must specify. Indeed, each synchronous monitor has to tell how it combines with others, then we get a combinatorial number of constraints with respect to the number of synchronous monitors and inputs of the critical component. To be adaptive with efficiency, we must face this problem. We first introduced some default rules to avoid the user to express a large number of constraints. We also studied how Abstract Interpretation technique can help us to reduce this complexity. This approach works if we forbid some “non monotonic” constraints, but this is a strong limitation. Thus, it is still a challenge for us. This drawback is a popular challenge in adaptive middleware and some results exist relying on controller synthesis methods. We are not in the exact framework where these techniques apply, but we plan to study if we can rely on some extension of these techniques.

On another hand, we also want to complement our preservation result in studying how the proof of a global property can be decomposed into the proof of local ones. In general, this decomposition (known as assume-
guarantee paradigm) is difficult to apply but there is no communication between synchronous monitors and so the decomposition could be tractable. Moreover, some works address this problem and we can rely on them.

6.24. The Clem Workflow

Participants: Annie Ressouche, Daniel Gaffé.

This research axis concerns the theoretical study of a synchronous language LE with modular compilation and the development of a toolkit (see figure 6 ) around the language to design, simulate, verify and generate code for programs. The novelty of the approach is the ability to manage both modularity and causality. Indeed, only few approaches consider a modular compilation because there is a deep incompatibility between causality and modularity. Thus, relying on semantics to compile a language ensures a modular approach but requires to complete the compilation process with a global causality checking. To tackle this problem, we introduced a new way to check causality from already checked sub programs and the modular approach we infer. The equational semantics compute a Boolean equation system and we ensure both modularity and causality in computing all the partial orders valid for a system and we define a way to merge two partial orders. The algorithm which computes partial orders rely on the computation of two dependency graphs: the upstream (downstream) dependency graph computes the dependencies of each variable of the system starting from the input (output) variables. This way of compiling is the corner stone of our approach. We defined three different approaches to compute the partial orders valid for an equation system:

1. apply PERT method: inputs (resp. outputs) have date 0 and recursively increase of dates for each vertice in the upstream (resp downstream) dependencies graph;
2. apply graph theory:
   – compute the adjacency matrix $\mathbb{U}$ of upstream (resp. downstream) dependencies graph;
   – the length of the maximal path from a variable $v$ to system inputs is characterized by the maximal $k$ such that $\mathbb{U}^k[v, i] \neq 0$ for all inputs $i$.
3. apply fix point theory: the vector of earliest (resp. lastest) dates can be computed as the least fix point of a monotonic increasing function.

The fix point characterization helps us to prove that the merge algorithm is correct (i.e we get the same partial orders using the merge algorithm on two previously sorted equation systems or when sorting the union of the two equation systems considered).

To be modular, we defined a technique to compose two already sorted equation systems: first, we memorize the two dependency graphs of equation systems. Second, we define two merge algorithms relying on two different techniques:

1. propagation of common variables dates adjustment;
2. fix point characterization starting with the vectors of already computed dates and considering only the variables in the dependencies (upstream and downstream) of common variables

This year we began the implementation of a separated compilation of LE programs, according to these theoretical results. We define a new intermediate format (lea) to record partially compiled module, i.e module whose Boolean equation systems may be composed of non defined variables (we called them abstract). Then we are implementing a refinement operation which replaces these abstract variables by their definition and performs adjustment of the dates. According to our theoretical results, we know that the resulting sorting is the same as with a global approach. After the termination of this separated compilation of LE programs, the challenge will be to use Clem to design a large application in the domain of smart cards. The application needs more than forty LE automata in parallel and the compiled code will have more than 500 registers and thousands variables. Only a separated compilation will work.

The Clem toolkit is completely described in [28]
TEMICS Project-Team

6. New Results

6.1. Analysis and modeling for compact representation and navigation

6.1.1. Joint projection/filling method for virtual view synthesis

Participants: Christine Guillemot, Vincent Jantet.

This study is carried out in collaboration with INSA/IETR (Luce Morin). Associated with a view synthesis method, a multi-view plus depth video allows the generation of virtual views of the scene from any viewpoint. Many algorithms have thus been developed to synthesize virtual views from one or several input views video plus depth data. These rendering algorithms are either based on Image-Based Rendering (IBR) techniques or Geometry-Based Rendering (GBR) techniques, according to the amount of 3D information they use. IBR techniques require limited geometric information to synthesize intermediate views and allow the generation of photo-realistic virtual views at the expense of virtual camera freedom. GBR techniques require detailed 3D models of the scene to synthesize arbitrary viewpoints (points of view). GBR techniques are sensitive to the accuracy of the 3D model, which is difficult to estimate from real multi-view videos. Depth-Image-Based Rendering (DIBR) techniques include hybrid rendering methods between IBR and GBR techniques. DIBR methods are based on warping equations, which project a reference view onto a virtual viewpoint. Each input view is defined by a "color" (or "texture") map and a "depth" map, which associates a depth value to each image pixel.

In classical DIBR schemes, the rendering proceeds in several distinct steps, each one designed to solve a specific problem. First, the input depth map is warped onto the virtual viewpoint. The obtained warped depth map contains disocclusions, cracks and ghosting artifacts. Second, this virtual depth map is filtered with a median filter, in order to remove the cracks, and then to dilate disocclusion areas on the background side, in order to avoid ghosting artifacts during view synthesis. Third, the filtered depth map is used in a backward warping to compute the color of each pixel of the virtual view. Fourth, this resulting depth map is inpainted, to fill in disocclusion areas. Finally, this complete depth map is used by a depth-aided inpainting algorithm to fill in disocclusions in the color map. However, all these steps are inter-dependent, and errors introduced by each one are amplified by the following one. Connectivity information is lost during the first projection step, as shown in Fig. 2 . Without this connectivity information, every inpainting method fails to fill in background disocclusions if the disoccluded area is surrounded by foreground objects. This case may happen each time a foreground object is not convex, and contains holes, as shown in Fig. 2 -(a). As a result, depth-aided inpainting uses wrong foreground patches to fill in background disocclusions, producing annoying artifacts, as shown in Fig. 2 -(b).

We have developed two DIBR techniques, both based on a novel forward projection technique, called the Joint Projection Filling (JPF) method [ 16 ]. The JPF method performs forward projection, using connectivity information to fill in disocclusions in a single step. The first proposed DIBR method is designed to extrapolate virtual views from a single input view plus depth video sequence. The synthesis of virtual depth maps by the JPF method avoids the use of dedicated filtering and inpainting processes and leads to synthesized depth maps of higher quality. The second proposed DIBR method is designed to interpolate intermediate views from multiple input view plus depth sequences. This interpolation method uses the Floating Texture approach to register multiple inputs view plus depth sequences before blending. The JPF method fills in disocclusion areas during the projection, to ensure that geometrical structures are well preserved. The method uses the occlusion-compatible ordering presented by McMillan, which uses epipolar geometry to select a pixel scanning order. The occlusion-compatible ordering is used to handle disocclusions gracefully. Cracks are filled in by interpolation of neighboring pixels, whereas disocclusions are only filled in by background pixels. This technique can also be used with non-rectified views, avoiding prior creation of parallax maps.
6.1.2. 2D/3D image inpainting for virtual view synthesis

Participants: Josselin Gauthier, Christine Guillemot, Mouid Keskes, Olivier Le Meur.

Inpainting methods play an important role in a wide range of applications. Removing text and advertisements (such as logos), removing undesired objects, noise reduction and image reconstruction from incomplete data are the key applications of inpainting methods. Algorithms can be classified into two categories: PDE (Partial Derivative Equation)-based schemes and exemplar-based schemes. The former uses diffusion schemes in order to propagate structures in a given direction. Their drawback is the introduction of blur due to diffusion. The latter relies on the sampling and the copying of texture from the known parts of the picture.

We have proposed a novel inpainting algorithm combining the advantages of both aforementioned methods. As in Criminisi et al’s approach, the proposed method involves two steps: first, a filling order is defined to favor the propagation of structure in the isophote direction. Second, a template matching is performed in order to find the best candidates to fill in the hole. Compared to previous approaches, the main contributions concern the use of structure tensors to define the filling order instead of field gradients. The structure tensor is defined as follow:

\[
J = \sum_{i \in \{R, G, B\}} \nabla I_i \nabla I_i^T \tag{2}
\]

\(J\) is the sum of the scalar structure tensors \(\nabla I_i \nabla I_i^T\) of each image channel \(I_i\) \((i \in \{R, G, B\})\). Information about local geometry can be deduced by computing the eigenvalues and eigenvectors of \(J\). The local vector geometry is computed from the structure tensor \(J\). Its eigenvectors \(v_{1,2}\) \((v_i \in \mathbb{R}^n)\) define an oriented orthogonal basis and its eigenvalues \(\lambda_{1,2}\) define the amount of structure variation. \(v_1\) is the orientation with the highest fluctuations (orthogonal to the image contours), and \(v_2\) gives the preferred local orientation. This

---

eigenvector (having the smallest eigenvalue) indicates the isophote orientation. The use of structure tensor allows to retrieve a more coherent local geometry. The computation of the filling order as proposed by Criminisi et al. is then replaced by a term coming from PDE-based schemes, called Coherence Enhancing Diffusion. The use of structure tensor in an exemplar-based approach leads to a more robust algorithm that visually improves the quality of the inpainted areas.

Additionally, the simple template matching originally used in previous methods has been improved by using a K-nearest neighbor approach. The weights of the linear combination of the first $K$ best candidate are adjusted by taking into account that all candidate patches are not equally reliable. Note that the number $K$ is also locally adjusted in function of the local spatial complexity.

The 2D inpainting algorithm described above has been extended to deal with 3D content. In this work, the goal is to synthesize novel views directly from the original images. Image-based rendering (IBR) is commonly used to render a virtual view. It generates a nearby viewpoint image by projecting a point from the reference view to the virtual view using the depth data. However, when the viewpoint is shifted, occluded regions in the original viewpoint are disoccluded. Handling these disocclusions (holes) is a difficult problem. We propose to use an extension of the 2D inpainting method to fill in these holes. For this goal, we have modified the computation of the structure tensor by adding the depth information. Equation (2) is simply modified as follow:

$$J = \sum_{i=\{R,G,B,Z\}} \nabla I_i \nabla I_i^T$$

where $Z$ represents the depth map. As previously, the tensor is used to compute the filling order. A directional term is also included in order to favor a filling direction. Specifically, when the viewpoint is shifted from left to right in the horizontal direction, occluded regions in the left image appear in the right image around the right side of the object. Therefore, it is recommended to start the filling from the right to the left. This filling is performed by a modified template matching using texture information as well as depth data. Figure 3 illustrates the inpainting quality for different approaches.

**Figure 3. Virtual synthesized view. From left to right: original view projected into the new viewpoint; disocclusions filled by Criminisi’s approach, Daribo’s approach and the proposed method.**

### 6.1.3. Computational modelling of visual attention

**Participants:** Josselin Gauthier, Olivier Le Meur.

#### 6.1.3.1. Eye-movement study:

In 2011, we have investigated whether two populations of visual fixation exist in 2D context. The question is simple: do all visual fixations have the same role in the free viewing of natural scenes? Recent studies suggest that there are at least two types of visual fixations: focal and ambient fixations. The former is believed to be used to inspect local areas accurately, whereas the latter is used to obtain the context of the scene.
From a collaboration with Technicolor (P. Guillotel and C. Chamaret) and LUTIN (T. Baccino), we found new evidence to support a focal-ambient dichotomy. Our results published in the journal i-Perception [14] indicate that the determining factor to classify the visual fixations is the saccade amplitude. We proposed an automatic system to cluster visual fixations in two groups using four types of natural scene images. From this automatic classification, the terms focal saliency map and ambient saliency map have been introduced. The dependence on the low-level visual features and the time course of these two kinds of visual fixations were examined. Our results demonstrate that there is an interplay between both fixation populations and that focal fixations are more dependent on low-level visual features than ambient fixations. These results might have a strong impact on both the computational modelling of visual attention and their performance assessment.

A second study related to eye-movement dealt with the role of the binocular disparity depth cue in the deployment of visual attention. To address this point, we compared eye tracking data recorded while observers viewed natural images in 2D and 3D conditions. The influence of disparity on the inter-observers congruency, saliency, center and depth bias was first examined. Results show that visual exploration in depth layer detection task is affected by the binocular disparity. In particular, participants tend to look first at closer areas just after the stimuli onset with the introduction of disparity, and then direct their gaze to more widespread locations. Our results has been submitted in the journal Cognitive Computation.

6.1.3.2. Model of visual attention:

Since 1998 with the publication of the influential work of Itti, Kock and Niebur[11], the computational modelling of the visual attention has known an increasing interest. The former models only used the low-level visual features for getting a saliency map. They perform well in a number of cases in predicting where an observer would look at. However, to improve the quality of the prediction, it seems unavoidable to use high-level information in order to account for visual deployment.

This work aims at designing a computational model mixing low-level and high-level features. Among the different factors influencing our gaze, we have focused our works on two cues: the dominant depth and the horizon line position. The dominant depth and the spatial position of the horizon line were inferred from the low-level visual features. A training database has been set up to perform a learning. Results indicate that the proposed model outperforms state-of-the-art models [37].

From behavioural studies on eye-movement in a 3D context, we have proposed a model of visual attention able to predict saliency of 3D pictures. The method developed aims at using the depth cue, the central bias and the low-level visual features. The predicted saliency is obtained by linearly combining these cues. The weights of the linear combination are learnt from a training database and are time-dependent. This study is under revision in the journal Cognitive Computation.

6.1.3.3. Predicting the inter-observer visual congruency:

This work aims at predicting the inter-observer visual congruency (IOVC), indicating the congruence or the variability among different subjects looking at the same image [35]. Predicting this congruence is of interest for image processing applications where the visual perception of a picture matters such as website design, advertisement, etc. We proposed a computational model of the IOVC. This new model is a mixture of low-level visual features extracted from the input picture. Model’s parameters are learned by using a large eye-tracking database. In this study, we also proposed a new scheme to compute the depth of field of a picture. Finally, once the training and the feature extraction have been carried out, a score ranging from 0 (minimal congruency) to 1 (maximal congruency) is computed. A value of 1 indicates that observers would focus on the same locations and suggests that the picture presents strong locations of interest. To illustrate the interest of the proposed model, we have used it to automatically rank personalized photograph. Figure 4 illustrates the proposed approach.

6.1.4. Visual cues analysis and modelling

Participants: Safa Cherigui, Christine Guillemot.

This work is carried out in collaboration with Technicolor (D. Thoreau, Ph. Guillotel, P. Perez) and aims at designing a compression algorithm based on the concept of epitomes. An epitome is a condensed representation of an image (or a video) signal containing the essence of the textural properties of this image. Different forms of epitomes have been proposed, such as a patch-based probability model learned either from still image patches or from space-time texture cubes taken from the input video. These probability models together with appropriate inference algorithms, are useful for content analysis inpainting or super-resolution. Another family of approaches makes use of computer vision techniques, like the KLT tracking algorithm, in order to recover self similarities within and across images. In parallel, another type of approach consists in extracting epitome-like signatures from images using sparse coding and dictionary learning.

The method developed aims at tracking self-similarities within an image using a block matching (BM) algorithm [25]. The epitome is constructed from disjoint pieces of texture (“epitome charts”) taken from the original image and a transform map which contains translational parameters (see Fig. 5-middle row). Those parameters keep track of the correspondences between each block of the input image and a block of the epitome. An Intra image compression scheme based on the epitome has been developed showing a rate saving of up to 12.

6.2. Representation and compression of large volumes of visual data

6.2.1. 3d representations for multi-view video sequences

Participants: Christine Guillemot, Vincent Jantet.

Multi-view plus depth video content represent very large volumes of input data which need to be compressed for storage and transmission to the rendering device. The huge amount of data contained in multi-view sequences indeed motivates the design of efficient representation and compression algorithms. In collaboration with INSA/IETR (Luce Morin), we have studied layered depth image (LDI) and layered depth video (LDV) representations as a possible compact representation format of multi-view video plus depth data. LDI give compact representations of 3D objects, which can be efficiently used for photo-realistic image-based rendering (IBR) of different scene viewpoints, even with complex scene geometry. The LDI extends the 2D+Z representation, but instead of representing the scene with an array of depth pixels (pixel color with associated depth values), each position in the array may store several depth pixels, organised into layers.

Various approaches exist to construct LDI, which all organize layers by visibility. The first layer contains all pixels visible from a chosen reference viewpoint. The other layers contain pixels hidden by objects in previous layers. With classical construction solutions, each layer may contain pixels from the background and pixels from objects in a same neighbourhood, creating texture and depth discontinuities within the same layer. These
Figure 5. Epitome and reconstructed images: (top row) Original images (columns a and b); (middle) Epitomes; (bottom) Reconstructed images from epitome texture and transform map.
discontinuities are blurred during the compression process which in turn significantly reduces the rendering quality.

We have thus developed a novel object-based LDI representation which is more tolerant to compression artifacts, as well as being compatible with fast mesh-based rendering techniques [34]. This representation organises LDI pixels into two separate layers (foreground and background) to enhance depth continuity (see Fig. 6). The construction of this object-based LDI makes use of a foreground-background region-growing segmentation algorithm followed by inpainting of both colour and texture images to have a complete background layer (without the holes corresponding to disocclusion areas). The costly inpainting algorithm is thus processed once, during the LDI classification, and not during each view synthesis, which helps to speed up the rendering step.

![Figure 6. Object-based LDI: (top) Foreground and background layers; (bottom) Rendering results classical and object LDI.](image)

### 6.2.2. From sparse to spread representations

**Participant:** Jean Jacques Fuchs.

Sparse representations, where one seeks to represent a vector on a redundant basis using the smallest number of basis vectors, appear to have numerous applications. The other extreme, where one seeks a representation that uses all the basis vectors, might be of interest if one manages to spread the information nearly equally over all of them. Minimizing the $\ell_\infty$-norm of the vector of weights is one way to find such a representation. Properties of the solution and a dedicated fast algorithm have been developed. While the application of such models in robust data coding and in improving achievable data rates over amplitude constrained channels seems to be wishful thinking, its use in indexing techniques appears to be promising. In this context, one further replaces the optimal vector by its sign vector (potentially associated with a re-evaluated scalar weight) to get a binary vector that is not only cheap to store and (somehow) easy to search for but also allows for an explicit reconstruction unlike all other Hamming embedding functions used to map real vectors into binary vectors.
6.2.3. **On-line dictionary learning methods for prediction**

**Participants:** Christine Guillemot, Mehmet Turkan.

One crucial question to the problem of sparse approximation, and hence also of prediction based on sparse approximation, is the choice of the dictionary. Various advanced dictionary learning schemes have been proposed in the literature for the sparse signal approximation problem, so that the dictionary used is well suited to the data at hand. The popular dictionary learning algorithms include the K-SVD, the Method of Optimal Directions (MOD), Sparse Orthonormal Transforms (SOT), and (Generalized) Principle Component Analysis (PCA). However, the above learning methods are often used off-line since their computational complexity, which results from the number and the dimension of training samples, makes them inappropriate for online learning. In addition, these methods are adapted to the learning of basis to be used for approximating input data vectors, but not to the problem of predicting unknown samples from noisy observed samples in a causal neighborhood.

In 2011, we have developed a method for on-line training dictionaries adapted to the prediction problem [41]. Let \( A \) be the input dictionary, which is divided into two sub-dictionaries: \( A_c \) and \( A_t \). The goal is to have a simple on-line dictionary learning method which is adapted to the prediction problem, i.e., which will learn both sub-dictionaries so that sparse vectors found by approximating the known samples (the template) using the first sub-dictionary \( A_c \) will also lead to a good approximation of the block to be predicted when used together with the second sub-dictionary \( A_t \). When dealing with the prediction problem, the sparse signal approximation is indeed first run with a set of masked basis functions (dictionary \( A_c \)), the masked samples corresponding to the location of the pixels to be predicted. The principle of the approach is to first search for the linear combination of basis functions which best approximates known sample values in a causal neighborhood, and keep the same linear combination of basis functions, but this time with the unmasked samples (dictionary \( A_t \)) to approximate the block to be predicted. The decoder similarly runs the algorithm with the masked basis functions and taking the previously decoded neighborhood as the known support. The use of masked basis functions converts the complete approximation problem into an overcomplete approximation problem. Because of its simplicity, of the limited number of training samples it requires, can be used for online learning of dictionaries, i.e. while doing the block-wise encoding of the image. The training samples are all possible previously coded/decoded texture patches (blocks of pixels) within a search window located in a causal neighborhood of the block to be predicted.

6.2.4. **Neighbor embedding methods for image prediction and inpainting**

**Participants:** Christine Guillemot, Mehmet Turkan.

The problem of texture prediction as well as image inpainting can be regarded as a problem of texture synthesis. Given observations, or known samples in a spatial neighborhood, the goal is to estimate unknown samples of the block to be predicted or of the patch to be filled in inpainting. We have developed texture prediction methods as well as a new inpainting algorithm based on neighbor embedding techniques which come from the area of data dimensionality reduction. The methods which we have more particularly considered are Locally Linear Embedding (LLE) and Non-negative Matrix Factorization (NMF). The first step in the developed methods consists in searching, within the known part of the image, for the \( K \) nearest (KNN) patches to the set of known samples in the neighborhood of the block to be predicted or of samples to be estimated in the context of inpainting. This first step can be seen as constructing a dictionary matrix by stacking in the matrix columns the vectorized K-NN texture patches. The non-negative dictionary matrix \( A \in \mathbb{R}^{N \times M} \), is formed by \( K \) nearest neighbors to the vector formed by the known samples in the neighborhood of the samples to be predicted. These \( K \) nearest neighbors are texture patches of the same shape taken from the known part of the image. This dictionary can then be used for approximating the known samples by masking the rows of the matrix which correspond to the position of the unknown samples, solving a least squares problem under the constraint of sum-to-one of the weights in the case of LLE, or under the constraint of non-negativity of the weights for NMF. It is actually a variant of NMF since one of the components matrices is fixed (the one corresponding to the dictionary matrix) and only the matrix containing the weights of the linear approximation must then be found. The approaches are thus intended to explore the properties of the manifolds on which the
input texture patches are assumed to reside. The underlying assumption is that the corresponding uncomplete and complete patches have similar neighborhoods on some nonlinear manifolds. The new prediction methods give RD performances which are significantly better than the ones given by the H.264 Intra prediction modes, in particular for highly textured images \[ 21 \], the highest gain being achieved with NMF.

A new examplar-based inpainting algorithm using neighbor embedding techniques has been developed. A new priority order has been proposed in order to inpaint first patches containing structures or contour information. The methods have also been shown to enhance the quality of inpainted images when compared to classical examplar-based solutions using simple template matching techniques to estimate the missing pixels, (see Fig. 7).

![Inpainting results](image)

**Figure 7.** Inpainting results: (left) mask of the image to be inpainted; (middle) Inpainting results with classical examplar-based inpainting; (right) Inpainting results with LLE (right).

### 6.2.5. Lossless coding for medical images

**Participants:** Claude Labit, Jonathan Taquet.

Last year, we developed a hierarchical oriented prediction (HOP) algorithm, for resolution scalable lossless and near lossless compression of biomedical images. In 2011, the algorithm has been slightly improved with an iterative optimization of the predictors in order to get better results on less noisy/smooth images \[ 39 \].

Recently, there have been a growing interest for the compression of an emerging imaging modality: the virtual microscopy (VM). It is used in anatomopathology and may produce huge images of more than 1 Gigabytes. We have studied the efficiency for lossless and lossy compression of our previously developed algorithms HOP and OWD (optimized wavelet decomposition) and of two extensions of OWD: near-lossless and/or region of interest (ROI) coding. The lossless results, which are slightly better than JPEG-LS and JPEG-2000 standards with about 3:1 compression ratio, show that lossless compression is not suited to VM. By compressing only the information area (ROI) which represents about 20 percents of the size of test images, 9:1 ratio could be obtained, and combined with near-lossless approach, depending on the required quality, ratio can reach 17:1 with no visual losses to more than 30:1 with some visual losses (or approximately about 6:1 for ROI only data). We have concluded that it would probably be better to use lossy or efficient quality scalable compression. Because those images have specific contents (cellular tissue for example) we have also introduced and investigated new learning based methods. We have developed an optimization process for designing multiple KLT (Karhunen-Loeve Transform) in order to get orthonormal bases that are optimal for decorrelation and quality scalability. This learning approach has been applied as an a-posteriori transform of a wavelet decomposition in order to propose transforms with no blocking artefacts. A fully quality-scalable coding algorithm allows to obtain interesting PSNR improvements compared to the optimized coding process.
of JPEG-2000. Gain is around 0.5 dB for 16:1 compression of ROI only data, and more than 1 dB for 8:1 compression ratio.

6.3. Distributed processing and robust communication

6.3.1. Loss concealment based on video inpainting

Participants: Mounira Ebdelli, Christine Guillemot, Olivier Le Meur.

In 2011, we have started developing a loss concealment scheme based on a new video examplar-based inpainting algorithm. The developed video inpainting approach relies on new patch priority functions as well as on a motion confidence-aided neighbor embedding techniques. Neighbor embedding approaches aim at approximating input vectors (or data points) as a linear combination of their neighbors. The search of the weights of the linear combination (i.e. of the embedding) are formulated as constrained least squares problems. When using the locally linear embedding, the constraint is that the sum of the weights is equal to 1. We have also considered non-negative matrix factorization to solve the problem, in which case the constraint is that the weights and the other vector are non-negative. The motion confidence introduced in the neighbor embedding improves the robustness of the algorithm in the sense that it limits the error propagation effects which otherwise result from uncertainties on the motion information of the unknown pixels to be estimated. A new patch similarity measure which accounts for the correlation between motion information has been defined for the $K$-NN search inherent to neighbor embedding techniques. Evaluations of the algorithm in a context of video editing (object removal) are on-going. The next step will be to assess the performance of the approach in a context of loss concealment, that is to estimate unknown pixels after decoding when the corresponding transport packets have been lost on the transmission network.

6.3.2. Unequal Erasure Protection and Object Bundle Protection

Participant: Aline Roumy.

In 2011, we started a new collaboration in the framework of the Joint INRIA/Alcatel Lucent lab. In this work, carried out with V. Roca (Planete, INRIA), B. Sayadi and R. Imad (Alcatel Lucent), we proposed and analyzed a novel technique capable of providing both an unequal erasure protection service and an object bundle protection service.

Unequal Erasure Protection: When a data flow contains information of different priority levels, it is natural to try to offer an unequal protection where the high priority data benefits from a higher protection than the rest of data. In this work we focused on the “erasure channel”, for instance the Internet where the UDP/IP datagram integrity is guaranteed by the physical layer FCS (or CRC) and the UDP checksum. In this context UEP refers to an Unequal Erasure Protection (rather than Error) and the FEC code being used is one of the various Application-Layer Forward Erasure Correction (AL-FEC) codes that have been designed and standardized in the past years, like Reed-Solomon, one of the LDPC variants, or Raptor(Q) codes. Offering an unequal protection in this context can be achieved by one of the following three general approaches: by using dedicated UEP-aware FEC codes, by using a dedicated UEP-aware packetization scheme, or by using an UEP-aware signaling scheme. In this work we ignored the first approach as we wanted to reuse existing AL-FEC codes. Instead we focused on and compared the last two approaches and more precisely the well known Priority Encoding Transmission (PET) scheme that belongs to the UEP-aware packetization category and a Generalized Object Encoding (GOE) scheme, we proposed [53], that belongs to the UEP-aware signaling category. Through a careful modeling of both proposals [55], whose accuracy has been confirmed by simulations, we have demonstrated that the protection performance (i.e. erasure resiliency and average decoding delay) of both approaches are equivalent, not only asymptotically but also in finite length conditions. In fact the key differences between these approaches become apparent when applying them in practical systems. Such metrics as the simplicity of the solution, the number of packets processed, the maximum memory requirements, the number of FEC encoding and decodings, as well as the system of linear equations complexity (number of variables) are in favor of the GOE approach.
**Object Bundle Protection:** We considered the use of PET, more precisely an extension called Universal Object Delivery (UOD), and GOE in situations where one needs to send a bundle of small object (e.g. files). If both solutions can address this need, we showed that once again the GOE scheme is highly recommendable for practical realizations. This is mostly due to the lack of flexibility of the PET/UOD approach. For instance the limited size of a packet creates an upper bound to the number of objects that can be considered together (e.g. UOD limits this number to 255), the symbol size has a coarse granularity (e.g. UOD requires symbols to be multiple of 4 bytes when used with RaptorQ codes) which can create rounding problems with certain sets of objects (i.e. the actual packet size may be significantly shorter than the target, and/or the actual code rate significantly different than its target). GOE has no such constraints. In particular GOE offers the possibility to adjust the packet interleaving to the use-case and channel erasure features. One can easily trade robustness in front of long erasure bursts for very short decoding delays of high priority objects and low memory requirements, which can be a key asset in case of small, lightweight terminals or timely delivery services. This feature may be sufficiently important to justify by itself the use of a GOE FEC Scheme [55].

### 6.3.3. Distributed compressed sensing

**Participants:** Aline Roumy, Velotiaray Toto-Zarasoa.

This work has been performed in collaboration with E. Magli and G. Coluccia (Politecnico di Torino) in the framework of the FP7 IST NOE NEWCOM++ (Jan. 2008 - Apr. 2011). A new lossy compression scheme for distributed and sparse sources under a low complexity encoding constraint has been proposed in [26]. This problem naturally arises in wireless sensor networks. Indeed, nodes of a sensor network may acquire temperature readings over time. The temperature may vary slowly over time, and hence consecutive readings have similar values. However, they also have inter-sensor correlation, as the sensors may be in the same room, in which the temperature is rather uniform. The question hence arises of how to exploit intra- and inter-sensor correlations without communication between the sensors and with a low complexity acquisition process in order to save energy consumption at the sensor. Therefore, we consider continuous, correlated, distributed and sparse (in some domain) sources and perform lossy universal compression under a low encoding complexity constraint.

In order to meet the low complexity encoding constraint, the encoding stage is performed by a lossy distributed compressed sensing (CS). More precisely, the proposed architecture is based on the joint use of CS to capture memory of a signal, and DSC to take advantage of inter-sensor correlations. First, we showed that the resilience of CS to quantization error also holds in the distributed setup. Moreover, the optimal number of measurements can be chosen as the one guaranteeing (close-to-)perfect reconstruction. In addition, using joint decoding, dequantization and reconstruction techniques allows to boost performance even further. The joint use of CS and DSC allows to save 1.18 bit per source sample for the same PSNR quality w.r.t. the non-distributed CS scheme. Compared to the DSC scheme (without CS), we observe a gain increasing with the rate for the same PSNR quality. All these results makes the proposed scheme an attractive choice for environments such as sensor networks, in which the devices performing acquisition and processing are severely constrained in terms of energy and computations.

### 6.3.4. Super-resolution as a communication tool

**Participants:** Marco Bevilacqua, Christine Guillemot, Raul Martinez-Noriega, Aline Roumy.

In 2011, we started a new collaboration in the framework of the Joint INRIA/Alcatel Lucent lab. In this work, carried out with M.-L. Alberi (Alcatel Lucent), we proposed a novel technique capable of producing a high-resolution (HR) image from a single low-resolution (LR) image. This method that belongs to the class of single-image super-resolution (SR), offers the promise of overcoming the inherent limitations of the video acquisition and transmission systems. More precisely, one can think of sending a low resolution video to adapt to the complexity constraint of the encoder and/or the bandwidth limitation of the network, while the decoder reconstructs a high-resolution video.
As a first step toward the more ambitious goal of compressing video through SR, we proposed a novel method for single-image super-resolution based on a neighbor embedding technique. Each low-resolution input patch is approximated by a linear combination of nearest neighbors taken from a dictionary. This dictionary stores low-resolution and corresponding high-resolution patches taken from natural images and is thus used to infer the HR details of the super-resolved image. The entire neighbor embedding procedure is carried out in a feature space. Features which are either the gradient values of the pixels or the mean-subtracted luminance values are extracted from the LR input patches, and from the LR and HR patches stored in the dictionary. The algorithm thus searches for the $K$ nearest neighbors of the feature vector of the LR input patch and then computes the weights for approximating the input feature vector. The so-obtained weights are finally used to compute a linear combination of the corresponding HR patches, which yields the super-resolved image. The use of a positive constraint for computing the weights of the linear approximation is shown to have a more stable behavior than the use of sum-to-one constraint and lead to significantly higher PSNR values for the super-resolved images.
6. New Results

6.1. Advanced algorithms of data analysis, description

6.1.1. Advanced description techniques

6.1.1.1. Image joint description and compression

Participant: Ewa Kijak.

This is a joint work with the TEMICS project-team (J. Zepeda and C. Guillemot).

In the context of ANR project ICOS-HD ended at december 2010, in collaboration with Christine Guillemot from TEMICS, we investigated sparse representations methods for local image description. We have developed methods for learning dictionaries to be used for sparse signal representations. These methods lead to dictionaries which have been called Iteration-Tuned Dictionaries (ITDs), Basic ITD (BITD), Tree-Structured ITD (TSITD) and Iteration-Tuned and Aligned Dictionaries (ITAD). All three proposed ITD schemes (BITD, TSITD and ITAD) have been shown to outperform the state-of-the-art learned dictionaries in terms of PSNR versus sparsity. The performance of these dictionaries has also been assessed for both compression and de-noising applications. ITAD in particular has been used to produce a new image codec that outperforms JPEG2000 for a fixed image class and leads in 2011 to two new publications [49], [20].

6.1.1.2. Bag-of-colors

Participant: Hervé Jégou.

This is joint work with Christian Wengert (Kooaba) and Matthijs Douze (INRIA LEAR and SED project-teams).

This work investigates [48] the use of color information when used within a state-of-the-art large scale image search system. We introduce a simple color signature generation procedure, used either to produce global or local descriptors. As a global descriptor, it outperforms several state-of-the-art color description methods, in particular the bag-of-words method based on color SIFT. As a local descriptor, our signature is used jointly with SIFT descriptors (no color) to provide complementary information.

6.1.1.3. Aggregating local image descriptors into compact codes

Participant: Hervé Jégou.

This is joint work with Matthijs Douze (INRIA LEAR and SED project-teams), Patrick Pérez (Technicolor), Florent Perronnin (Xérox Research Center Europe) and Cordelia Schmid (INRIA LEAR).

This work [19] addresses the problem of large-scale image search and consolidates and extends results from a previous work [78]. Different ways of aggregating local image descriptors into a vector are compared, and the Fisher vector is shown to achieves better performance than the reference bag-of-visual words approach for any given vector dimension. We then jointly optimize dimensionality reduction and indexing in order to obtain a precise vector comparison as well as a compact representation. The evaluation shows that the image representation can be reduced to a few dozen bytes. Searching a 100 million image dataset takes about 250 ms on one processor core.

6.1.2. Browsing multimedia databases

Participant: Laurent Amsaleg.

This is a joint work with Björn Pór Jónsson and Grímur Tómasson from the School of Computer Science, Reykjavik University, Iceland.
Since the introduction of personal computers, personal collections of digital media have been growing ever larger. It is therefore increasingly important to provide effective browsing tools for such collections. We have proposed a multi-dimensional model for media browsing, called ObjectCube, based on the multi-dimensional model commonly used in OLAP applications. We implemented a prototype of a media browser based on the ObjectCube model. We then ran evaluations of its performance using three different underlying data stores and photo collections of up to one million photos.

6.1.3. Advanced data analysis techniques

6.1.3.1. Factorial analysis and output display for text and textual streams mining

Participant: Annie Morin.

Textual data can be easily transformed in frequency tables and any method working on contingency tables can be used to process them. Besides, with the important amount of available textual data, we need to find convenient ways to process the data and to get invaluable information. It appears that the use of factorial correspondence analysis allows us to get most of the information included in the data. We start exploring temporal changes in textual data and mainly focus on the visualization of results: we try to detect the topics if they have not already been identified and to study the evolution of the previous vocabulary inside a topic through time. In fact, as with economical datasets, we try to find seasonal components and cycling components in the documents and to characterize these components.

6.1.3.2. Intensive use of SVM for text and image mining

Participants: François Poulet, Thanh Nghi Doan.

Support Vector Machines (SVM) and kernel methods are known to provide accurate models but the learning task usually needs a quadratic program, so this task for very large datasets requires a large memory capacity and a long time. We have developed new algorithms. The first versions of the algorithms were based on a CPU distributed software program, then we have used GP-GPU (General Purpose GPU) versions to significantly improve the algorithm speed (130 times faster than the CPU one, 2500 times faster than libSVM, SVMPerf or CB-SVM). We have extended the least squares SVM algorithm (LS-SVM) to adapt the algorithm to datasets having a very large number of dimensions and have applied boosting to LS-SVM for datasets having simultaneously a very large number of vectors and dimensions on standard computers. In image classification, the usual frameworks involve three steps: feature extraction, building codebook by feature quantization and training the classifier with a standard classification algorithm (eg. SVM). However, task complexity becomes very large when applying this approach on large scale datasets like the ImageNet dataset containing more than 14 million images and 21,000 classes. The complexity is both about the time needed to perform each task and the memory and disk usage (eg. 11TB are needed to store the SIFT descriptors computed on the full datasets). Efficient algorithms must be used into these three steps: - obviously, the descriptors computed for one image are independant of the other image ones, so they can be computed in a parallel way, - the quantization step usually uses a k-means algorithm, we have developed different versions of parallel k-means algorithms to use on GPU or a cluster of CPUs, - for the learning task, we have developed a parallel version of LibSVM. The first results on the ten largest classes of ImageNet dataset are promising [55], we have developed a fast and efficient framework for large scale image classification.

6.1.4. Security of media

6.1.4.1. Security of content based image retrieval

Participants: Thanh Toan Do, Ewa Kijak, Laurent Amsaleg, Teddy Furon.

Over the years, the level of maturity reached by content-based retrieval systems (CBRSs) has significantly increased. CBRSs have so far been used in very friendly settings where cultural enrichments are paramount. CBRSs are also used in quite different settings where the control, the surveillance and the filtering of multimedia information are central, such as for copyright enforcement systems. While an abundant literature assesses that today’s CBRSs are robust against general-purpose attacks, we address in this work the security of content-based retrieval systems. Because of our expertise, we focus on security of content-based image retrieval, where images are described by SIFT descriptors and indexed by NV-Tree. We proved in one
preliminary study that a real system fails to match a specifically attacked image and its quasi-copy, breaking its otherwise excellent copyright protection performances. After proposing specific attacks that aim to disturb the descriptor detection stage by both prevent some key-points of being detected and create new ones [75], [74], we pursue the work by considering attacks dedicated to the description computation stage.

6.1.4.2. Estimation of the false alarm probability in watermarking and fingerprinting

Participant: Teddy Furon.

A key issue in watermarking and fingerprinting applications is to satisfy the requirement on the probability of false detection or false accusation. Assume commercial contents are encrypted and watermarked and that future consumer electronics storage devices have a watermark detector. These devices refuse to record a watermarked content since it is copyrighted material. The probability of false alarm is the probability that the detector considers an original piece of content (which has not been watermarked) as protected. The movie that a user shot during his holidays could be rejected by his storage device. This absolutely non-user-friendly behavior really scares consumer electronics manufacturers.

In fingerprinting, users’ identifiers are embedded in purchased contents. When this content is found in an illegal place (e.g. a P2P network), the copyright holders decode the hidden message, find an identifier, and thus they can trace the traitor, i.e. the customer who has illegally broadcast his copy. However, the task is not that simple because dishonest users might collude. For security reason, anti-collusion codes have to be employed. Yet, these solutions have a non-zero probability of error (defined as the probability of accusing an innocent). This probability should be, of course, extremely low, but it is also a very sensitive parameter: anti-collusion codes get longer (in terms of the number of bits to be hidden in content) as the probability of error decreases. Fingerprint designers have to strike a trade-off, which is hard to conceive when only rough estimation of the probability of error is known. The major issue for fingerprinting algorithms is the fact that embedding large sequences implies also assessing reliability on a huge amount of data, which may be practically unachievable without using rare event analysis.

In collaboration with the team-projects ASPI and ALEA, we developed a novel strategy for simulating rare events and an associated Monte Carlo estimation of tail probabilities. Our method uses a system of interacting particles and exploits a Feynman-Kac representation of that system to analyze their fluctuations. Our precise analysis of the variance of a standard multilevel splitting algorithm reveals an opportunity for improvement. This leads to a novel method that relies on adaptive levels and produces, in the limit of an idealized version of the algorithm, estimates with optimal variance. Some numerical results show performance close to the idealized version of our technique for these practical applications. This work has been published in the journal Statistics and computing [13]. Algorithms for estimating extreme probabilities and quantiles are implemented as a Matlab package.

6.1.4.3. New decoders for fingerprinting

Participant: Teddy Furon.

So far, the accusation process of a Tardos fingerprinting code is based on single decoders which compute a score per user. Users with the highest score or whose scores is above a threshold are then deemed guilty. In the past years, we have contributed to this approach bringing two improvements; the ‘learn and match’ strategy aims at estimating the collusion process and using the matched score function; a rare event analysis translates this score into a more meaningful probability of being guilty. A fast implementation computes the scores of one million of users within 0.2 second on a regular laptop. Therefore, contrary to common belief, although a single decoder is exhaustive with a linear complexity in $O(n)$, it is not slow.

This fast implementation allows us to propose iterative decoders. A first idea is that conditioning by the identities of some colluders bring more discrimination power to the score function. The first iteration is thus a single decoder, users we are extremely confident to accuse are enrolled as side information. The next iteration computes new scores for the remaining users etc. A second idea is that information theory proves that a joint decoder computing scores for pairs, triplets, or in general $t$-tuples is more powerful than single decoders working with scores for single users. However, nobody did try them for large scale setups since the number of $t$-tuples is in $O(n^t)$. We propose in a first iteration to use a single decoder, to prune out users who are
definitely innocents (because their scores are low) and keeping \(O(\sqrt{n})\) individual suspects. The second iteration is a joint decoding working on pairs of users etc. Iteratively, we prune out enough users such that it is manageable to run a joint decoder on bigger \(t\)-tuples. This work has been done under a collaboration of TEMICS, and published in a series of conference papers [37], [38], [36]. A journal version has been submitted to IEEE Trans. on Information Forensics and Security. A Tardos code software suite (generation of code, collusion attacks, accusation algorithms) is available as a C package.

6.1.4.4. Protocols for fingerprinting
**Participant:** Teddy Furon.

A key assumption of the fingerprinting schemes developed so far is that the colluders may know their own codewords but they ignore the codeword of any other innocent user. Otherwise, the collusion can very easily forge a pirated content framing an innocent user because it contains a sequence close enough to his/her codeword. This puts a lot of pressure on the versioning mechanism which creates the personal copy of the content in accordance to a codeword. For instance, suppose that the versioning is done in the user’s setup box, the unique codeword being loaded into this device at the manufacture. If the code matrix ends up in the hands of an untrustworthy employee, then the whole fingerprinting system is pulled down. This is one argument of the motivation for designing cryptographic protocols for the construction, the versioning and the accusation. We have proposed a new asymmetric fingerprinting protocol dedicated to the state-of-the-art Tardos codes. We believe that this is the first such protocol, and that it is practically efficient. The construction of the fingerprints and their embedding within pieces of content is based on oblivious transfer and do not need a trusted third party. Note, however, that during the accusation stage, a trusted third party, like a Judge, is necessary like in any asymmetric fingerprinting scheme we are aware of. This work has been done in collaboration with the team-project TEMICS, Lab-STICC Telecom Bretagne and University College London, and presented at Information Hiding [22]. Ana Charpentier defended her PhD. thesis in October 2011 [72].

6.1.4.5. Reconstructing an image from its local descriptors
**Participant:** Hervé Jégou.

We show [47] that an image can be approximately reconstructed based on the output of a black-box local description software such as those classically used for image indexing. Our approach consists first in using an off-the-shelf image database to find patches which are visually similar to each region of interest of the unknown input image, according to associated local descriptors. These patches are then warped into input image domain according to interest region geometry and seamlessly stitched together. Final completion of still missing texture-free regions is obtained by smooth interpolation. As demonstrated in our experiments, visually meaningful reconstructions are obtained just based on image local descriptors like SIFT, provided the geometry of regions of interest is known. The reconstruction allows most often the clear interpretation of the semantic image content. As a result, this work raises critical issues of privacy and rights when local descriptors of photos or videos are given away for indexing and search purpose.

6.2. Multi-dimensional indexing and clustering
6.2.1. Improved NV-tree
**Participant:** Laurent Amsaleg.

This is a joint work with Björn Pór Jónsson from the School of Computer Science, Reykjavik University, Iceland and with Herwig Lejsek, Videntifier Technologies, Iceland.

We have further improved the NV-Tree (Nearest Vector Tree) indexing techniques. It addresses the specific, yet important, problem of efficiently and effectively finding the approximate \(k\)-nearest neighbors within a collection of a few billion high-dimensional data points. The NV-Tree is a very compact index, as only six bytes are kept in the index for each high-dimensional descriptor. It thus scales extremely well when indexing large collections of high-dimensional descriptors. The NV-Tree efficiently produces results of good quality, even at such a large scale that the indices cannot be kept entirely in main memory any more. We have demonstrated this with extensive experiments using a collection of 2.5 billion SIFT (Scale Invariant Feature Transform) descriptors. Additional experiments involving more than 30 billion SIFT descriptors show results are still of a good quality and that disks are handled as efficiently as they can be.
6.2.2. Indexation of time series

Participants: Laurent Amsaleg, Romain Tavenard.

Dynamic Time Warping (DTW) is the most popular approach for evaluating the similarity of time series, but its computation is costly. Therefore, simple functions lower bounding DTW distances have been designed, accelerating searches by quickly pruning sequences that could not possibly be best matches. The tighter the bounds, the more they prune and the better the performance. Designing new functions that are even tighter is difficult because their computation is likely to become complex, canceling the benefits of their pruning. It is possible, however, to design simple functions with a higher pruning power by relaxing the no false dismissal assumption, resulting in approximate lower bound functions. We have discovered how very popular approaches accelerating DTW such as LB_Keogh and LB_PAA can be made more efficient via approximations. The accuracy of approximations can be tuned, ranging from no false dismissal to potential losses when aggressively set for great response time savings. At very large scale, indexing time series is mandatory. These approximate lower bound functions can be used with iSAX. Furthermore, we have also observed that a k-means-based quantization step for iSAX gives significant performance gains.

6.2.3. Improved image indexing with asymmetric Hamming embedding

Participants: Patrick Gros, Mihir Jain, Hervé Jégou.

We have proposed [28] an improved asymmetric Hamming Embedding scheme for large scale image search based on local descriptors. The comparison of two descriptors relies on a vector-to-binary code comparison, which limits the quantization error associated with the query compared with the original Hamming Embedding method. The approach is used in combination with an inverted file structure that offers high efficiency, comparable to that of a regular bag-of-features retrieval systems, and consistently improves the search quality over the symmetric version on the two datasets used for the evaluation.

6.2.4. Compression techniques for nearest neighbor search

Participants: Laurent Amsaleg, Teddy Furon, Hervé Jégou, Romain Tavenard.

Part of this work on this topic was done in cooperation with Matthijs Douze and Cordelia Schmid (INRIA/LEAR).

6.2.4.1. Re-ranking with source coding

An extension of our previous work on source coding techniques for high-dimensional indexing has been proposed [29]. The goal is to index a large set of vectors, as large as 1 billion vectors, with limited CPU and memory usage. Based on the product quantization-based indexing technique [18], we show that it is interesting to add an additional level of processing to refine the estimated distances. It consists in quantizing the difference vector between a point and the corresponding centroid. When combined with an inverted file, this gives three levels of quantization. Experiments performed on SIFT and GIST image descriptors show excellent search accuracy outperforming three state-of-the-art approaches. Compared with the original work [18], the proposed re-ranking technique is shown to obtain better trade-off with respect to memory, efficiency and search quality.

6.2.4.2. Anti-sparse coding for approximate nearest neighbor search

Following recent works on Hamming Embedding techniques, we propose [67] a binarization method that aim at addressing the problem of nearest neighbor search for the Euclidean metric by mapping the original vectors into binary vectors ones, which are compact in memory, and for which the distance computation is more efficient.

Our method is based on the recent concept of anti-sparse coding, which exhibits here excellent performance for approximate nearest neighbor search. Unlike other binarization schemes, this framework allows, up to a scaling factor, the explicit reconstruction from the binary representation of the original vector. We also show that random projections which are used in Locality Sensitive Hashing algorithms, are significantly outperformed by regular frames for both synthetic and real data if the number of bits exceeds the vector dimensionality, i.e., when high precision is required.
6.2.5. Architecture-aware indexing techniques for solid state disks

**Participants:** Laurent Amsaleg, Gylfi Gudmundsson.

This is a joint work with Björn Pór Jónsson from the School of Computer Science, Reykjavik University, Iceland.

The scale of multimedia data collections is expanding at a very fast rate. In order to cope with this growth, the high-dimensional indexing methods used for content-based multimedia retrieval must adapt gracefully to secondary storage. Recent progress in storage technology, however, means that algorithm designers must now cope with a spectrum of secondary storage solutions, ranging from traditional magnetic hard drives to state-of-the-art solid state disks. We have analyzed the impact of storage technology on a simple, prototypical high-dimensional indexing method for large scale query processing. We found that while the algorithm implementation deeply impacts the performance of the indexing method, the setup of the underlying storage technology is equally important.

6.3. New techniques for linguistic information acquisition and use

6.3.1. NLP for document description

6.3.1.1. Semantic annotation of multimedia documents based on textual data

**Participants:** Ali Reza Ebadat, Vincent Claveau, Pascale Sébillot, Ewa Kijak.

This work is done in the framework of the Quaero project (see below).

On this subject, TEXMEX is implied in three tasks of the Quaero project.

The first task concerns the extraction of terminology from document. The objective of this work is to study the development and the adaptation of methods to automate the acquisition and the structuring of terminologies. In this context, in 2011, we have undergone a new evaluation of terminology extraction systems. Here again, our system, relying on TermoStat (see previous reports) ranked first for the tracks in which we participated. We have also continued our work the use of morphology for biomedical terminologies. This approach relies on the decomposition of terms into morphemes and the translation of these morphemes into japanese (kanji) sub-words. The kanji characters thus offer a semantic way to access the semantics of the morpheme and allow us to detect semantic relations between them. We have tested this approach on more languages and have proved its relevance for information retrieval problems.

The second task aims at extracting semantic and ontological relations from documents. Indeed, detecting semantic and ontological relations in texts is a key to describe a domain and thus manipulate cleverly documents. In 2011, we developed a new relation extraction system based on k-nearest-neighbors and language modeling. It has been tested in the framework of the Quaero evaluation campaign and ranked first or second for all tracks. We have also developed a clustering technique for named entities. It relies on new representation schemes called bag-of-vectors (or bag-of-bags-of-features), which perform better than the classical bag-of-word approach.

The last task directly deals with the semantic annotation of multimedia documents based on textual data, for, very often, many textual or language-related data can be found in multimedia documents or come along such documents. For example, a TV-broadcast, contains speech that can transcribed, Electronic Program Guide and standard program guide information, closed captions, associated websites, etc. All these sources offers a way to exploit complementary information that can be used to semantically annotate multimedia document. During this year, we finished the development of a football multimedia corpus. It contains the video of several matches, the speech transcripts, associated textual data from specialized websites. All these media have been manually annotated in terms of events, named entities, specialized relations (fouls, replacements, etc) and other relevant information. This corpus will be distributed under LGPL-LR license.

6.3.1.2. Text recognition in videos

**Participants:** Khaoula Elagouni, Pascale Sébillot.
This work is done in the context of a joint TEXMEX/Orange Ph.D. thesis supported by a CIFRE grant with Orange Labs.

We aim at helping multimedia content understanding by obtaining benefit from textual clues embedded in digital video data. In 2011, we proposed an Optical Character Recognition-based method to recognize natural scene texts in images, avoiding the conventional character segmentation step. The text image is scanned with multi-scale windows and a robust recognition model is applied on each window, relying on a neural classification approach, to identify non valid characters and recognize valid ones. A graph model is used to represent spatial constraint between recognition results, and to determine the best sequence of characters. Some linguistic knowledge is also incorporated in the graph to remove errors due to recognition confusions. The method was evaluated on the ICDAR 2003 database of scene text images and outperforms state-of-the-art approaches. This work will be presented at DAS2012.

6.3.1.3. DEFT evaluation campaign participation

Participants: Vincent Claveau, Christian Raymond.

Christian Raymond and Vincent Claveau participated to DEFT (http://deft2011.limsi.fr/). Two tasks were proposed: the first one was called ”the diachronic variation task” whose objective was to identify the writing year of some OCR newspapers from 1801 to 1944. The second one was a abstract/article pairing task. Their approaches based on boosting and k-nearest neighbors was ranked first on the difficult diachronic task.

6.3.2. Oral and textual information retrieval

6.3.2.1. Graded-inclusion-based Information retrieval systems

Participants: Vincent Claveau, Laurent Ughetto.

Our work on this topic is done in close collaboration with Olivier Pivert from the PILGRIM project-team of IRISA Lannion.

Databases (DB) querying mechanisms, and more particularly the division of relations was at the origin of the Boolean model for Information Retrieval Systems (IRSs). This model has rapidly shown its limitations and is no more used in Information retrieval (IR). Among the reasons, the Boolean approach do not allow to represent and use the relative importance of terms indexing the documents or representing the queries. However, this notion of importance can be captured by the division of fuzzy relations. This division, modeled by fuzzy implications, corresponds to graded inclusions. Theoretical work conducted by the PILGRIM project-team have shown the interest of this operator in IR.

Our first work was to investigate the use of graded inclusions to model the information retrieval process. In this framework, documents and queries are represented by fuzzy sets, which are paired with operations like fuzzy implications and T-norms. Through different experiments, we have shown that only some among the wide range of fuzzy operations are relevant for information retrieval. When appropriate settings are chosen, it is possible to mimic classical systems, thus yielding results rivaling those of state-of-the-art systems. These positive results have validated the proposed approach, while negative ones have given some insights on the properties needed by such a model.

More recently, the links between our fuzzy model and other classical IR models have been studied. It has been shown that our fuzzy implication-based model can be interpreted as several classical models: an Extended Boolean Model, a Logical Model, a Vector Space Model or a Language Model in IR.

6.3.2.2. Information retrieval in TV streams using automatic speech recognition

Participants: Guillaume Gravier, Patrick Gros, Julien Fayolle, Fabienne Moreau, Christian Raymond.
Automatic speech recognition outputs are by nature incomplete and uncertain, so much that lexical indexes of speech are not sufficient to overcome the errors due to out-of-vocabulary words and to most of the named entities, consisting in important semantic information from the discourse. Using if necessary a phonetic index is a solution to retrieve partially the mis-recognized words but at the price of a lower precision because the phonetic representation is also noisy. We proposed this year (still to be submitted) an indexation method which jointly model lexical and phonetic levels with finite-state transducers, offering the possibility to take a lexical path or a phonetic path between two synchronization nodes. The edges are weighted by a vector of features (edition scores, confidence measures, durations) that will be used in a supervised manner to estimate the reliability of the returned result at the search step. The experiments have shown the complementary of lexical-phonetic representations and their contribution for a task of spoken utterance retrieval using named entity queries.

6.4. New processing tools for audiovisual documents

6.4.1. TV stream structuring

6.4.1.1. Repetition detection-based TV structuring

**Participants:** Vincent Claveau, Guillaume Gravier, Patrick Gros, Emmanuelle Martienne, Abir Ncibi.

We work on the issue of structuring large TV streams. More precisely, we focus on the problem of labeling the segments of a stream according to their types (e.g., programs, commercial breaks, sponsoring, etc). Contrary to existing techniques, we wanted to take into account the sequential aspect of the data, and thus we used Conditional Random Fields (CRF), a classifier which has proved useful to handle sequential data in other domains like computational linguistics or computational biology. During this year, we proved the relevance of CRF in the framework of TV segments labeling. We conducted different experiments, either on manually or automatically segmented streams, with different label granularities, and demonstrated that this approach rivals existing ones. The use of this model for semi-supervised and unsupervised learning are under study.

6.4.2. Program structuring

6.4.2.1. Audiovisual models for event detection in videos

**Participants:** Guillaume Gravier, Patrick Gros, Cédric Penet.

*This work was performed in close collaboration with Technicolor as external partner.*

Following our work on the detection of audio concepts related to violence in movie soundtracks [58], we developed a system for the detection of violent scenes in movies, combining multimodal features. We investigated multimodal fusion strategies and temporal integration exploiting Bayesian networks as a joint distribution model. Several strategies for learning the structure of the Bayesian networks were compared, resulting in a complete system for violence detection. The system was evaluated on the Violent Scenes Detection task of the MediaEval 2011 international evaluation [42] that we co-organized with Technicolor and the University of Geneva [62]. A fair amount of time was dedicated this year to the organization of the evaluation campaign which includes defining the task and metrics, supervising the annotation, recruiting participants, analyzing the results and organizing the corresponding workshop session.

6.4.2.2. Unsupervised multimedia content mining

**Participants:** Guillaume Gravier, Anh Phuong Ta.

*This work on audio content discovery was partially carried out in collaboration with Armando Muscariello and Frédéric Bimbot from the Metiss project-team.*

As an alternative to supervised approaches for multimedia content analysis, where predefined concepts are searched for in the data, we investigate content discovery approaches where knowledge emerge from the data. Following this general philosophy, we pursued work on motif discovery in audio and video content.
Audio motif discovery is the task of finding out, without any prior knowledge, all pieces of signals that repeat, eventually allowing variability. In 2011, we extended our recent work on seeded discovery to near duplicate detection and spoken document retrieval from examples. First, we proposed algorithmic speed ups for the discovery of near duplicate motifs (low variability) in large (several days long) audio streams, exploiting subsampling strategies [39]. Second, we investigated the use of previously proposed efficient pattern matching techniques to deal with motif variability in speech data [40] in a different setting, that of spoken document retrieval from an audio example. We demonstrated the potential of model-free approaches for efficient spoken document retrieval on a variety of data sets, in particular in the framework of the Spoken Web Search task of the MediaEval 2011 international evaluation [41].

Video structure is often enforced through editing rules which result in a set of shots defining an event that repeats throughout the video with a high visual and audio similarity. Typical such shots are anchor persons and close-up on guests in talk-shows. We recently proposed an unsupervised multimodal approach to discover such events exploiting audio and visual consistency between two sets of independent nested clusters, one for each modality [21]. In 2011, we extended the approach in two directions. First, we improved the selection of consistent audio and visual clusters and the unsupervised selection of positive and negative examples exploiting redundancy between nested clusters. Second, we extended the method to discover several audio-visualy consistent events rather than a single one in our previous work, thus enabling the use of unsupervised mining as a pre-processing step for video structure analysis.

6.4.2.3. Topic segmentation with vectorization and morpho-mathematics

Participant: Vincent Claveau.

Our work on this topic is done in close collaboration with Sébastien Lefèvre from the SEASIDE project-team of IRISA Vannes.

Segmenting a program into topics is an important step for fine-grained structuring of TV streams. Based on our work on vectorization (see previous reports), we have developed a new segmentation technique using speech transcripts. Making an analogy with image segmentation, we have adapted the watershed transform to handle these textual data and more precisely the distances computed by vectorization between possible segments.

This method has been tested on different TV collections (news, reports) as well as more usual text collection used for segmentation evaluation. In every cases, our technique has outperformed any state-of-the-art approaches.

6.4.3. Using speech to describe and structure video

Participants: Camille Guinaudeau, Guillaume Gravier, Ludivine Kuznik, Bogdan Ludusan, Pascale Sébillot.

Speech can be used to structure and organize large collections of spoken documents (videos, audio streams, etc) based on semantics. This is typically achieved by first transforming speech into text using automatic speech recognition (ASR), before applying natural language processing (NLP) techniques on the transcripts. Our research focuses firstly on the adaptation of NLP methods designed for regular texts to account for the specific aspects of automatic transcripts. In particular, we investigate a deeper integration between ASR and NLP, i.e., between the transcription phase and the semantic analysis phase.

In 2011, we mostly focused on robust transcription, hierarchical topic segmentation and collection structuring. On the one hand, we investigated the use of broad phonetic landmarks and syllable prominence to improve large vocabulary speech recognition by guiding the Viterbi search process. Several mechanisms to incorporate landmarks into the search space were studied. Significant improvements were observed on radio broadcast news data in the French language. On the other hand, we pursued our work on unsupervised topic adaptation, focusing on the automatic selection of out-of-vocabulary words combining phonetic and morpho-syntactic criteria.
Linear topic segmentation has been widely studied for textual data and recently adapted to spoken contents. However, most documents exhibit a hierarchy of topics which cannot be recovered using linear segmentation. We investigated hierarchical topic segmentation of TV programs exploiting the spoken material. Recursively applying linear segmentation methods is one solution but fails at the lowest levels of the hierarchy when small segments are targeted, in particular when transcription errors jeopardize lexical cohesion. We proposed new probabilistic measures of the lexical cohesion to emphasize the contribution of words that appears only locally, thus attenuating the impact of words which contributed to the segments at an upper level of the hierarchy [11].

Finally, we initiated work in collaboration with INA on structuring a large collection of news reports. The idea is to automatically create links and threads between reports in several months of broadcast news shows, based either on the documentary records of the shows and/or on the automatic transcripts. As preliminary step towards this goal, we investigated distances between documentary records in an information retrieval setting so as to construct a nearest neighbor graph. The next step consists in exploiting graph clustering methods.

Our research in speech for TV content structuring was illustrated through the Texmix demonstration (see Section 5.2) which exploits most of our achievements in the field, including transcription, topic segmentation and collection structuring.
6. New Results

6.1. 3D object and scene modeling, analysis, and retrieval

6.1.1. Quantitative image analysis for archeology

Participants: Bryan Russell, Jean Ponce, Josef Sivic, Helene Dessales [ENS Archeology laboratory].

Accurate indexing and alignment of images is an important problem in computer vision. A successful system would allow a user to retrieve images with similar content to a query image, along with any information associated with the image. Prior work has mostly focused on techniques to index and match photographs depicting particular instances of objects or scenes (e.g. famous landmarks, commercial product labels, etc.). This has allowed progress on tasks, such as the recovery of a 3D reconstruction of the depicted scene.

However, there are many types of images that cannot be accurately aligned. For instance, for many locations there are drawings and paintings made by artists that depict the scene. Matching and aligning photographs, paintings, and drawings is extremely difficult due to various distortions that can arise. Examples include perspective and caricature distortions, along with errors that arise due to the difficulty of drawing a scene by hand.

In this project, we seek to index and align a database of images, paintings, and drawings. The focus of our work is the Championnet house in the Roman ruins at Pompeii, Italy. Given an alignment of the images, paintings, and drawings, we wish to explore tasks that are of interest to archaeologists and curators who wish to study and preserve the site. Example applications include: (i) digitally restoring paintings on walls where the paintings have disappeared over time due to erosion, (ii) geometrically reasoning about the site over time through the drawings, (iii) indexing and searching patterns that exist throughout the site.
Figure 2. Final alignment between the paintings and 3D model. For each example, left: painting; middle: 3D model contours projected onto painting; right: synthesized viewpoint from 3D model using recovered camera parameters. For the examples in (a-c), note how the final alignment is close to the painting. Our system handles paintings that depict the 3D structure of the scene over time and span different artistic styles and mediums (e.g. water colors, cross-hatching, copies of originals on engravings). Notice how the site changes over time, with significant structural changes (e.g. the wall murals decay over time, the columns change). Example failure cases are shown in (d,e).
Recently, we have addressed the problem of automatically aligning historical architectural paintings with 3D models obtained using multi-view stereo technology from modern photographs. This is a challenging task because of the variations in appearance, geometry, color and texture due to environmental changes over time, the nonphotorealistic nature of architectural paintings, and differences in the viewpoints used by the painters and photographers. Our alignment procedure consists of two novel aspects: (i) we combine the gist descriptor with the view-synthesis/retrieval of Irschara et al. to obtain a coarse alignment of the painting to the 3D model, and (ii) we have developed an ICP-like viewpoint refinement procedure, where 3D surface orientation discontinuities (folds and creases) and view-dependent occlusion boundaries are rendered from the automatically obtained and noisy 3D model in a view-dependent manner and matched to gPB contours extracted from the paintings. We demonstrate the alignment of XIXth Century architectural watercolors of the Casa di Championnet in Pompeii with a 3D model constructed from modern photographs using the PMVS public-domain multi-view stereo software. Figure 1 shows some of the captured photographs and snapshots of the 3D reconstruction of the site. Notice that the 3D reconstruction captures much detail of the walls and structures. Example painting to 3D model alignments are shown in figure 2.

This work resulted in a workshop publication [16].

6.1.2. Visual localization by linear combination of image descriptors

Participants: Josef Sivic, Akihiko Torii [Tokyo Institute of Technology], Tomas Pajdla [CTU in Prague].

In this work, we seek to predict the GPS location of a query image given a database of images localized on a map with known GPS locations. The contributions of this work are three-fold: (1) we formulate the image-based localization problem as a regression on an image graph with images as nodes and edges connecting close-by images; (2) we design a novel image matching procedure, which computes similarity between the query and pairs of database images using edges of the graph and considering linear combinations of their feature vectors. This improves generalization to unseen viewpoints and illumination conditions, while reducing the database size; (3) we demonstrate that the query location can be predicted by interpolating locations of matched images in the graph without the costly estimation of multi-view geometry. We demonstrate benefits of the proposed image matching scheme on the standard Oxford building benchmark, and show localization results on a database of 8,999 panoramic Google Street View images of Pittsburgh.

This work resulted in a publication [18].

6.2. Category-level object and scene recognition

6.2.1. Task-Driven Dictionary Learning

Participants: Julien Mairal, Jean Ponce, Francis Bach [INRIA SIERRA].

Modeling data with linear combinations of a few elements from a learned dictionary has been the focus of much recent research in machine learning, neuroscience and signal processing. For signals such as natural images that admit such sparse representations, it is now well established that these models are well suited to restoration tasks. In this context, learning the dictionary amounts to solving a large-scale matrix factorization problem, which can be done efficiently with classical optimization tools. The same approach has also been used for learning features from data for other purposes, e.g., image classification, but tuning the dictionary in a supervised way for these tasks has proven to be more difficult. In this paper, we present a general formulation for supervised dictionary learning adapted to a wide variety of tasks, and present an efficient algorithm for solving the corresponding optimization problem. Experiments on handwritten digit classification, digital art identification, nonlinear inverse image problems, and compressed sensing demonstrate that our approach is effective in large-scale settings, and is well suited to supervised and semi-supervised classification, as well as regression tasks for data that admit sparse representations.

This work has resulted in a publication [4].

6.2.2. Ask the locals: multi-way local pooling for image recognition

Participants: Y-Lan Boureau, Jean Ponce, Nicolas Le Roux [INRIA SIERRA], Francis Bach [INRIA SIERRA], Yann LeCun [New York University].
Invariant representations in object recognition systems are generally obtained by pooling feature vectors over spatially local neighborhoods. But pooling is not local in the feature vector space, so that widely dissimilar features may be pooled together if they are in nearby locations. Recent approaches rely on sophisticated encoding methods and more specialized codebooks (or dictionaries), e.g., learned on subsets of descriptors which are close in feature space, to circumvent this problem. In this work, we argue that a common trait found in much recent work in image recognition or retrieval is that it leverages locality in feature space on top of purely spatial locality. We propose to apply this idea in its simplest form to an object recognition system based on the spatial pyramid framework, to increase the performance of small dictionaries with very little added engineering. State-of-the-art results on several object recognition benchmarks show the promise of this approach.

This work has resulted in a publication [7].

6.2.3. A Graph-matching Kernel for Object Categorization

**Participants:** Olivier Duchenne, Armand Joulin, Jean Ponce.

This paper addresses the problem of category-level image classification. The underlying image model is a graph whose nodes correspond to a dense set of regions, and edges reflect the underlying grid structure of the image and act as springs to guarantee the geometric consistency of nearby regions during matching. A fast approximate algorithm for matching the graphs associated with two images is presented. This algorithm is used to construct a kernel appropriate for SVM-based image classification, and experiments with the Caltech 101, Caltech 256, and Scenes datasets demonstrate performance that matches or exceeds the state of the art for methods using a single type of features.

This work has resulted in an ICCV 2011 publication [9] (oral presentation).

6.2.4. A Tensor-Based Algorithm for High-Order Graph Matching

**Participants:** Olivier Duchenne, Jean Ponce, Francis Bach [INRIA SIERRA], Inso Kweon [KAIST, Korea].

This paper addresses the problem of establishing correspondences between two sets of visual features using higher-order constraints instead of the unary or pairwise ones used in classical methods. Concretely, the corresponding hypergraph matching problem is formulated as the maximization of a multilinear objective function over all permutations of the features. This function is defined by a tensor representing the affinity between feature tuples. It is maximized using a generalization of spectral techniques where a relaxed problem is first solved by a multi-dimensional power method, and the solution is then projected onto the closest assignment matrix. The proposed approach has been implemented, and it is compared to state-of-the-art algorithms on both synthetic and real data.

This work has resulted in an PAMI publication [2].

6.2.5. Clusterpath: an algorithm for clustering using convex fusion penalties

**Participants:** Armand Joulin, Toby Hocking [INRIA SIERRA], Francis Bach [INRIA SIERRA], Jean-Philippe Vert [Mines ParisTech].

We present a new clustering algorithm by proposing a convex relaxation of hierarchical clustering, which results in a family of objective functions with a natural geometric interpretation. We give efficient algorithms for calculating the continuous regularization path of solutions, and discuss relative advantages of the parameters. Our method experimentally gives state-of-the-art results similar to spectral clustering for non-convex clusters, and has the added benefit of learning a tree structure from the data.

This work has resulted in an publication [10].

6.2.6. An MRF model for binarization of natural scene text

**Participants:** Kartteek Alahari, Anand Mishra [IIT India], C.V. Jawahar [IIT India].
Scene text recognition has gained significant attention from the computer vision community in recent years. Recognizing text in the wild is a challenging problem, even more so than the recognition of scanned documents. In this work, we focus on the problem of cropped word recognition. We present a framework that exploits both bottom-up and top-down cues. The bottom-up cues are derived from individual character detections from the image. We build a Conditional Random Field model on these detections to jointly model the strength of the detections and the interactions between them. We impose top-down cues obtained from a lexicon-based prior, i.e. language statistics, on the model. The optimal word represented by the text image is obtained by minimizing the energy function corresponding to the random field model.

We show very significant improvements in accuracies on two challenging public datasets, namely Street View Text (over 15%) and ICDAR 2003 (over 10%).

This work has resulted in an publication [12].

6.2.7. Strongly-supervised deformable part model for object detection

**Participants:** Hossein Azizpour [KTH Stockholm], Ivan Laptev, Stefan Carlsson [KTH Stockholm].

Deformable part models achieve state-of-the-art performance for object detection while relying on the greedy initialization during training. The goal of this paper is to investigate limitations of such initialization and to improve the model for the case when part locations are known at the training time. To this end, we deploy part-level supervision and demonstrate improved detection results when learning models with manually-initialized part locations. We further explore the benefits of the strong supervision and learn model structure by minimizing the variance among adjacent model parts. Our method can simultaneously handle samples with and without part-level annotation making benefit even from a fraction of fully-annotated training samples. Experimental results are reported for the detection of six animal classes in PASCAL VOC 2007 and 2010 datasets. We demonstrate significantly improved performance of our model compared to the state-of-the-art LSVM object detector and poselet detector. Example learnt models are shown in figure 3.

This work has resulted in a submission to CVPR 2012.

6.2.8. Exploiting Photographic Style for Category-Level Image Classification by Generalizing the Spatial Pyramid

**Participant:** Jan van Gemert [University of Amsterdam].

This paper investigates the use of photographic style for category-level image classification. Specifically, we exploit the assumption that images within a category share a similar style defined by attributes such as colorfulness, lighting, depth of field, viewpoint and saliency. For these style attributes we create correspondences across images by a generalized spatial pyramid matching scheme. Where the spatial pyramid groups features spatially, we allow more general feature grouping and in this paper we focus on grouping images on photographic style. We evaluate our approach in an object classification task and investigate style differences between professional and amateur photographs. We show that a generalized pyramid with style-based attributes improves performance on the professional Corel and amateur Pascal VOC 2009 image datasets.

This work has resulted in a publication [20].

6.2.9. Generalized Fast Approximate Energy Minimization via Graph Cuts: Alpha-Expansion Beta-Shrink Moves

**Participants:** Karteek Alahari, Mark Schmidt [INRIA SIERRA].

We present alpha-expansion beta-shrink moves, a simple generalization of the widely-used alpha beta-swap and alpha-expansion algorithms for approximate energy minimization. We show that in a certain sense, these moves dominate both alpha beta-swap and alpha-expansion moves, but unlike previous generalizations the new moves require no additional assumptions and are still solvable in polynomial-time. We show promising experimental results with the new moves, which we believe could be used in any context where alpha-expansions are currently employed.
6.3. Image restoration, manipulation and enhancement

6.3.1. Non-uniform Deblurring for Shaken Images

Participants: Oliver Whyte, Josef Sivic, Andrew Zisserman, Jean Ponce.

We argue that blur resulting from camera shake is mostly due to the 3D rotation of the camera, causing a blur that can be significantly non-uniform across the image. However, most current deblurring methods model the observed image as a convolution of a sharp image with a uniform blur kernel. We propose a new parametrized geometric model of the blurring process in terms of the rotational velocity of the camera during exposure. We apply this model in the context of two different algorithms for camera shake removal: the first uses a single blurry image (blind deblurring), while the second uses both a blurry image and a sharp but noisy image of the same scene. We show that our approach makes it possible to model and remove a wider class of blurs than previous approaches, and demonstrate its effectiveness with experiments on real images.

The project resulted in a publication [5].

6.3.2. Deblurring shaken and partially saturated images

Participants: Oliver Whyte, Josef Sivic, Andrew Zisserman.

We address the problem of deblurring images degraded by camera shake blur and saturated or over-exposed pixels. Saturated pixels are a problem for existing non-blind deblurring algorithms because they violate the assumption that the image formation process is linear, and often cause significant artifacts in deblurred outputs. We propose a forward model that includes sensor saturation, and use it to derive a deblurring algorithm properly treating saturated pixels. By using this forward model and reasoning about the causes of artifacts.
in the deblurred results, we obtain significantly better results than existing deblurring algorithms. Further we propose an efficient approximation of the forward model leading to a significant speed-up. Example result is shown in figure 4.

The project resulted in a publication [19].

**Figure 4.** Deblurring saturated images. Note that the ringing around saturated regions, visible in columns (b) and (c) is removed by our method (d), without causing any loss in visual quality elsewhere.

### 6.3.3. Dictionary Learning for Deblurring and Digital Zoom

**Participants:** Florent Couizinie, Julien Mairal, Jean Ponce, Francis Bach [INRIA SIERRA].

This work proposes a novel approach to image deblurring and digital zooming using sparse local models of image appearance. These models, where small image patches are represented as linear combinations of a few elements drawn from some large set (dictionary) of candidates, have proven well adapted to several image restoration tasks. A key to their success has been to learn dictionaries adapted to the reconstruction of small image patches. In contrast, recent works have proposed instead to learn dictionaries which are not only adapted to data reconstruction, but also tuned for a specific task. We introduce here such an approach to deblurring and digital zoom, using pairs of blurry/sharp (or low-/high-resolution) images for training, as well as an effective stochastic gradient algorithm for solving the corresponding optimization task. Although this learning problem is not convex, once the dictionaries have been learned, the sharp/high-resolution image can be recovered via convex optimization at test time. Experiments with synthetic and real data demonstrate the effectiveness of the proposed approach, leading to state-of-the-art performance for non-blind image deblurring and digital zoom.

This work has resulted in a publication [1].

### 6.3.4. Sparse Image Representation with Epitomes

**Participants:** Louise Benoit, Julien Mairal, Jean Ponce, Francis Bach [INRIA SIERRA].

Sparse coding, which is the decomposition of a vector using only a few basis elements, is widely used in machine learning and image processing. The basis set, also called dictionary, is learned to adapt to specific data. This approach has proven to be very effective in many image processing tasks. Traditionally, the dictionary is an unstructured "flat" set of atoms. In this work, we study structured dictionaries which are
obtained from an epitome, or a set of epitomes. The epitome is itself a small image, and the atoms are all the patches of a chosen size inside this image. This considerably reduces the number of parameters to learn and provides sparse image decompositions with shift invariance properties. We propose a new formulation and an algorithm for learning the structured dictionaries associated with epitomes, and illustrate their use in image denoising tasks.

This work has resulted in a CVPR’11 publication [6].

6.3.5. Proximal Methods for Hierarchical Sparse Coding

Participants: Julien Mairal, Rodolphe Jenatton [INRIA SIERRA], Guillaume Obozinski [INRIA SIERRA], Francis Bach [INRIA SIERRA].

Sparse coding consists in representing signals as sparse linear combinations of atoms selected from a dictionary. We consider an extension of this framework where the atoms are further assumed to be embedded in a tree. This is achieved using a recently introduced tree-structured sparse regularization norm, which has proven useful in several applications. This norm leads to regularized problems that are difficult to optimize, and in this paper, we propose efficient algorithms for solving them. More precisely, we show that the proximal operator associated with this norm is computable exactly via a dual approach that can be viewed as the composition of elementary proximal operators. Our procedure has a complexity linear, or close to linear, in the number of atoms, and allows the use of accelerated gradient techniques to solve the tree-structured sparse approximation problem at the same computational cost as traditional ones using the l1-norm. Our method is efficient and scales gracefully to millions of variables, which we illustrate in two types of applications: first, we consider fixed hierarchical dictionaries of wavelets to denoise natural images. Then, we apply our optimization tools in the context of dictionary learning, where learned dictionary elements naturally self-organize in a prespecified arborescent structure, leading to better performance in reconstruction of natural image patches. When applied to text documents, our method learns hierarchies of topics, thus providing a competitive alternative to probabilistic topic models.

This work has resulted in a publication [3].

6.4. Human activity capture and classification

6.4.1. Track to the future: Spatio-temporal video segmentation with long-range motion cues

Participants: Jose Lezama, Karteek Alahari, Ivan Laptev, Josef Sivic.

Video provides rich visual cues such as motion and appearance but also much less explored long-range temporal interactions among objects. We aim to capture such interactions and to construct powerful intermediate-level video representation for subsequent recognition. Motivated by this goal, we seek to obtain spatio-temporal oversegmentation of the video into regions that respect object boundaries and, at the same time, associate object pixels over many video frames. The contributions of this paper are twofold. First, we develop an efficient spatio-temporal video segmentation algorithm, that naturally incorporates long-range motion cues from the past and future frames in the form of clusters of point tracks with coherent motion. Second, we devise a new track clustering cost-function that includes occlusion reasoning, in the form of depth ordering constraints, as well as motion similarity along the tracks. We evaluate the proposed approach on a challenging set of video sequences of office scenes from feature length movies.

This work resulted in a publication [11].

6.4.2. Density-aware person detection and tracking in crowds

Participants: Mikel Rodriguez, Ivan Laptev, Josef Sivic, Jean-Yves Audibert [INRIA SIERRA].
We address the problem of person detection and tracking in crowded video scenes. While the detection of individual objects has been improved significantly over the recent years, crowd scenes remain particularly challenging for the detection and tracking tasks due to heavy occlusions, high person densities and significant variation in people’s appearance. To address these challenges, we propose to leverage information on the global structure of the scene and to resolve all detections jointly. In particular, we explore constraints imposed by the crowd density and formulate person detection as the optimization of a joint energy function combining crowd density estimation and the localization of individual people. We demonstrate how the optimization of such an energy function significantly improves person detection and tracking in crowds. We validate our approach on a challenging video dataset of crowded scenes. The proposed approach is illustrated in figure 5.

This work has resulted in a publication [14].

![Figure 5. Individual head detections provided by state-of-the-art object detector (Felzenswalb et al. 2009) (bottom-left; green: true positives; red: false positives) are improved significantly by our method (bottom-right; yellow: new true positives) using the crowd density estimate (top-right) obtained from the original frame (top-left).]

### 6.4.3. Data-driven Crowd Analysis in Videos

**Participants:** Mikel Rodriguez, Josef Sivic, Ivan Laptev, Jean-Yves Audibert [INRIA SIERRA].

In this work, we present a new crowd analysis algorithm powered by behavior priors that are learned on a large database of crowd videos gathered from the Internet. The algorithm works by first learning a set of crowd behavior priors off-line. During testing, crowd patches are matched to the database and behavior priors are transferred. We adhere to the insight that despite the fact that the entire space of possible crowd behaviors is infinite, the space of distinguishable crowd motion patterns may not be all that large. For many individuals in a crowd, we are able to find analogous crowd patches in our database which contain similar patterns of behavior that can effectively act as priors to constrain the difficult task of tracking an individual in a crowd. Our algorithm is data-driven and, unlike some crowd characterization methods, does not require us to have seen the test video beforehand. It performs like state-of-the-art methods for tracking people having common crowd behaviors and outperforms the methods when the tracked individual behaves in an unusual way.

This work has resulted in a publication [15].

### 6.4.4. Learning person-object interactions for action recognition in still images

**Participants:** Vincent Delaitre, Josef Sivic, Ivan Laptev.
In this work, we investigate a discriminatively trained model of person-object interactions for recognizing common human actions in still images. We build on the locally order-less spatial pyramid bag-of-features model, which was shown to perform extremely well on a range of object, scene and human action recognition tasks. We introduce three principal contributions. First, we replace the standard quantized local HOG/SIFT features with stronger discriminatively trained body part and object detectors. Second, we introduce new person-object interaction features based on spatial co-occurrences of individual body parts and objects. Third, we address the combinatorial problem of a large number of possible interaction pairs and propose a discriminative selection procedure using a linear support vector machine (SVM) with a sparsity inducing regularizer. Learning of action-specific body part and object interactions bypasses the difficult problem of estimating the complete human body pose configuration. Benefits of the proposed model are shown on human action recognition in consumer photographs, outperforming the strong bag-of-features baseline. The proposed model is illustrated in figure 6.

This work has resulted in a publication [8].

![Figure 6. Representing person-object interactions by pairs of body part (cyan) and object (blue) detectors. To get a strong interaction response, the pair of detectors (here visualized at positions pi and pj) must fire in a particular relative 3D scale-space displacement (given by the vector v) with a scale-space displacement uncertainty (deformation cost) given by diagonal 3x3 covariance matrix C (the spatial part of C is visualized as a yellow dotted ellipse). Our image representation is defined by the max-pooling of interaction responses over the whole image, solved efficiently by the distance transform.](image)

### 6.4.5. People Watching: Human Actions as a Cue for Single View Geometry

**Participants:** David Fouhey [CMU], Vincent Delaitre, Abhinav Gupta [CMU], Ivan Laptev, Alexei Efros [CMU], Josef Sivic.

We present an approach which exploits the coupling between human actions and scene geometry. We investigate the use of human pose as a cue for single-view 3D scene understanding. Our method builds upon recent advances in still-image action recognition and pose estimation, to extract functional and geometric constraints about the scene from people detections. These constraints are then used to improve state-of-the-art single-view 3D scene understanding approaches. The proposed method is validated on a collection of single-viewpoint time-lapse image sequences as well as a dataset of still images of indoor scenes. We demonstrate that observing people performing different actions can significantly improve estimates of scene geometry and 3D layout. The main idea of this work is illustrated in figure 7.
This work is in submission to CVPR 2012.

Figure 7. What can human actions tell us about the 3D structure of the scene? Quite a lot, actually. Consider the two person detections and their estimated pose in (a). They were detected in a time-lapse sequence of one of the three scenes (b-d). Can you guess which one? Most people can easily see that it is (b). Even though this is only a static image, the actions and the pose of the disembodied figures reveal a lot about the geometric structure of the scene. The pose of the left figure reveals a horizontal surface right under its pelvis, which ends abruptly at the knees. The right figure’s pose reveals a ground plane under its feet as well as a likely horizontal surface near the hand location. In both cases we observe a strong physical and functional coupling that exists between people and the 3D geometry of the scene. Our aim in this work is to exploit this coupling.

6.4.6. Joint pose estimation and action recognition in image graphs

Participants: K. Raja [INRIA Rennes], Ivan Laptev, Patrick Perez [Technicolor], L. Osei [INRIA Rennes].

Human analysis in images and video is a hard problem due to the large variation in human pose, clothing, camera view-points, lighting and other factors. While the explicit modeling of this variability is difficult, the huge amount of available person images motivates for the implicit, datadriven approach to human analysis. In this work we aim to explore this approach using the large amount of images spanning a subspace of human appearance. We model this subspace by connecting images into a graph and propagating information through such a graph using a discriminatively trained graphical model. We particularly address the problems of human pose estimation and action recognition and demonstrate how image graphs help solving these problems jointly. We report results on still images with human actions from the KTH dataset.

This work has resulted in a publication [13].

6.5. Creation of the SIERRA project-team

6.5.1. From WILLOW alone to WILLOW and SIERRA

The WILLOW team officially started in the Spring of 2007. From the start, it was clear that machine learning was a key ingredient to new breakthroughs, and our activities have steadily grown in this area. In three short years, WILLOW has grown into a mature group of about 30 people, and it divides its activities between computer vision, machine learning, and the cross-pollination of the two fields, with video as one of the core research areas. We have been very successful, with many publications in all the major international conferences and leading journals in both areas, but we are a large group with very diverse interests, ranging from camera
geometry to statistics, and from image retrieval to bioinformatics applications of structured sparse coding. With the creation of the SIERRA project-team, the core machine learning activities of WILLOW have been transferred to the new group.

The two teams continue collaborating with each other (they remain co-located at the INRIA site in central Paris), but have a sharper focus on their respective computer vision and machine learning activities.

6.5.2. SIERRA

The SIERRA project-team was created by the INRIA on January 1st 2011 and is headed by Francis Bach, who received in 2009 a Jr. ERC grant.