IN PARTNERSHIP WITH:
CNRS

Institut polytechnique de Grenoble

Université Joseph Fourier (Grenoble 1)

Activity Report 2011

Project-Team PERCEPTION

Interpretation and Modelling of Images and Videos

IN COLLABORATION WITH: Laboratoire Jean Kuntzmann (LJK)

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Vision, Perception and Multimedia Understanding

Table of contents

Project-Team PERCEPTION

Keywords: Computer Vision, Auditory Analysis, Audio-Visual Fusion, Robot Vision, Robot Hearing

1. Members

Research Scientists
Radu Horaud [Team leader, Research Director (DR), HdR]
Peter Sturm [Research Director (DR), member until 1 May 2011, HdR]

Technical Staff
Michel Amat [Development Engineer]

PhD Students
Xavier Alameda-Pineda [MESR grant]
Visesh Chari [INRIA grant]
Antoine Deleforge [MESR grant]
Mauricio Diaz [Alban-EU grant]
Kaustubh Kulkarni [INRIA grant]
Régis Perrier [INRIA grant]
Jordi Sanchez-Riera [INRIA grant]
Avinash Sharma [INRIA grant]

Post-Doctoral Fellows
Miles Hansard [Research Engineer]
Jan Cech [Research Engineer]
Simone Gasparini [Research Engineer]

2. Overall Objectives

2.1. Introduction

The overall objective of the PERCEPTION group is to develop theories, models, methods, and systems allowing computers to see, to hear and to understand what they see and what they hear. A major difference between classical computer systems and computer perception systems is that while the former are guided by sets of mathematical and logical rules, the latter are governed by the laws of nature. It turns out that formalizing interactions between an artificial system and the physical world is a tremendously difficult task.

A first objective is to be able to gather images and videos with one or several cameras, to calibrate them, and to extract 2D and 3D geometric information. This is difficult because the cameras receive light stimuli and these stimuli are affected by the complexity of the objects (shape, surface, color, texture, material) composing the real world. The interpretation of light in terms of geometry is also affected by the fact that the three dimensional world projects onto two dimensional images and this projection alters the Euclidean nature of the observed scene.

A second objective is to gather sounds using several microphones, to localize and separate sounds composed of several auditory sources, and to analyse and interpret them. Sound localization, separation and recognition is difficult, especially in the presence of noise, reverberant rooms, competing sources, overlap of speech and prosody, etc.

A third objective is to analyse articulated and moving objects. Solutions for finding the motion fields associated with deformable and articulated objects (such as humans) remain to be found. It is necessary to introduce prior models that encapsulate physical and mechanical features as well as shape, aspect, and behaviour. The ambition is to describe complex motion as “events” at both the physical level and at the semantic level.

A fourth objective is to combine vision and hearing in order to disambiguate situations when a single modality is not sufficient. In particular we are interested in defining the notion of *audio-visual object* (AVO) and to deeply understand the mechanisms allowing to associate visual data with auditory data.

A fifth objective is to build vision systems, hearing systems, and audio-visual systems able to interact with their environment, possibly in real-time. In particular we are interested in building the concept of an audio-visual robot that communicates with people in the most natural way.

2.2. Highlights

2.2.1. The European project Humavips – Humanoids with Auditory and Visual Abilities in Populated Spaces

HUMAVIPS (http://humavips.inrialpes.fr) is a 36 months FP7 STREP project coordinated by Radu Horaud and which started in 2010. The project addresses multimodal perception and cognitive issues associated with the computational development of a social robot. The ambition is to endow humanoid robots with audiovisual (AV) abilities: exploration, recognition, and interaction, such that they exhibit adequate behavior when dealing with a group of people. Proposed research and technological developments will emphasize the role played by multimodal perception within principled models of human-robot interaction and of humanoid behavior.

2.2.2. Collaboration with SAMSUNG – 3D Capturing and Modeling from Scalable Camera Configurations

In 2010 started a 30 months collaboration with the Samsung Advanced Institute of Technology (SAIT), Seoul, Korea. Whithin this project we develop a methodology able to combine data from several types of visual sensors (2D high-definition color cameras and 3D range cameras) in order to reconstruct, in real-time, an indoor scene without any constraints in terms of background, illumination conditions, etc.

2.2.3. Outstanding paper award at ICMI'11

Our article "Finding Audio-Visual Events in Informal Social Gatherings" [29] received the "Outstanding Paper Award" (best paper) at the IEEE/ACM 13th International Conference on Multimodal Interaction (ICMI), Alicante, Spain, November 2011. The paper is co-authored by members of both PERCEPTION and MISTIS, Xavi Alameda-Pineda, Vasil Khalidov, Radu Horaud and Florence Forbes. The paper addresses the problem of detecting and localizing audio-visual events (such as people) in a complex/cluttered scenario such as a cocktail party. The work is carried out within the collaborative European project HUMAVIPS.

3. Scientific Foundations

3.1. The geometry of multiple images

Computer vision requires models that describe the image creation process. An important part (besides e.g. radiometric effects), concerns the geometrical relations between the scene, cameras and the captured images, commonly subsumed under the term “multi-view geometry”. This describes how a scene is projected onto an image, and how different images of the same scene are related to one another. Many concepts are developed and expressed using the tool of projective geometry. As for numerical estimation, e.g. structure and motion calculations, geometric concepts are expressed algebraically. Geometric relations between different views can for example be represented by so-called matching tensors (fundamental matrix, trifocal tensors, ...). These tools and others allow to devise the theory and algorithms for the general task of computing scene structure and camera motion, and especially how to perform this task using various kinds of geometrical information: matches of geometrical primitives in different images, constraints on the structure of the scene or on the intrinsic characteristics or the motion of cameras, etc.

3.2. The photometry component

In addition to the geometry (of scene and cameras), the way an image looks like depends on many factors, including illumination, and reflectance properties of objects. The reflectance, or “appearance”, is the set of laws and properties which govern the radiance of the surfaces . This last component makes the connections between the others. Often, the “appearance” of objects is modeled in image space, e.g. by fitting statistical models, texture models, deformable appearance models (...) to a set of images, or by simply adopting images as texture maps.

Image-based modelling of 3D shape, appearance, and illumination is based on prior information and measures for the coherence between acquired images (data), and acquired images and those predicted by the estimated model. This may also include the aspect of temporal coherence, which becomes important if scenes with deformable or articulated objects are considered.

Taking into account changes in image appearance of objects is important for many computer vision tasks since they significantly affect the performances of the algorithms. In particular, this is crucial for feature extraction, feature matching/tracking, object tracking, 3D modelling, object recognition etc.

3.3. Shape Acquisition

Recovering shapes from images is a fundamental task in computer vision. Applications are numerous and include, in particular, 3D modeling applications and mixed reality applications where real shapes are mixed with virtual environments. The problem faced here is to recover shape information such as surfaces, point positions, or differential properties from image information. A tremendous research effort has been made in the past to solve this problem and a number of partial solutions had been proposed. However, a fundamental issue still to be addressed is the recovery of full shape information over time sequences. The main difficulties are precision, robustness of computed shapes as well as consistency of these shapes over time. An additional difficulty raised by real-time applications is complexity. Such applications are today feasible but often require powerful computation units such as PC clusters. Thus, significant efforts must also be devoted to switch from traditional single-PC units to modern computation architectures.

3.4. Motion Analysis

The perception of motion is one of the major goals in computer vision with a wide range of promising applications. A prerequisite for motion analysis is motion modelling. Motion models span from rigid motion to complex articulated and/or deformable motion. Deformable objects form an interesting case because the models are closely related to the underlying physical phenomena. In the recent past, robust methods were developed for analysing rigid motion. This can be done either in image space or in 3D space. Image-space analysis is appealing and it requires sophisticated non-linear minimization methods and a probabilistic framework. An intrinsic difficulty with methods based on 2D data is the ambiguity of associating a multiple degree of freedom 3D model with image contours, texture and optical flow. Methods using 3D data are more relevant with respect to our recent research investigations. 3D data are produced using stereo or a multiple-camera setup. These data (surface patches, meshes, voxels, etc.) are matched against an articulated object model (based on cylindrical parts, implicit surfaces, conical parts, and so forth). The matching is carried out within a probabilistic framework (pair-wise registration, unsupervised learning, maximum likelihood with missing data).

Challenging problems are the detection and segmentation of multiple moving objects and of complex articulated objects, such as human-body motion, body-part motion, etc. It is crucial to be able to detect motion cues and to interpret them in terms of moving parts, independently of a prior model. Another difficult problem is to track articulated motion over time and to estimate the motions associated with each individual degree of freedom.

3.5. Multiple-camera acquisition of visual data

Modern computer vision techniques and applications require the deployment of a large number of cameras linked to a powerful multi-PC computing platform. Therefore, such a system must fulfill the following requirements: The cameras must be synchronized up to the millisecond, the bandwidth associated with image transfer (from the sensor to the computer memory) must be large enough to allow the transmission of uncompressed images at video rates, and the computing units must be able to dynamically store the data and/to process them in real-time.

Current camera acquisition systems are all-digital ones. They are based on standard network communication protocols such as the IEEE 1394. Recent systems involve as well depth cameras that produce depth images, i.e. a depth information at each pixel. Popular technologies for this purpose include the Time of Flight Cameras (TOF cam) and structured light cameras, as in the very recent Microsoft's Kinect device.

3.6. Auditory and audio-visual scene analysis

In 2010, PERCEPTION started to investigate a new research topic, namely the analysis of auditory information and the fusion between auditory and visual data. In particular we are interested in analysing the acoustic layout of a scene (how many sound sources are out there and where are they located, what is the semantic content of each auditory signal). For that purpose we use microphones that are mounted onto a human-like head. This allows the extraction of several kinds of auditory cues, either based on the time difference of arrival or based on the fact that the head and the ears modify the spectral properties of the sounds perceived with the left and right microphones. Both the temporal and spectral binaural cues can be used to locate the most proeminent sound sources, and to separate the perceived signal into several sources. This is however an extremely difficult task because of the inherent ambiguity due resemblance of signals, and of the presence of acoustic noise and reverberations. The combination of visual and auditory data allows to solve the localization and separation tasks in a more robust way, provided that the two stimuli are available. One interesting yet unexplored topic is the development of hearing for robots, such as the role of head and body motions in the perception of sounds.

4. Application Domains

4.1. 3D modeling, rendering and interaction

3D modeling and image-based rendering are two technologies that, when combined together, produce extremely realistic visual representations of objects, animals, humans, etc. The employment of advanced computer vision techniques for media applications is a dynamic area that will benefit from scientific findings and developments. There is a huge potential in the spheres of TV and film productions, interactive TV, multimedia database retrieval, and so forth.

Vision research provides solutions for real-time recovery of studio models (3D scene, people and their movements, etc.) in realistic conditions compatible with artistic production (several moving people in changing lighting conditions, partial occlusions). In particular, the recognition of people and their motions will offer a whole new range of possibilities for creating dynamic situations and for immersive/interactive interfaces and platforms in TV productions. These new and not yet available technologies involve integration of action and gesture recognition techniques for new forms of interaction between, for example, a TV moderator and virtual characters and objects, two remote groups of people, real and virtual actors, etc.

4.2. Human motion capture and analysis

We are particularly interested in the capture and analysis of human motion, which consists in recovering the motion parameters of the human body and/or human body parts, such as the hand. In the past researchers have concentrated on recovering constrained motions such as human walking and running. We are interested in recovering unconstrained motion. The problem is difficult because of the large number of degrees of freedom, the small size of some body parts, the ambiguity of some motions, the self-occlusions, etc. Human motion capture methods have a wide range of applications: human monitoring, surveillance, gesture/action analysis, motion recognition, computer animation, etc.

4.3. Human-robot interaction (HRI)

Robots have gradually moved from factory floors to populated spaces. There is a need for novel methodologies and associated technologies enabling robots to deal with complex and unstructured environments and to communicate with people in the most natural way. There are many applications that will benefit from HRI, such as human helpers, entertainment robots, rescue robots, etc.

5. Software

5.1. Mixed camera platform

We started to develop a multiple camera platform composed of both high-definition color cameras and low-resolution depth cameras. This platform combines the advantages of the two camera types. On one side, depth (time-of-flight) cameras provide relatively accurate 3D scene information. On the other side, color cameras provide information allowing for high-quality rendering. The software package developed during the year 2011 contains the calibration of TOF cameras, alignment between TOF and color cameras, and image-based rendering. These software developments are performed in collaboration with the Samsung Advanced Institute of Technology. The multi-camera platform and the basic software modules are products of 4D Views Solutions SAS, a start-up company issued from the PERCEPTION group.

5.2. Audiovisual robot head

We have developed two audiovisual (AV) robot heads: the POPEYE head and the NAO stereo head. Both are equipped with a binocular vision system and four microphones. The software modules comprise stereo matching and reconstruction, sound-source localization and audio-visual fusion. POPEYE has been developed within the European project POP (http://perception.inrialpes.fr/POP in collaboration with the project-team MISTIS and with two other POP partners: the Speech and Hearing group of the University of Sheffield and the Institute for Systems and Robotics of the University of Coimbra. The NAO stereo head is being developed under the European project HUMAVIPS (http://humavips.inrialpes.fr) in collaboration with Aldebaran Robotics (which manufactures the humanoid robot NAO) and with the University of Bielefeld, the Czech Technical Institute, and IDIAP. The software modules that we develop are compatible with both these robot heads.

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 pointcorrespondences 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 highresolution 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.

7. Contracts and Grants with Industry

7.1. Contract with Samsung Electronics

We continued a 12 months collaboration with the Samsung Advanced Institute of Technology (SAIT), Seoul, South Korea. Whithin this project we develop a methodology able to combine data from several types of visual sensors (2D high-definition color cameras and 3D range cameras) in order to reconstruct, in real-time, an indoor scene without any constraints in terms of background, illumination conditions, etc. The final software package was successfully installed in October 2011 at Samsung.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR Project

8.1.1.1. ROM

Program: ANR CONTINT

Project acronym: ROM

Project title: Realtime Onset Matchmoving

Duration: 2008 – 2011

Coordinator: Duran Duboi SA

Other partners: VORTEX (IRI Toulouse, France)

Abstract: This industrial R&D project concerns the generation of special effects for movie or other film productions. In particular, the goal is to provide tools for successful onset matchmoving, that is the estimation of camera trajectories during acquisition, with immediate pre-visualization of special effects superimposed on acquired sequences. Besides this real-time aspect of matchmoving, the project also addresses the problem of preparing a shooting, by analyzing if matchmoving with natural features is possible and if not, by instrumenting the scene with artificial markers in appropriate positions.

8.2. European Initiatives

8.2.1. FP7 Project

8.2.1.1. HUMAVIPS

Title: Humanoids with audiovisual skills in populated spaces

Type: COOPERATION (ICT)

Defi: Cognitive Systems and Robotics

Instrument: Specific Targeted Research Project (STREP)

Duration: February 2010 - January 2013

Coordinator: INRIA (France)

Others partners: CTU Prague (Czech Republic), University of Bielefeld (Germany), IDIAP (Switzerland), Aldebaran Robotics (France)

See also: http://humavips.inrialpes.fr

Abstract: Humanoids expected to collaborate with people should be able to interact with them in the most natural way. This involves significant perceptual, communication, and motor processes, operating in a coordinated fashion. Consider a social gathering scenario where a humanoid is expected to possess certain social skills. It should be able to explore a populated space, to localize people and to determine their status, to decide to join one or two persons, to synthetize appropriate behavior, and to engage in dialog with them. Humans appear to solve these tasks routinely by integrating the often complementary information provided by multi sensory data processing, from low-level 3D object positioning to high-level gesture recognition and dialog handling. Understanding the world from unrestricted sensorial data, recognizing people? intentions and behaving like them are extremely challenging problems. The objective of HUMAVIPS is to endow humanoid robots with audiovisual (AV) abilities: exploration, recognition, and interaction, such that they exhibit adequate behavior when dealing with a group of people. Proposed research and technological developments will emphasize the role played by multimodal perception within principled models of human-robot interaction and of humanoid behavior. An adequate architecture will implement auditory and visual skills onto a fully programmable humanoid robot. An open-source software platform will be developed to foster dissemination and to ensure exploitation beyond the lifetime of the project.

8.2.2. ESA project

8.2.2.1. ITI 3D

Program: ESA ITI (European Space Agency Triangular Initiatives)

Project acronym: ITI 3D

Project title: Multi-View 3D Reconstruction of Asteroids

Duration: 2010 – 2011

Coordinator: EADS Astrium

Abstract: The goal of the project is to implement and validate algorithms for image-based 3D modeling of asteroids. The algorithms combine multi-view stereo and shape-from-shading.

9. Dissemination

9.1. Animation of the scientific community

- Radu Horaud is a member of the following editorial boards:
  - advisory board member of the *International Journal of Robotics Research*,
  - associate editor of the *International Journal of Computer Vision*, and
  - area editor of *Computer Vision and Image Understanding*.
- Peter Sturm is associate editor of the journals IVC, JMIV, JCST, and CVA.
- Peter Sturm was Area Chair of CVPR'11 and program committee member of 4 international workshops.
- Peter Sturm has been Program Chair of the IEEE International Conference on Computer Vision (ICCV'11), Barcelona, Spain.

9.2. Invited talks

R. Horaud gave the following invited talks:

- Invited keynote speaker at the "Global 3D Technology Forum", organized by Korea? 3D Fusion Industry Association, Seoul, Korea, 10-11 October 2011.
- Invited review lecture "3D Shape Representation Using Graph Kernels" at the Third International Conference on Scale-Space and Variational Methods in Computer Vision (SSVM?1), May 2011.
- Invited lecture at the "3D Forum" of the Samsung Advanced Institute of Technology, Seoul, South Korea, April 2011

9.3. Teaching

Master: Visual computing, 30 hours, P. Sturm

Master: Computer Vision, 30 hours, P. Sturm

Master: 3D Computer Vision, 6h, P. Sturm

Doctorate : Data Analysis and Manifold Learning, 24 hours, University of Grenoble, R. Horaud.

9.4. PhD defenses

PhD: Amaël Delaunoy, *Multi-view Shape Reconstruction from Images : Contributions Towards Generic and Practical Solutions using Deformable Meshes*, University of Grenoble, December 2nd, 2011, advised by Peter Sturm and Emmanuel Prados

PhD: Mauricio Diaz, *Estimating Illumination and Photometric Properties using Photo Collections*, University of Grenoble, October 26th, 2011, advised by Peter Sturm

PhD: Régis Perrier, *Estimation de l'attitude d'un satellite à l'aide de caméras pushbroom et de capteurs stellaires*, University of Grenoble, September 27th, 2011, advised by Peter Sturm and Elise Arnaud

10. Bibliography

Major publications by the team in recent years

[1] N. ANDREFF, B. ESPIAU, R. HORAUD. *Visual Servoing from Lines*, in "International Journal of Robotics Research", August 2002, vol. 21, n^{o} 8, p. 679-700, http://perception.inrialpes.fr/Publications/2002/AEH02.

[2] A. BARTOLI, N. DALAL, R. HORAUD. *Motion Panoramas*, in "Computer Animation and Virtual Worlds", 2004, vol. 15, p. 501-517, http://perception.inrialpes.fr/Publications/2004/BDH04.

[3] Y. DUFOURNAUD, C. SCHMID, R. HORAUD. *Image Matching with Scale Adjustment*, in "Computer Vision and Image Understanding", February 2004, vol. 93, n^{o} 2, p. 175-194, http://perception.inrialpes.fr/Publications/2004/DSH04.

[4] J.-S. FRANCO, E. BOYER. *Efficient Polyhedral Modeling from Silhouettes*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", March 2009, vol. 31, n^{o} 3, p. 414–427,, http://perception.inrialpes.fr/Publications/2009/FB09.

[5] P. GARGALLO, E. PRADOS, P. STURM. *Minimizing the Reprojection Error in Surface Reconstruction from Images*, in "Proceedings of the International Conference on Computer Vision, Rio de Janeiro, Brazil", IEEE Computer Society Press, 2007, http://perception.inrialpes.fr/Publications/2007/GPS07.

[6] M. HANSARD, R. HORAUD. *Cyclopean Geometry of Binocular Vision*, in "Journal of the Optical Society of America A", September 2008, vol. 25, n^{o} 9, p. 2357-2369, http://perception.inrialpes.fr/Publications/2008/HH08.

[7] R. HORAUD, G. CSURKA, D. DEMIRDJIAN. *Stereo Calibration from Rigid Motions*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", December 2000, vol. 22, n^{o} 12, p. 1446-1452, http://perception.inrialpes.fr/Publications/2000/HCD00.

[8] R. HORAUD, M. NISKANEN, G. DEWAELE, E. BOYER. *Human Motion Tracking by Registering an Articulated Surface to 3-D Points and Normals*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", January 2009, vol. 31, n^{o} 1, p. 158-164, http://perception.inrialpes.fr/Publications/2009/HNDB09.

[9] D. KNOSSOW, R. RONFARD, R. HORAUD. *Human Motion Tracking with a Kinematic Parameterization of Extremal Contours*, in "International Journal of Computer Vision", September 2008, vol. 79, n^{o} 2, p. 247-269, http://perception.inrialpes.fr/Publications/2008/KRH08.

[10] D. MATEUS, R. HORAUD, D. KNOSSOW, F. CUZZOLIN, E. BOYER. *Articulated Shape Matching Using Laplacian Eigenfunctions and Unsupervised Point Registration*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", 2008, http://perception.inrialpes.fr/Publications/2008/MHKCB08.

[11] S. RAMALINGAM, S. LODHA, P. STURM. *A Generic Structure-from-Motion Framework*, in "Computer Vision and Image Understanding", sep 2006, vol. 103, n^{o} 3, p. 218–228, http://perception.inrialpes.fr/Publications/2006/RLS06.

[12] C. SMINCHISESCU, B. TRIGGS. *Estimating Articulated Human Motion with Covariance Scaled Sampling*, in "International Journal of Robotics Research", 2003, http://perception.inrialpes.fr/Publications/2003/ST03.

[13] J.-P. TARDIF, P. STURM, M. TRUDEAU, S. ROY. *Calibration of Cameras with Radially Symmetric Distortion*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2009, vol. 31, n^{o} 9, p. 1552-1566, http://perception.inrialpes.fr/Publications/2009/TSTR09.

[14] D. WEINLAND, R. RONFARD, E. BOYER. *Free Viewpoint Action Recognition using Motion History Volumes*, in "Computer Vision and Image Understanding", November/December 2006, vol. 104, n^{o} 2-3, p. 249–257, http://perception.inrialpes.fr/Publications/2006/WRB06a.

[15] M. WILCZKOWIAK, P. STURM, E. BOYER. *Using Geometric Constraints Through Parallelepipeds for Calibration and 3D Modelling*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", feb 2005, vol. 27, n^{o} 2, p. 194-207, http://perception.inrialpes.fr/Publications/2005/WSB05.

[16] K.-J. YOON, E. PRADOS, P. STURM. *Joint Estimation of Shape and Reflectance using Multiple Images with Known Illumination Conditions*, in "International Journal of Computer Vision", 2010, vol. 86, n^{o} 2-3, p. 192–210, http://perception.inrialpes.fr/Publications/2010/YPS10.

[17] A. ZAHARESCU, E. BOYER, K. VARANASI, R. HORAUD. *Surface Feature Detection and Description with Applications to Mesh Matching*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", Miami Beach, Florida, June 2009, http://perception.inrialpes.fr/Publications/2009/ZBVH09.

[18] A. ZAHARESCU, R. HORAUD. *Robust Factorization Methods Using A Gaussian/Uniform Mixture Model*, in "International Journal of Computer Vision", March 2009, vol. 81, n^{o} 3, p. 240-258, http://perception.inrialpes.fr/Publications/2009/ZH09.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[19] A. DELAUNOY. *Modélisation 3D à partir d'images: contributions en reconstruction photométrique à l'aide de maillages déformables*, Université de Grenoble, December 2011, http://tel.archives-ouvertes.fr/tel-00654188/en/.

[20] M. DIAZ. *Analyse de l'Illumination et des Proprietes de Reflectance en Utilisant des Collections d'Images*, Universite de Grenoble, October 2011, http://hal.inria.fr/tel-00641467/en.

[21] B. PETIT. *Telepresence, immersion et interactions pour le reconstruction 3D temps-reel*, Universite de Grenoble, February 2011, http://hal.inria.fr/tel-00584001/en.

Articles in International Peer-Reviewed Journal

[22] A. DELAUNOY, E. PRADOS. *Gradient Flows for Optimizing Triangular Mesh-based Surfaces: Applications to 3D Reconstruction Problems dealing with Visibility*, in "International Journal of Computer Vision", 2011, vol. 95, n^{o} 2, p. 100–123 [*DOI :* 10.1007/s11263-010-0408-9], http://hal.inria.fr/inria-00546105/en.

[23] M. HANSARD, R. HORAUD. *A Differential Model of the Complex Cell*, in "Neural Computation", 2011, vol. 23, n^{o} 9, p. 2324-2357 [*DOI :* 10.1162/NECO_a_00163], http://hal.inria.fr/inria-00590266/en.

[24] R. HORAUD, F. FORBES, M. YGUEL, G. DEWAELE, J. ZHANG. *Rigid and Articulated Point Registration with Expectation Conditional Maximization*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", March 2011, vol. 33, n^{o} 3, p. 587–602 [*DOI :* 10.1109/TPAMI.2010.94], http://hal.inria.fr/inria-00590265/en.

[25] V. KHALIDOV, F. FORBES, R. HORAUD. *Conjugate Mixture Models for Clustering Multimodal Data*, in "Neural Computation", February 2011, vol. 23, n^{o} 2, p. 517–557 [*DOI :* 10.1162/NECO_a_00074], http://hal.inria.fr/inria-00590267/en.

[26] L. PUIG, Y. BASTANLAR, P. STURM, J. GUERRERO, J. BARRETO. *Calibration of Central Catadioptric Cameras Using a DLT-Like Approach*, in "International Journal of Computer Vision", May 2011, vol. 93, n^{o} 1, p. 101–114 [*DOI :* 10.1007/s11263-010-0411-1], http://hal.inria.fr/inria-00590268/en.

[27] P. STURM, S. RAMALINGAM, J.-P. TARDIF, S. GASPARINI, J. BARRETO. *Camera Models and Fundamental Concepts Used in Geometric Computer Vision*, in "Foundations and Trends in Computer Graphics and Vision", 2011, vol. 6, n^{o} 1-2, p. 1–183 [*DOI :* 10.1561/0600000023], http://hal.inria.fr/inria-00590269/en.

[28] A. ZAHARESCU, E. BOYER, R. HORAUD. *Topology-Adaptive Mesh Deformation for Surface Evolution, Morphing, and Multi-View Reconstruction*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", April 2011, vol. 33, n^{o} 4, p. 823–837 [*DOI :* 10.1109/TPAMI.2010.116], http://hal.inria.fr/inria-00590271/en.

International Conferences with Proceedings

[29] X. ALAMEDA-PINEDA, V. KHALIDOV, R. HORAUD, F. FORBES. *Finding Audio-Visual Events in Informal Social Gatherings*, in "IEEE/ACM International Conference on Multimodal Interfaces", Alicante, Spain, November 2011, Outstanding Paper Award, http://hal.inria.fr/inria-00623489/en.

[30] J. CECH, R. HORAUD. *Joint Disparity and Optical Flow by Correspondence Growing*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)", Prague, Czech Republic, 2011 [*DOI :* 10.1109/ICASSP.2011.5946548], http://hal.inria.fr/inria-00614408/en.

[31] J. CECH, J. SANCHEZ-RIERA, R. HORAUD. *Scene Flow Estimation by Growing Correspondence Seeds*, in "IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)", Colorado Springs, CO, IEEE Computer Society, 2011, http://hal.inria.fr/inria-00590274/en.

[32] M. DIAZ, P. STURM. *Exploiting Image Collections for Recovering Photometric Properties*, in "International Conference on Computer Analysis of Images and Patterns (CAIP)", Seville, Spain, Springer, 2011, http://hal.inria.fr/inria-00606149/en.

[33] M. DIAZ, P. STURM. *Radiometric Calibration using Photo Collections*, in "IEEE International Conference on Computational Photography (ICCP '11)", Pittsburgh, IEEE Computer Society, 2011, p. 1–8 [*DOI :* 10.1109/ICCPHOT.2011.5753117], http://hal.inria.fr/inria-00590275/en.

[34] M. HANSARD, R. HORAUD, M. AMAT, S. LEE. *Projective Alignment of Range and Parallax Data*, in "IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)", Colorado Springs, CO, IEEE Computer Society Press, 2011, p. 3089–3096, http://hal.inria.fr/inria-00590277/en.

[35] A. HOODA, M. BRONSTEIN, A. BRONSTEIN, R. HORAUD. *Shape Palindromes: Analysis of Intrinsic Symmetries in 2D Articulated Shapes*, in "3rd International Conference on Scale Space and Variational Methods in Computer Vision (SSVM '11)", Ein-Gedi, Israel, Springer-Verlag, 2011, http://hal.inria.fr/inria-00590276/en.

[36] S. RAMALINGAM, S. BOUAZIZ, P. STURM. *Pose estimation using both Points and Lines for Geo-Localization*, in "IEEE International Conference on Robotics and Automation", Shanghai, Chine, IEEE Computer Society, 2011, http://hal.inria.fr/inria-00590279/en.

[37] S. RAMALINGAM, S. BOUAZIZ, P. STURM, P. TORR. *The Light-Path Less Traveled*, in "IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", Colorado Springs, États-Unis, 2011, http://hal.inria.fr/inria-00582846/en.

[38] A. SHARMA, R. HORAUD, J. CECH, E. BOYER. *Topologically-Robust 3D Shape Matching Based on Diffusion Geometry and Seed Growing*, in "IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)", Colorado Springs, CO, IEEE Computer Society Press, 2011, p. 2481–2488, http://hal.inria.fr/inria-00590280/en.

[39] K. VARUN RAJ, P. STURM. *Detection and Avoidance of Semi-Transparent Obstacles using a Collective-Reward Based Approach*, in "IEEE International Conference on Robotics and Automation (ICRA '11)", Shanghai, China, IEEE Computer Society, 2011, http://hal.inria.fr/inria-00590278/en.

National Conferences with Proceeding

[40] L. CALVET, P. GURDJOS, V. CHARVILLAT, S. GASPARINI, P. STURM. *Suivi de camera a partir de marqueurs plans composes de cercles concentriques : paradigme et algorithmes*, in "ORASIS - Congres des jeunes chercheurs en vision par ordinateur", Praz-sur-Arly, France, INRIA Grenoble Rhône-Alpes, 2011, http://hal.inria.fr/inria-00596248/en.

[41] M. DIAZ, P. STURM. *Calibrage Radiometrique en Utilisant des Collections de Photos*, in "ORASIS - Congres des jeunes chercheurs en vision par ordinateur", Praz-sur-Arly, France, INRIA Grenoble Rhône-Alpes, 2011, http://hal.inria.fr/inria-00595363/en.

Scientific Books (or Scientific Book chapters)

[42] A. SHARMA, R. HORAUD, D. MATEUS. *3D Shape Registration Using Spectral Graph Embedding and Probabilistic Matching*, in "Image Processing and Analysing With Graphs: Theory and Practice", O. LEZORAY, L. GRADY (editors), CRC Press, 2011, http://hal.inria.fr/inria-00590273/en.

Research Reports

[43] X. ALAMEDA-PINEDA, J. SANCHEZ-RIERA, V. FRANCH, J. WIENKE, J. CECH, K. KULKARNI, A. DELEFORGE, R. HORAUD. *The Ravel data set*, INRIA, August 2011, n^{o} RR-7709, http://hal.inria.fr/inria-00614483/en.

[44] A. DELEFORGE, R. HORAUD. *Learning the Direction of a Sound Source Using Head Motions and Spectral Features*, INRIA, February 2011, n^{o} RR-7529, http://hal.inria.fr/inria-00564708/en.

[45] E. PRADOS, N. JINDAL, S. SOATTO. *A non-local approach to shape from ambient shading*, INRIA, November 2011, n^{o} RR-7783, http://hal.inria.fr/inria-00637531/en.

[46] M. SAPIENZA, M. HANSARD, R. HORAUD. *Real-time 3D reconstruction and fixation with an active binocular head*, INRIA, July 2011, n^{o} RR-7682, http://hal.inria.fr/inria-00608769/en.

[47] A. ZAHARESCU, E. BOYER, R. HORAUD. *Scale-Space Representation of Scalar Functions on 2D Manifolds*, INRIA, July 2011, n^{o} RR-7672, http://hal.inria.fr/inria-00607252/en.