INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team perception

Interpretation and Modelling of Images and Videos

Grenoble - Rhône-Alpes

Theme : Vision, Perception and Multimedia Understanding

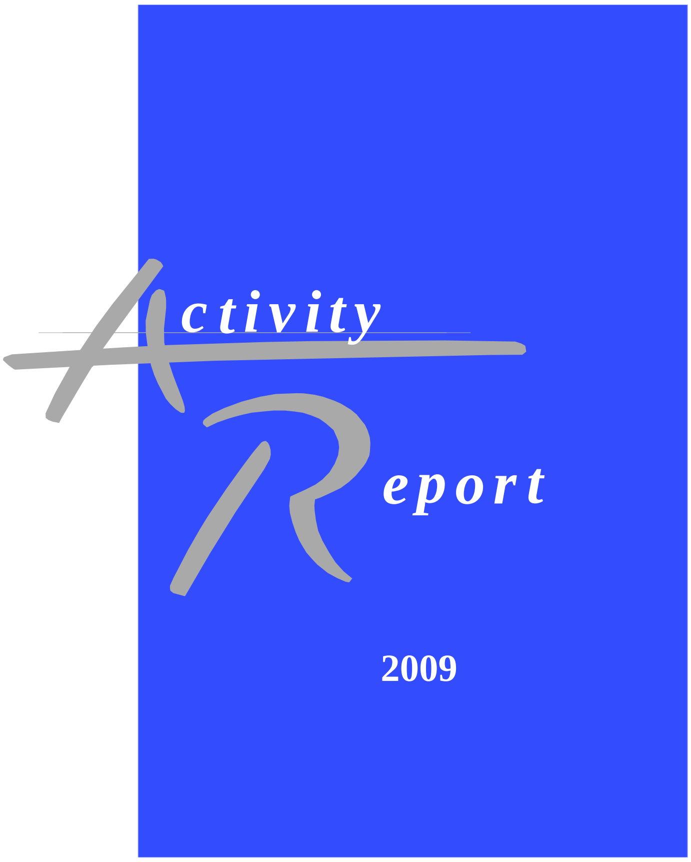

Table of contents

Perception is a common project with INRIA and LJK (Laboratoire Jean Kuntzmann).

1. Team

Research Scientist

Radu Horaud [Team leader, Research Director (DR), HdR]
Emmanuel Prados [Research Associate (CR)]
Peter Sturm [Research Director (DR), HdR]

Faculty Member

Elise Arnaud [Université Joseph Fourier Grenoble]
Edmond Boyer [Université Joseph Fourier Grenoble, HdR]

Technical Staff

Florent Lagaye [Development Engineer]
David Knossow [Development Engineer]

PhD Student

Amaël Delaunoy [INRIA grant]
Mauricio Diaz [Alban-EU grant]
Diana Mateus [Marie-Curie grant]
Ramya Narasimha [INRIA grant]
Régis Perrier [INRIA grant]
Benjamin Petit [INRIA grant]
Kiran Varanasi [INRIA grant]
Visesh Chari [INRIA grant]
Avinash Sharma [INRIA grant]
Jamil Draréni [Co-supervision agreement with Université de Montréal]
Antoine Letouzey [INRIA grant]
Sacha Bernet [Visiting PhD student, Université Haute-Alsace, Mulhouse, France]

Post-Doctoral Fellow

Etienne Von Lavante
Simone Gasparini
Miles Hansard

Visiting Scientist

Hiroshi Kawasaki [Visiting professor, Saitama University, Japan]
Julian Quiroga [Assistant Lecturer, Universidad Javierana Bogota, Colombia]
João Barreto [Associate Professor, University of Coimbra, Portugal]

Administrative Assistant

Anne Pasteur [Secretary (SAR) Inria, until sept 09]
Marie-Eve Morency [Secretary (SAR) Inria, from sept 09]

2. Overall Objectives

2.1. Introduction

The overall objective of the PERCEPTION research team is to develop theories, models, methods, and systems in order to allow computers to see and to understand what they see. A major difference between classical computer systems and computer vision 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 from these images and videos. This is an extremely difficult task 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 analyse articulated and moving objects. The real world is composed of rigid, deformable, and articulated 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 third objective is to describe and interpret images and videos in terms of objects, object categories, and events. In the past it has been shown that it is possible to recognize a single occurrence of an object from a single image. A more ambitious goal is to recognize object classes such as people, cars, trees, chairs, etc., as well as events or *objects evolving in time*. In addition to the usual difficulties that affect images of a single object there is also the additional issue of the variability within a class. The notion of statistical shape must be introduced and hence statistical learning should be used. More generally, learning should play a crucial role and the system must be designed such that it is able to learn from a small training set of samples. Another goal is to investigate how an object recognition system can take advantage from the introduction of non-visual input such as semantic and verbal descriptions. The relationship between images and meaning is a great challenge.

A fourth objective is to build vision systems that encapsulate one or several objectives stated above. Vision systems are built within a specific application. The domains at which vision may contribute are numerous:

- Multi-media technologies and in particular film and TV productions, database retrieval;
- Visual surveillance and monitoring;
- Augmented and mixed reality technologies and in particular entertainment, cultural heritage, telepresence and immersive systems, image-based rendering and image-based animation;
- Embedded systems for television, portable devices, defense, space, etc.

2.2. Highlights

2.2.1. The European project POP – Perception on Purpose

We coordinated the European project POP http://perception.inrialpes.fr/POP/. Together with four other partners we developed a new concept and associated technology that enables a robot to combine data from sound and vision to create multimodal purposeful perception. In collaboration with the EPI MISTIS, PERCEPTION developed a new methodology based on unsupervised statistical machine learning: a conjugate mixture model of multimodal clustering. The associated algorithm is able to cluster data acquired with physically different sensors such as cameras and microphones [55]. This three year European project (2006-2009) was highlighed on the *ICT results* website dedicated to promoting outstanding European research, http://cordis.europa.eu/ictresults/index.cfm?section=news&tpl=article&BrowsingType=Features&ID=90953.

2.2.2. The European Marie-Curie research training network VISIONTRAIN

This RTN Marie-Curie action successfully ended in 2009. It gathered eleven academic partners from eleven European countries and it allowed the training of eleven PhD students and 12 Postdoctoral researchers through collaborative research. Radu Horaud was the network coordinator, http://visiontrain.inrialpes.fr/.

2.2.3. The Virtualization gate presented at SIGGRAPH’09

VGate is a new immersive environment that allows full-body immersion and interaction with virtual worlds. A live demonstration of VGate was presented at Siggraph Emerging Technologies 2009, New Orleans, 3-7 August 2009, http://grimage.inrialpes.fr/vgate/VGate/VGate.html.

2.2.4. Eurographics'09 best third paper award

The best third paper price of Eurographics 2009 was awarded to Elise Arnaud and co-authors Maxime Tournier, Xiaomao Wu, Nicolas Courty and Lionel Reveret for their paper "Motion compression using principal geodesics analysis [19].

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.

Until recently, the vast majority of systems were based on hybrid analog-digital camera systems. Current systems are all-digital ones. They are based on network communication protocols such as the IEEE 1394. Current systems deliver 640×480 grey-level/color images but in the near future 1600×1200 images will be available at 30 frames/second.

Camera synchronization may be performed in several ways. The most common one is to use special-purpose hardware. Since both cameras and computers are linked through a network, it is possible to synchronize them using network protocols, such as NTP (network time protocol).

4. Application Domains

4.1. 3D modeling and rendering

3D modeling from images can be seen as a basic technology, with many uses and applications in various domains. Some applications only require geometric information (measuring, visual servoing, navigation) while more and more rely on more complete models (3D models with texture maps or other models of appearance) that can be rendered in order to produce realistic images. Some of our projects directly address potential applications in virtual studios or “edutainment” (e.g. virtual tours), and many others may benefit from our scientific results and software.

4.2. Mixed and Augmented Reality

Mixed realities consist in merging real and virtual environments. The fundamental issue in this field is the level of interaction that can be reached between real and virtual worlds, typically a person catching and moving a virtual object. This level depends directly on the precision of the real world models that can be obtained and on the rapidity of the modeling process to ensure consistency between both worlds. A challenging task is then to use images taken in real-time from cameras to model the real world without help from intrusive material such as infrared sensors or markers.

Augmented reality systems allow an user to see the real world with computer graphics and computer animation superimposed and composited with it. Applications of the concept of AR basically use virtual objects to help the user to get a better understanding of her/his surroundings. Fundamentally, AR is about augmentation of human visual perception: entertainment, maintenance and repair of complex/dangerous equipment, training, telepresence in remote, space, and hazardous environments, emergency handling, and so forth. In recent years, computer vision techniques have proved their potential for solving key-problems encountered in AR: real-time pose estimation, detection and tracking of rigid objects, etc. However, the vast majority of existing systems use a single camera and the technological challenge consisted in aligning a prestored geometrical model of an object with a monocular image sequence.

4.3. 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 analysis, motion recognition, computer animation, etc.

4.4. Multi-media and interactive applications

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.

4.5. Car driving technologies

In the long term (five to ten years from now) all car manufacturers foresee that cameras with their associated hardware and software will become parts of standard car equipment. Cameras' fields of view will span both outside and inside the car. Computer vision software should be able to have both low-level (alert systems) and high-level (cognitive systems) capabilities. Forthcoming camera-based systems should be able to detect and recognize obstacles in real-time, to assist the driver for manoeuvering the car (through a verbal dialogue), and to monitor the driver's behaviour. For example, the analysis and recognition of the driver's body gestures and head motions will be used as cues for modelling the driver's behaviour and for alerting her or him if necessary.

4.6. Defense technologies

The PERCEPTION project has a long tradition of scientific and technological collaborations with the French defense industry. In the past we collaborated with Aérospatiale SA for 10 years (from 1992 to 2002). During these years we developed several computer vision based techniques for air-to-ground and ground-to-ground missile guidance. In particular we developed methods for enabling 3D reconstruction and pose recovery from cameras on-board of the missile, as well as a method for tracking a target in the presence of large scale changes.

4.7. Satellite imaging

The study of advanced computer technique to analyse satelitte images is a growing area as cameras are widely used for earth observation. More specifically, the PERCEPTION project is developping an expertise in the study of linear puschbroom cameras. These cameras are used in passive remote sensing from space as they provide high resolution images. The pushbroom camera is a linear sensor that takes 1-D images at several time instants. The sensor sweeps out a region of space; stitching together all 1-D images gives a complete 2-D image of the observed scene. Several interesting computer vision issues are of interest, such as the calibration of such cameras, or the estimation of satellite vibrations from these sensors.

5. Software

5.1. Platforms

5.1.1. The Grimage platform

The Grimage platform is an experimental laboratory dedicated to multi-media applications of computer vision. It hosts a multiple-camera system connected to a PC cluster, as well as to a multi-video projection system. This laboratory is shared by several research groups, most proeminently PERCEPTION and MOAIS. In particular, Grimage allows challenging real-time immersive applications based on computer vision and interactions between real and virtual objects, Figure 1.

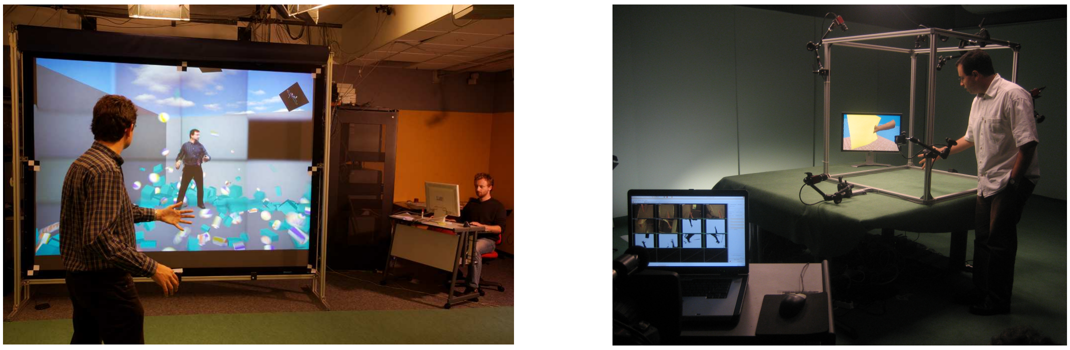

Figure 1. Left: The Grimage platform allows immersive/interactive applications such as this one. The real character is reconstruced in real-time and immersed in a virtual world, such that he/she can interact with virtual objects. Right: The mini-Grimage platform holds on a table top. It uses six cameras connected to six mini-PCs and to a laptop.

5.1.2. The mini-Grimage platform

We also deveoped a miniaturized version of Grimage. Based on the same algorithms and software, this mini-Grimage platform can hold on a desk top and/or can be used for various experiments involving fast and realistic 3-D reconstruction of objects, Figure 1.

5.1.3. Virtualization Gate

Vgate is a new immersive environment that allows full-body immersion and interaction with virtual worlds. It is a joint initiative of computer scientists from computer vision, parallel computing and computer graphics from several research groups at INRIA Grenoble RhÃ´ne-Alpes, and in collaboation with the company 4D View Solutions. The PERCEPTION team is leading this project. Vgate was demonstrated at Siggraph'09. [41]

5.1.4. POPEYE: an audiovisual robotic head

We have developed an audiovisual (AV) robot head that supports software for AV fusion based on binocular vision and binaural audition (see below). The vision module is composed of two digital cameras that form a stereoscopic pair with control of vergence (one rotational degree of freedom per camera). The auditory module is composed of two microphones. The head can perform pan and tilt rotations as well. All the sensors are linked to a PC. POPEYE computes ITD (interaural time difference) signals at 100 Hz and stereo disparities at 15 Hz. These audio and visual observations are then fused by a AV clustering technique. 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.

5.2. Software packages

5.2.1. TransforMesh: Mesh evolution with applications to dense surface reconstruction

We completed the development of TransforMesh, started in 2007. It is a mesh-evolution software developed within the thesis of Andrei Zaharescu and recently submitted for journal publication [57]. It is a provably correct mesh-based surface evolution method. It is able to handle topological changes and self-intersections without imposing any mesh sampling constraints. The exact mesh geometry is preserved throughout, except for the self-intersection areas. Typical applications, including mesh morphing and 3-D reconstruction using variational methods, are currently handled. TransforMesh will is available as open source with LGPL on http://mvviewer.gforge.inria.fr/

5.2.2. 3D feature detector (MeshDOG) and descriptor (MeshHOG) for uniformly triangulated meshes

This is a C++ implementation of a 3D feature detector (MeshDOG) and a 3D feature descriptor (MeshHOG) for uniformly triangulated meshes, invariant to changes in rotation, translation, and scale. The descriptor is able to capture the local geometric and/or photometric properties in a succinct fashion. Moreover, the method is defined generically for any scalar function, e.g., local curvature. Typical applications include mesh description for "bag-of-features" kind of representations, mesh matching and mesh tracking. Both MeshDOG and MeshHOG are available as open source with LGPL on http://mvviewer.gforge.inria.fr/.

5.2.3. Shape and discrete-surface registration based on spectral graph matching

We started to develop a software package that registers shapes based on either their volumetric (voxels) or surface (meshes) representations. The software implements a spectral graph matching method combined with non-linear dimensionality reduction and with rigid point registration, as described in several recent publications and in [53] as well as in the PhD thesis of Diana Mateus [11]. The SpecMatch software package is publicly available as open source with GPL on http://open-specmatch.gforge.inria.fr/index.php.

5.2.4. *Real-time shape acquisition and visualization*

This software can be paraphrased as *from pixels to meshes*. It is a complete package that takes as input uncompressed image sequences grabbed with synchronized cameras. The software typically handles between 8 and 20 HDTV cameras, i.e., 2 million pixels per image. The software calibrates the cameras, segments the images into foreground (silhouettes) and background, and converts the silhouettes into a 3D meshed surface. The latter is smoothed and visualized using an image-based rendering technique. Currently this software package is commercialized by our start-up company, 4D View Solutions (http://www.4dviews.com). We continue to collaborate with this company. The latest version of the software is available for INRIA researchers and it runs on the GrImage platform.

5.2.5. *Audio-visual localization of speakers*

We developed a software package that uses binocular vision and binaural audition to spatially localize speakers. The software runs on the POPEYE platform (see above) and it has been developed in collaboration with the MISTIS project-team and with the Speech and Hearing group of the University of Sheffield. The software combines

stereo, interaural time difference, and expectation-maximization algorithms. It is developed within the European project POP.

5.3. Database

5.3.1. *Audio-visual database*

The University of Sheffield and INRIA have gathered synchronized auditory and visual datasets for the study of audio-visual fusion. The idea was to record a mix of scenarios where the audio-visual tasks of tracking the speaking face, where either the visual or auditory cues add disambiguating information; or more varied scenarios (eg. sitting in at a coffee break meeting) with a large amount of challenging audio and visual stimuli such as multiple speakers, varied amount of background noise, occulting objects, faces turned away and getting obscured, etc. Central to all scenarios is the state of the audio-visual perceiver and we have been very interesed in getting hold of some data recored with an active perceiver, so we propose that the perceiver is either static, panning or moving (probably limited to rotating its head) so as to mimic attending to the most interesting source at the moment. The calibrated data collection is freely accessible for research purposes at http://perception.inrialpes.fr/CAVA_Dataset/Site/

5.3.2. *4D repository (http://4drepository.inrialpes.fr/)*

This website hosts dynamic mesh sequences reconstructed from images captured using a multi-camera set up. Such mesh-sequences offer a new promising vision of virtual reality, by capturing real actors and their interactions. The texture information is trivially mapped to the reconstructed geometry, by back-projecting from the images. These sequences can be seen from arbitrary viewing angles as the user navigates in 4D (3D geometry + time) . Different sequences of human / non-human interaction can be browsed and downloaded from the data section. A software to visualize and navigate these sequences is also available for download.

6. New Results

6.1. Segmentation and registration of 3D shapes

One of the fundamental problems in 2D and 3D shape analysis is their segmentation and registration. We address both these issues within the framework of spectral graph theory. Namely, we started by developing a shape registration method based on *Spectral graph matching* and we applied this method for registering shapes described by meshes [53]. The method is based on the Laplacian embedding of meshes into an Euclidean space. Hence, the problem of matching two meshes is equivalent to the problem of registering two clouds of points. We thoroughly addressed the problem of point registration and we devised a probabilistic registration method based on Gaussian mixtures and convex optimization [54].

In parallel, we started to investigate the problem of shape segmentation using a combination of non-parametric and parametric clustering techniques, namely we combined spectral clustering with Gaussian mixture models [44], [22]. The ongoing work investigates constrained clustering methods and their use to the problem of learning shape segmentation. This completes our work on articulated object tracking [14].

6.2. Audio-visual perception

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 that 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 thouroughly 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 [55]

6.3. Stereoscopic vision

Current approaches to dense stereo matching estimate the disparity by maximizing its a posteriori probability, given the images and the prior probability distribution of the disparity function. This is done within a Markov random field model that makes tractable the computation of the joint probability of the disparity field. In this framework, we investigated the link between intensity-based stereo and contour-based stereo. In particular, we properly described surface-discontinuity contours for both piecewise planar objects and objects with smooth surfaces, and injected these contours into the probabilistic framework and the associated minimization methods. One drawback of such an approach, and of traditional stereo algorithms, is the use of the frontal parallel assumption that bias the results towards frontal parallel plane solution. To overcome this issue, we have investigated the use of a joint random Markov field, so that to each pixel is associated a disparity value and a surface normal. The estimation of the two field is done alternatively using minimization methods described above [56]

6.4. Analysis and Exploitation of the reflectance properties and lighting

6.4.1. Image-based modelling exploiting the surface normal vector field and the visibility

We have developed multiview 3D reconstruction algorithms (recovering the 3D shape and the reflectance of a scene surface) which explicitly exploit the visibility and the properties of the surface normals. Our approach thus allows to naturally combine stereo, silhouette and shading cues in a single framework. This method applies to a number of classical scenarios - classical stereovision, multiview photometric stereo, and multiview shape from shading [23]. This method which has been first developed in level-set framework [20], [46], has then been adapted to mesh representations which provide more accurate results and more robust and more stable algorithms. This last work has been submitted to the International Journal of Computer Vision in october 2009. In this framework, we are also working on several specific applications and scenarios such as Helmholtz Stereopsis and dynamic photometric stereo. This work is still in progress.

6.4.2. Shape from ambient shading

We study the mathematical and numerical aspects of the estimation of the 3-D shape of a Lambertian scene seen under diffuse illumination. This problem is known as âshape from ambient shadingâ (SFAS), and its solution consists of integrating a strongly non-local and non-linear Integro-Partial Differential Equation (I-PDE). We provide a first analysis of this global I-PDE, whereas previous work had focused on a local version that ignored effects such as occlusion of the light field. We also design an original approximation scheme which, following Barles and Souganidis' theory, ensures the correctness of the numerical approximations, and discuss about some numerical issues. For more details, see our SSVM'09 paper [42]. This work is still in progress.

6.4.3. Finding Images with Similar Lighting Conditions in Large Photo Collections.

In our previous works on the analysis and exploitation of reflectance properties and lighting, the goal was to obtain accurate 3D surface and reflectance models of objects. This required controlled image acquisition conditions. We have recently started another line of work concerning lighting and reflectance, for less controlled settings. Concretely, like in the PhotoSynth system by Microsoft, we wish to exploit the large amount and variety of photographs available for free on the internet, via commodity collection such as flickR. For many major monuments, it is easy to download hundreds or thousands of images. Whereas PhotoSynth exploits them to generate 3D information from the images, we intend to use them to tell as much as possible about the appearance of objects. Appearance depends on reflectance properties and the lighting – in the above photo collections, we usually find photos representative for many different lighting conditions. If we are able to extract models for the appearance of objects, they may be used for various applications, e.g. relighting, without necessarily having to handle and estimate detailed physics-based reflectance models. Our first work along these lines consists in developing methods for clustering photographs into sets of images with similar lighting conditions [35], [51]. To do so, we automatically segment the sky region in images and analyze the distribution of colors within it, leading to a clustering based on histograms of colors.

6.4.4. Multiple-View Geometry of the Refractive Plane.

Another new line of work concerns the analysis of images of semi-transparent objects. Our first result concerns the basic case of a planar refractive surface [28]. We have shown that images taken through such a surface, inspite the refractions, still have an epipolar geometry, and have shown how to formulate it analytically, *via* a fundamental matrix. Methods for estimating this and the relative position of cameras, have been proposed. Future work will be directed towards modeling of non-planar refractive semi-transparent surfaces.

6.4.5. Reconstructing specular surfaces.

In joint work with Jingyi Yu and Yuanyuan Ding from Delaware University, we have proposed a new method for reconstructing specular surfaces in 3D, from images [32]. The method exploits the fact that locally, specular surface patches act as reflectors that map 3D lines into conic curves on camera images. Observed conic curves reveal information about the specular surface, which is exploited here in a way analoguous to calibration of catadioptric cameras, albeit for more complicated shapes than the usual surfaces of revolution used for catadioptric cameras.

6.5. Multi-Region Segmentation on Manifolds

We have addressed the problem of segmenting data defined on a manifold into a set of regions with uniform properties. This issue is fundamental in a number computer vision and graphics applications. Here we have considered this issue in order to introduce segmentation constraints on the reflectance properties in 3D shape reconstruction problems.

We propose a numerical method when the manifold is represented by a triangular mesh. Based on recent image segmentation models, our method minimizes a convex energy and then enjoys significant favorable properties: it is robust to initialization and avoid the problem of the existence of local minima present in many variational models. Our contributions are threefold: firstly we adapt the convex image labeling model to manifolds; in particular the total variation formulation. Secondly we show how to implement the proposed method on triangular meshes, and finally we show how to use and combine the method in other computer vision problems, such as 3D reconstruction. For more details, see our ICCV'09 paper [29].

6.6. Surface tracking

We have proposed a new method to capture the temporal evolution of a surface from multiple videos. By contrast to most current methods, our algorithm does not use any prior information on the nature of the tracked surface. In addition, it does not require sparse features to constrain the deformation but only relies on purely geometric information: a target set of 3D points and normals. Our approach is inspired by the Iterative Closest Point algorithm but handles large deformations of non-rigid surfaces. To this aim, a mesh is iteratively deformed while enforcing local rigidity with respect to a reference model. This rigidity is preserved by diffusing it on local patches randomly seeded on the surface. The iterative nature of the algorithm combined with the softly enforced local rigidity allows to progressively evolve the mesh to fit the target data. The proposed method is validated and evaluated on several standard and challenging surface data sets acquired using real videos. [27]

6.7. Surface Feature Detection and Description with Applications to Mesh Matching

In this work we revisit local feature detectors/descriptors developed for 2D images and extend them to the more general framework of scalar fields defined on 2D manifolds. We provide methods and tools to detect and describe features on surfaces equiped with scalar functions, such as photometric information. This is motivated by the growing need for matching and tracking photometric surfaces over temporal sequences, due to recent advancements in multiple camera 3D reconstruction. We propose a 3D feature detector (MeshDOG) and a 3D feature descriptor (MeshHOG) for uniformly triangulated meshes, invariant to changes in rotation, translation, and scale. The descriptor is able to capture the local geometric and/or photometric properties in a succinct fashion. Moreover, the method is defined generically for any scalar function, e.g., local curvature. Results with rigid and non-rigid mesh matching demonstrate the interest of the proposed framework.[47].

6.8. Telepresence

Real-time multi-camera 3D modeling provides full-body geometric and photometric data on the objects present in the acquisition space. It can be used as an input device for rendering textured 3D models, and for computing interactions with virtual objects through a physical simulation engine. We are currently working on a collaborative environment where two distant users, each one 3D modeled in real-time, interact in a shared virtual world. [40]

6.9. Omnidirectional vision

6.9.1. Calibration of medical endoscopes.

Some endoscopes give a very wide field of view, in order to fully inspect e.g. vessels in the human body. Often, endoscopes are used for visualization purposes only, less frequently for making measurements. Nevertheless, an approximate calibration is required even for visualization, in order to present the wide-field of view images, which include strong distortion, in a more common way and without these distortions. In [26] we propose a very pragmatic method for endoscope calibration, well suited for non-expert users, requiring a single image of a planar grid. This is joint work with João Barreto from Coimbra University. In another joint work, we developed a more general calibration approach, requiring multiple images of a planar grid but allowing to calibrate a wider range of camera models, namely any central catadioptric camera [36]. It is based on a recent result of ours, showing the existence of fundamental matrices and plane homographies, for any such camera [1]

6.9.2. Geolocalization using Skylines from Omni-Images.

In this joint work with MERL [43], the goal is to provide an image-based method for geolocalization in cities, where the GPS is known to be unreliable. The developed method exploits the perhaps surprising fact that skylines, as seen from city streets, are rather characteristic for the camera location. We use upward-looking fisheye cameras with a roughly hemispheric field of view; in the acquired images, skylines are extracted relativelily reliable by segmenting the regions belonging to the sky. Nowadays, coarse 3D city models are easily available; given such a model, it is simple to generate the skyline that should be seen from any given location in the city. This is used in our geolocalization approach, where the camera location is found by comparing skylines extracted in images with those generated from candidate camera locations. This approach has been shown to be very robust and reasonable accurate, over image sequences corresponding to camera displacements of hundreds of meters.

6.10. Satellite imaging

6.10.1. Modeling, estimating and compensating the vibrations

Within a collaboration with ASTRIUM, we work on satellite imaging and linear pushbroom cameras. These cameras are widely used in passive remote sensing from space as they provide high resolution images. In earth observation applications, where several pushbroom sensors are mounted in a same focal plane, small dynamic disturbances of the satellite's orientation lead to noticeable geometrical distortions in the images. We have defined a global method to estimate those disturbances, which are effectively vibrations. We exploit the geometry of the focal plane and the stationary nature of the disturbances to recover undistorted images. To do so, we embed the estimation process in a Bayesian framework. An autoregressive model is used as a prior on the vibrations. The problem can be seen as a global image registration task where multiple pushbroom images are registered to the same coordinate system, the registration parameters being the vibration coefficients. An alternating maximisation procedure is designed to obtain Maximum a Posteriori estimates (MAP) of the vibrations as well as of the autoregressive model coefficients. A paper has been submitted to CVPR, and a patent has been applied.

6.11. Other results

6.11.1. Statistical approaches for computer animation

This research area is quite recent and very promising. This work has been done in close collaboration with the team EVASION of INRIA. We first addressed the problem of inverse kinematics, which is classically solved using deterministic numerical approaches. We proposed an original modelisation of the problem via a hidden Markov chain. Therefor, we provided a new algorithm based on a Monte Carlo filter that allows to solve sequentially the inverse kinematics problem. The work presented in [19] concerns the compression of motion capture sequences. We propose a novel, lossy compression algorithm that exploits both temporal and spatial coherence. This algorithm is based on approximation of pose manifold computed using a principal geodesics analysis.

6.11.2. Theoretical result on Sequential Monte Carlo

Sequential Monte Carlo approaches, also known as particle filters, are well-know tools for tracking appplications in computer vision. A longstanding problem in sequential Monte Carlo (SMC) is to mathematically prove the popular belief that resampling does improve the performance of the estimation (this of course is not always true, and the real question is to clarify classes of problems where resampling helps). A more pragmatic answer to the problem is to use adaptive procedures that have been proposed on the basis of heuristic considerations, where resampling is performed only when it is felt necessary, i.e. when some criterion (effective number of

---

[1] P. Sturm and J. Barreto, General Imaging Geometry for Central Catadioptric Cameras, ECCV'08

particles, entropy of the sample, etc.) reaches some prescribed threshold. It still remains to mathematically prove the efficiency of such adaptive procedures. The contribution of the paper [25] is to propose an approach, based on a representation in terms of multiplicative functionals (in which importance weights are treated as particles, roughly speaking) to obtain the asymptotic variance of adaptive resampling procedures, when the sample size goes to infinity. It is then possible to see the impact of the threshold on the asymptotic variance, at least in the Gaussian case, where the resampling criterion has an explicit expressions in the large sample asymptotics.

6.11.3. Geometric Video Projector Auto-Calibration.

We have developed several methods for the calibration and self-calibration of projector–camera system [50], [34], [33]. Typical classical methods have to augment the projection screen with markers. Our methods can do without this, but require moving the projector for the self-calibration procedure. In many cases, this is arguably less cumbersome to do. Our methods are based on the fact that projectors can be modeled like cameras, and adapt self-calibration methods designed for cameras, to projector–camera systems. This is joint work with Sébastien Roy from the University of Montréal.

6.11.4. Critical motions for camera self-calibration.

This is one of our historic topics in structure-from-motion research. This year, we have extended the existing results in two main ways [38]. First, a more general framework than previously was used, based on confocal quadric theory; this makes several results easier to obtain than previously. Second, we analyzed so-called artificial critical motions, i.e. degeneracies that arise due to the specific self-calibration method used and not due to a generic degeneracy. We have shown the artificial critical motions for the most representative family of self-calibration methods, which neglect a rank-constraint on the dual absolute quadric. We have also shown that enforcing that rank constraint *a posteriori*, allows to resolve all artificial degeneracies, i.e. allows self-calibration to be successful on more cases than previously possible with the considered methods. This is joint work with Pierre Gurdjos from IRI Toulouse and Adrien Bartoli from LASMEA, Clermont-Ferrand.

7. Contracts and Grants with Industry

7.1. Contract with ASTRIUM

In 2008, a three year contract has started with ASTRIUM. High-resolution satellite imagery is typically based on so-called push-broom cameras (one or a few rows of pixels, covering different spectral bands). High-resolution images are generated by stitching individual push-broom images, taken at successive time instants, together. The main goal of our work is to develop both theoretical models and algorithms for modeling, estimating and compensating these vibrations, using information contained in the images.

8. Other Grants and Activities

8.1. National initiatives

8.1.1. ANR project ROM

This is an ANR-funded "pre-industrial" project running for two years (2009-11). The coordinator is Duran Duboi, one of the leading French companies in (post-) production for movies and advertisement. The two academic partners are IRI Toulouse and PERCEPTION. The goal of the project is to develop tools for aiding the preparation of shooting sequences in complex settings, especially concerning a film camera that is moving during the shooting. A standard vision-based technique for generating special effects and other augmentations, is match-moving, with or without artificial markers. An important practical issue is that match-moving is typically performed off-line and if it fails, cumbersome manual work or even re-shooting becomes necessary. The tools we will develop will allow an efficient preparation of a shooting, by analyzing the scene before the shooting and automatically judging the feasibility of match-moving; if the feasibility is judged as too low, a tool will suggest to the operator where to augment the scene with artificial markers that will help the match-moving.

8.1.2. ANR project FLAMENCO

FLAMENCO is a 3-year project that has started on January 1, 2007. This project deals with the challenges of spatio-temporal scene reconstruction from several video sequences, i.e. from images captured from different viewpoints and at different time instants. This project tackles the following three important factors which limit the major problems in computer vision so far:

- the computational time / the poor resolution of the models: the acquisition of video sequences from multiple cameras generates a very large amount of data, which makes the design of efficient algorithms very important. The high computational cost of existing methods has limited the spatial resolution of the reconstruction and has allowed to handle video sequences of a few seconds only, which is prohibitive in real applications.
- the lack of spatio-temporal coherence: to our knowledge, none of the existing methods has been able to reconstruct coherent spatio-temporal models: Most methods build threedimensional models at each time step without taking advantage of the continuity of the motion and of the temporal coherence of the model. This issue requires elaborating new mathematical and algorithmic tools dedicated to four-dimensional representations (three space dimensions plus the time dimension).
- the simplicity of the models: the information available in multiple video sequences of a scene are not restricted to geometry and motion. Most reconstruction methods disregard such information as the illumination of the scene, and the reflectance, the materials and the textures of the objects. Our goal is to build more exhaustive models, by automatically estimating these parameters concurrently to geometry and motion. For example, in augmented reality, reflectance properties allow to synthesize novel views with higher photo-realism.

In this project, we are collaborating with the CERTIS laboratory (Ecole Nationale des Ponts et Chaussees) and the PRIMA group (INRIA Rhone-Alpes) via Frederic Devernay.

The team members directly involved in this project are Peter Sturm, Emmanuel Prados (INRIA researchers) and Amael Delaunoy (PhD thesis).

8.1.3. ADT GrimDev

GrimDev is an ADT (Action de Developpement Technologique) proposed in the context of the Grimage interactive and immersive platform. The objective of GrimDev is to organize and manage software developements around the Grimage platform in order to ensure their reusabilities and durations. GrimDev was proposed by the Perception team and involves the following teams from INRIA Grenoble Rhône-Alpes: Perception, Evasion, Moais and SED.

8.2. Projects funded by the European Commission

8.2.1. iGlance (European project MEDEA 2008-2011)

iGlance aims at developing new free viewpoint capabilities for the next TV generation. 10 partners are involved in this project including: ST microelectronics (France), Philips research (Holland), the university of Eindhoven(Holland), 4D View solutions (France), INRIA (France), Silicon Hive (Holland), Logica (France), Task 24 (Holland), Verum (Holland), Tima (France).

8.2.2. FP6/Marie-Curie RTN VISIONTRAIN

VISIONTRAIN is a 4 year Marie Curie Research Training Network, or RTN (2005-2009) coordinated by Radu Horaud. This network gathers 11 partners from 11 European countries and has the ambition to address foundational issues in computational and cognitive vision systems through an European doctoral and post-doctoral program.

VISIONTRAIN addresses the problem of understanding vision from both computational and cognitive points of view. The research approach is based on formal mathematical models and on the thorough experimental validation of these models. We intend to reduce the gap that exists today between biological vision (which performs outstandingly well and fast but not yet understood) and computer vision (whose robustness, flexibility, and autonomy remain to be demonstrated). In order to achieve these ambitious goals, 11 internationally recognized academic partners work cooperatively on a number of targeted research topics: computational theories and methods for low-level vision, motion understanding from image sequences, learning and recognition of shapes, categories, and actions, cognitive modelling of the action of seeing, and functional imaging for observing and modelling brain activity. There are three categories of researchers involved in this network: doctoral students, post-doctoral researchers, as well as highly experienced researchers. The work includes participation to proof-of-concept achievements, annual thematic schools, industrial meetings, attendance of conferences, etc.

8.2.3. FP6 IST STREP project POP

We are coordinators of the POP project (Perception on Purpose) involving the MISTIS and the PERCEPTION INRIA groups, as well as 4 other partners: University of Osnabruck (cognitive neuroscience), University Hospital Hamburg-Eppendorf (neurophysiology), University of Coimbra (robotics), and University of Sheffield (hearing and speech). POP proposes the development of a fundamentally new approach, perception on purpose, which is based on 5 principles. First, visual and auditory information should be integrated in both space and time. Second, active exploration of the environment is required to improve the audiovisual signal-to-noise ratio. Third, the enormous potential sensory requirements of the entire input array should be rendered manageable by multimodal models of attentional processes. Fourth, bottom-up perception should be stabilized by top-down cognitive function and lead to purposeful action. Finally, all parts of the system should be underpinned by rigorous mathematical theory, from physical models of low-level binocular and binaural sensory processing to trainable probabilistic models of audiovisual scenes.

8.3. Bi-lateral project

8.3.1. PHC project OMNILOC

This is a “Partenariat Hubert Curien” (PHC) between the University of Coimbra, Portugal (João Barreto) and PERCEPTION (2009-10). The goal is to study omnidirectional cameras and their use for camera localization.

9. Dissemination

9.1. Editorial boards and program committees

- Radu Horaud is a member of the following editorial boards:
  - advisory board member of the *International Journal of Robotics Research*,
  - editorial board member of the *International Journal of Computer Vision*,
  - area editor of *Computer Vision and Image Understanding*, and
  - associated editor of *Machine Vision Applications*.
- Peter Sturm is a member of the editorial boards of the *Image and Vision Computing* journal, the *Journal of Computer Science and Technology*, the *Journal of Mathematical Imaging and Vision*, the *International Journal on Intelligent Computing and Cybernetic* and the *Transactions on Computer Vision and Applications* – IPSJ (Information Processing Society of Japan)
- Peter Sturm has been Area Chair for CVPR, ICCV and ACCV.
- Peter Sturm has been Tutorial Chair for ICCV.
- Peter Sturm has has been a member of the Program Committees of:

  - Workshop and Symposium on Advanced Imaging Methods
  - Iberoamerican Congress on Pattern Recognition
  - International Conference on Computer Vision Theory and Applications
  - Iberian Conference on Pattern Recognition and Image Analysis
  - International Workshop on Combinatorial Image Analysis
  - Journés Jeunes Chercheurs ORASIS
  - Workshop on Video Event Categorization, Tagging and Retrieval
  - International Conference on Signal-Image Technology and Internet-Based Systems
  - Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras
  - Workshop on Community Based 3D Content and Its Applications In Mobile Internet Environments
- Edmond Boyer is a member of the editorial boards of the Image and Vision Computing journal and has been a member of the program committees of: cvpr2009, iccv2009, cvmp2009, mlvma'09 (Workshop iccv'09).

9.2. Services to the Scientific Community

- Peter Sturm is Chairman of theWorking Group “Imaging and Geometry” of the GdR ISIS (a French research network), 2006-09.
- Peter Sturm is Director of the "Geometry and Images" department of Laboratoire Jean Kuntzmann, 2008-09.
- Emmanuel Prados has been member of the Phd thesis of Mickael Péchaud, University Paris Diderot, oct. 2009

9.3. Teaching

- Representation de connaissance et inference, m2r, ujf, 15h, E. Arnaud
- Tutorat d'apprentis, m2p, ujf, 45 h, E. Arnaud
- image retrieval, m2p, ujf, 15h, E. Arnaud
- vision and robotics, m2r, 5h, E. Arnaud
- Outils mathematiques, m1, ujf, 10h, E. Arnaud
- Methodes statistiques pout la biologie, l2, ujf, 36h, E. Arnaud
- Informatique instrumentale et multimedia, l1, ujf, 66h, E. Arnaud
- Decouverte des mathematiques appliquees, l1, ujf, 18h, E. Arnaud
- 3D Modelling from Images or Videos, m2r, ujf, 18h, P. Sturm and E. Boyer
- Synthese d'images, m1, ujf, 60h, E. Boyer
- Projet, m1, ujf, 15h, E. Boyer
- Vision par Ordinateur, m2p, ujf, 40h, E. Boyer
- Synthese d'images, m1, Polytech, 36h, E. Boyer
- Modelisation 3D, m2r, inpg, 18h, E. Boyer
- Introduction aux techniques de l'images, l3, ujf, 15h, E. Boyer
- Master course (M2R) on Computer Vision, ujf, 20h, P. Sturm

9.4. Tutorials and invited talks

- Peter Sturm has given a tutorial on Computer Vision at the 14th Iberoamerican Congress on Pattern Recognition, Guadalajara, Mexico
- Peter Sturm has given invited conference talks at:
  - SCIA (Scandinavian Conference on Image Analysis, Oslo, Norway)
  - CIARP (Iberoamerican Congress on Pattern Recognition, Guadalajara, Mexico)
  - Two-yearly plenary meeting of the French GdR ISIS network (Batz-sur-Mer, France)
- Peter Sturm has given Invited seminars at:
  - KAIST (Korea Advanced Institute of Science and Technology), Daejeon, South Korea
  - Université de Montpellier, France
  - University of Barcelona, Spain
  - Technical University of Munich, Germany
- Edmond Boyer has given an invited talk at
  - Labri, Bordeaux (may 2009)
  - NAIST (Insitute of Science and Technology), Nara-Japan (October 2009)
- Edmond Boyer has been a keynote speaker at the International Conference on 3D Digital Imaging and Modeling 3DIM'09, Kyoto (octobre 2009)

9.5. Thesis

- Diana Mateus

10. Bibliography

Major publications by the team in recent years

[1] 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.

[2] R. HORAUD, G. CSURKA, D. DEMIRDJIAN. *Stereo Calibration from Rigid Motions*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 22, n^{o} 12, December 2000, p. 1446-1452, ftp://ftp.inrialpes.fr/pub/movi/publications/HoraudCsurkaDemirdjian-pami2000.ps.gz.

[3] S. LAZEBNIK, E. BOYER, J. PONCE. *On How to Compute Exact Visual Hulls of Object Bounded by Smooth Surfaces*, in "Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA", IEEE Computer Society Press, Dec 2001, http://perception.inrialpes.fr/publication.php3?bibtex=LBP01.

[4] S. PETITJEAN, E. BOYER. *Regular and Non-Regular Point Sets: Properties and Reconstruction*, in "Computational Geometry - Theory and Application", vol. 19, n^{o} 2-3, 2001, p. 101-126, http://perception.inrialpes.fr/publication.php3?bibtex=PB01.

[5] E. PRADOS, O. FAUGERAS. *Shape from Shading: a well-posed problem ?*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, California", vol. II, IEEE, jun 2005, p. 870–877, http://perception.inrialpes.fr/Publications/2005/PF05a.

[6] S. RAMALINGAM, P. STURM, S. LODHA. *Towards Complete Generic Camera Calibration*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California", vol. 1, jun 2005, p. 1093-1098, http://perception.inrialpes.fr/Publications/2005/RSL05a.

[7] A. RUF, R. HORAUD. *Visual Servoing of Robot Manipulators, Part I: Projective Kinematics*, in "International Journal of Robotics Research", vol. 18, n^{o} 11, November 1999, p. 1101-1118, http://hal.inria.fr/inria-00073002.

[8] P. STURM, S. MAYBANK. *On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications*, in "Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA", June 1999, p. 432-437.

[9] P. STURM, S. RAMALINGAM. *A Generic Concept for Camera Calibration*, in "Proceedings of the European Conference on Computer Vision, Prague, Czech Republic", vol. 2, Springer, May 2004, p. 1-13, http://perception.inrialpes.fr/Publications/2004/SR04.

[10] P. STURM. *Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico", Juin 1997, p. 1100-1105.

Year Publications

Doctoral Dissertations and Habilitation Theses

[11] D. MATEUS. *Spectral Tools for Unsupervised Modeling of Articulated Objects from Multiple-view Videos*, Institut National Polytechnique de Grenoble, Grenoble, France, September 2009, http://perception.inrialpes.fr/Publications/2009/Mat09, Ph. D. Thesis.

Articles in International Peer-Reviewed Journal

[12] D. COBZAS, M. JÄGERSAND, P. STURM. *3D SSD tracking with estimated 3D planes*, in "Image and Vision Computing", vol. 27, n^{o} 1, 2009, p. 69-79, http://hal.inria.fr/inria-00384230/en/CA.

[13] J.-S. FRANCO, E. BOYER. *Efficient Polyhedral Modeling from Silhouettes*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2009, http://hal.inria.fr/inria-00349103/en/.

[14] 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", vol. 31, n^{o} 1, 01 2009, p. 158-164, http://hal.inria.fr/inria-00446898/en/.

[15] C. LENGLET, E. PRADOS, J.-P. PONS, R. DERICHE, O. FAUGERAS. *Brain Connectivity Mapping using Riemannian Geometry, Control Theory and PDEs*, in "SIAM Journal on Imaging Sciences", vol. 2, n^{o} 2, 2009, p. 285-322, http://hal.inria.fr/inria-00377379/en/US.

[16] T. PRIBANIC, P. STURM, S. PEHAREC. *Wand Based Calibration of 3D Kinematic System*, in "IET Computer Vision", vol. 3, n^{o} 3, 2009, p. 124-129, http://hal.inria.fr/inria-00434345/en/HR.

[17] R. SOSA, E. ARNAUD, E. MEMIN, G. ARTANA. *Study of the Flow Induced by a Sliding Discharge*, in "IEEE Transactions on Dielectrics and Electrical Insulation", vol. 16, n^{o} 2, 2009, p. 305–311, http://hal.inria.fr/inria-00377307/en/AR.

[18] 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", vol. 31, n^{o} 9, 2009, p. 1552-1566, http://hal.inria.fr/inria-00384203/en/CA.

[19] M. TOURNIER, X. WU, N. COURTY, E. ARNAUD, L. REVERET. *Motion Compression using Principal Geodesics Analysis*, in "Computer Graphics Forum (Proceedings of Eurographics 2009)", 2009, http://hal.inria.fr/inria-00384213/en/.

[20] 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 (IJCV)", 2009, http://hal.inria.fr/inria-00266293/en/.

[21] A. ZAHARESCU, R. HORAUD. *Robust Factorization Methods Using A Gaussian/Uniform Mixture Model*, in "International Journal of Computer Vision", vol. 81, 03 2009, p. 240-258, http://hal.inria.fr/inria-00446987/en/.

Invited Conferences

[22] R. HORAUD. *Combining spectral and model-based clustering for shape segmentation and registration*, in "Working Group on Model-Based Clustering Summer Session, Paris", July 2009, http://perception.inrialpes.fr/Publications/2009/Hor09.

[23] P. STURM. *3D and appearance modeling from images*, in "Iberoamerican Congress on Pattern Recognition, Mexique Guadalajara", 2009, http://hal.inria.fr/inria-00386349/en/.

[24] P. STURM. *Complete photoconsistency optimization in multi-view stereo*, in "Scandinavian Conference on Image Analysis, Norvège Oslo", 2009, http://hal.inria.fr/inria-00386350/en/.

International Peer-Reviewed Conference/Proceedings

[25] E. ARNAUD, F. LE GLAND. *SMC with Adaptive Resampling: Large Sample Asymptotics*, in "IEEE Workshop on Statistical Signal Processing, Royaume-Uni Cardiff", 2009, http://hal.inria.fr/inria-00421619/en/.

[26] J. BARRETO, P. STURM. *Automatic Calibration of Medical Endoscopes using a Single Image of a Planar Grid*, in "20th British Machine Vision Conference, Royaume-Uni London", 2009, http://hal.inria.fr/inria-00434336/en/PT.

[27] C. CAGNIART, E. BOYER, I. SLOBODAN. *Iterative Mesh Deformation for Dense Surface Tracking*, in "3DIM'2009, Japon Kyoto", 2009, http://hal.inria.fr/inria-00440389/en/DE.

[28] V. CHARI, P. STURM. *Multiple-View Geometry of the Refractive Plane*, in "20th British Machine Vision Conference, Royaume-Uni London", 2009, http://hal.inria.fr/inria-00434342/en/.

[29] A. DELAUNOY, K. FUNDANA, E. PRADOS, A. HEYDEN. *Convex Multi-Region Segmentation on Manifolds*, in "The 12th IEEE International Conference on Computer Vision, Japon Kyoto", 2009, http://hal.inria.fr/inria-00435133/en/SE.

[30] Y. DING, J. YU, P. STURM. *Multiperspective Stereo Matching and Volumetric Reconstruction*, in "IEEE International Conference on Computer Vision, Japon Kyoto", 2009, http://hal.inria.fr/inria-00434341/en/US.

[31] Y. DING, J. YU, P. STURM. *Recovering Specular Surfaces Using Curved Line Images*, in "IEEE Conference on Computer Vision and Pattern Recognition, États-Unis d'Amérique Miami", 2009, http://hal.inria.fr/inria-00384250/en/US.

[32] Y. DING, J. YU, P. STURM. *Recovering Specular Surfaces Using Curved Line Images*, in "IEEE International Conference on Computer Vision, Japon Kyoto", 2009, http://hal.inria.fr/inria-00434344/en/US.

[33] J. DRARÉNI, S. ROY, P. STURM. *Geometric Video Projector Auto-Calibration*, in "IEEE International Workshop on Projector-Camera Systems, États-Unis d'Amérique Miami", 2009, http://hal.inria.fr/inria-00384221/en/CA.

[34] J. DRARÉNI, S. ROY, P. STURM. *Projector Calibration Using a Markerless Plane*, in "International Conference on Computer Vision Theory and Applications, Portugal Lisbon", vol. 2, 2009, p. 377-382, http://hal.inria.fr/inria-00384242/en/CA.

[35] M. DÍAZ, P. STURM. *Finding Images with Similar Lighting Conditions in Large Photo Collections*, in "Iberoamerican Congress on Pattern Recognition, Mexique Guadalajara", 2009, p. 53-60, http://hal.inria.fr/inria-00434337/en/.

[36] S. GASPARINI, P. STURM, J. BARRETO. *Plane-Based Calibration of Central Catadioptric Cameras*, in "IEEE International Conference on Computer Vision, Japon Kyoto", 2009, http://hal.inria.fr/inria-00434343/en/PT.

[37] A. GLENNERSTER, M. HANSARD, A. FITZGIBBON. *View-based approaches to Spatial Representation in Human Vision*, in "Proc. Dagstuhl Seminar on Statistical and Geometrical Approaches to Visual Motion Analysis, Allemagne Dagstuhl", 2009, http://hal.inria.fr/inria-00435556/en/GBUS.

[38] P. GURDJOS, A. BARTOLI, P. STURM. *Is Dual Linear Self-Calibration Artificially Ambiguous?*, in "IEEE International Conference on Computer Vision, Japon Kyoto", 2009, http://hal.inria.fr/inria-00434340/en/.

[39] Z. JANKO, J.-P. PONS. *Spatio-Temporal Image-Based Texture Atlases for Dynamic 3-D Models*, in "IEEE International Workshop on 3-D Digital Imaging and Modeling, Japon Kyoto", 2009, http://hal.inria.fr/inria-00435534/en/.

[40] B. PETIT, J.-D. LESAGE, E. BOYER, J.-S. FRANCO, B. RAFFIN. *Remote and Collaborative 3D Interactions*, in "IEEE 3DTV Conference, Allemagne Potsdam", 2009, http://hal.inria.fr/inria-00405810/en/.

[41] B. PETIT, J.-D. LESAGE, E. BOYER, B. RAFFIN. *Virtualization Gate*, in "Siggraph - Emerging Technologies, États-Unis d'Amérique New Orleans", 2009, http://hal.inria.fr/inria-00418917/en/.

[42] E. PRADOS, N. JINDAL, S. SOATTO. *A Non-Local Approach to Shape From Ambient Shading*, in "2nd International Conference on Scale Space and Variational Methods in Computer Vision (SSVM'09), Norvège Voss", Springer-Verlag, 2009, p. 696–708, http://hal.inria.fr/inria-00357071/en/US.

[43] S. RAMALINGAM, S. BOUAZIZ, P. STURM, M. BRAND. *Geolocalization using Skylines from Omni-Images*, in "IEEE Workshop on Search in 3D and Video, Japon Kyoto", 2009, http://hal.inria.fr/inria-00434339/en/US.

[44] A. SHARMA, R. HORAUD, D. KNOSSOW, E. VON LAVANTE. *Mesh Segmentation Using Laplacian Eigenvectors and Gaussian Mixtures*, in "AAAI Fall Symposium on Manifold Learning and its Applications, Arlington,

VA États-Unis d'Amérique", R. SOUVENIR (editor), AAAI Press, AAAI, 11 2009, p. 50-56, http://hal.inria.fr/inria-00446990/en/.

[45] P. STURM, A. DELAUNOY, P. GARGALLO I PIRACÉS, E. PRADOS, K.-J. YOON. *3D and Appearance Modeling from Images*, in "14th Iberoamerican Congress on Pattern Recognition, Mexique Guadalajara", 2009, http://hal.inria.fr/inria-00434333/en/KR.

[46] K.-J. YOON, E. PRADOS, P. STURM. *Generic Scene Recovery using Multiple Images*, in "2nd International Conference on Scale Space and Variational Methods in Computer Vision (SSVM'09), Norvège Voss", Springer, 2009, http://hal.inria.fr/inria-00357139/en/.

[47] A. ZAHARESCU, E. BOYER, V. KIRAN, R. HORAUD. *Surface Feature Detection and Description with Applications to Mesh Matching*, in "International Conference on Computer Vision and Pattern Recognition, États-Unis d'Amérique Miami", 2009, http://hal.inria.fr/inria-00440407/en/.

National Peer-Reviewed Conference/Proceedings

[48] C. CAGNIART, E. BOYER, I. SLOBODAN. *Suivi de Surfaces par Déformations Itératives*, in "22es journées de l'Association Française d'Informatique Graphique, France Arles", 2009, http://hal.inria.fr/inria-00440425/en/DE.

[49] A. DELAUNOY, E. PRADOS, P. GARGALLO, J.-P. PONS, P. STURM. *Stéréo multi-vues : erreur de reprojection et maillages triangulaires*, in "ORASIS'09 - Congrès des jeunes chercheurs en vision par ordinateur, France Trégastel, France", 2009, http://hal.inria.fr/inria-00404620/en/.

[50] J. DRARÉNI, S. ROY, P. STURM. *Calibration de projecteur par plan non-marqué*, in "ORASIS'09 - Congrès des jeunes chercheurs en vision par ordinateur, France Trégastel, France", 2009, http://hal.inria.fr/inria-00404608/en/CA.

[51] M. DÍAZ, P. STURM. *Recherche d'Images avec des Conditions d'Illumination Similaires dans une Collection de Photos*, in "ORASIS'09 - Congrès des jeunes chercheurs en vision par ordinateur, France Trégastel, France", 2009, http://hal.inria.fr/inria-00404615/en/.

Scientific Books (or Scientific Book chapters)

[52] S. GASPARINI, P. STURM. *Multi-View Matching Tensors from Lines for General Camera Models*, in "Tensors in Image Processing and Computer Vision", S. AJA-FERNANDEZ, R. DE LUIS GARCIA, D. TAO, X. LI (editors), vol. 33, Springer, 2009, http://hal.inria.fr/inria-00384237/en/.

[53] D. KNOSSOW, A. SHARMA, D. MATEUS, R. HORAUD. *Inexact Matching of Large and Sparse Graphs Using Laplacian Eigenvectors*, in "Graph-Based Representations in Pattern Recognition", A. TORSELLO, F. ESCOLANO, L. BRUN (editors), Lecture Notes in Computer Science, vol. 5534/2009, Springer Berlin / Heidelberg, 05 2009, p. 144-153, http://hal.inria.fr/inria-00446989/en/.

Research Reports

[54] R. HORAUD, F. FORBES, M. YGUEL, G. DEWAELE, J. ZHANG. *Rigid and Articulated Point Registration with Expectation Conditional Maximization*, INRIA, 2009, http://hal.inria.fr/inria-00435772/en/, RR-7114, Rapport de recherche.

[55] V. KHALIDOV, F. FORBES, R. HORAUD. *Conjugate Mixture Models for Clustering Multimodal Data*, INRIA, 2009, http://hal.inria.fr/inria-00436468/en/, RR-7117, Rapport de recherche.

[56] R. NARASIMHA, E. ARNAUD, F. FORBES, R. HORAUD. *A Joint Framework for Disparity and Surface Normal Estimation*, INRIA, 2009, http://hal.inria.fr/inria-00430635/en/, RR-7090, Rapport de recherche.

[57] A. ZAHARESCU, E. BOYER, R. HORAUD. *Topology-Adaptive Mesh Deformation for Surface Evolution, Morphing, and Multi-View Reconstruction*, INRIA, 2009, http://hal.inria.fr/inria-00438358/en/, RR-7136, Rapport de recherche.