INRIA

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

Activity Report

2010

Table of contents

1. Team

Research Scientists

Radu Horaud [Team leader, Research Director (DR), HdR]
Peter Sturm [Research Director (DR), sabbatical at TU Munich, HdR]

Faculty Member

Edmond Boyer [Université Joseph Fourier Grenoble, HdR]

Technical Staff

Michel Amat [Development Engineer, from September 2010]
Gaëtan Janssens [Development Engineer]
Simone Gasparini [Development Engineer]

PhD Students

Xavier Alameda-pineda [MESR grant]
Visesh Chari [INRIA grant]
Amaël Delaunoy [INRIA grant]
Antoine Deleforge [MESR grant]
Mauricio Diaz [Alban-EU grant]
Jamil Draréni [Co-supervision agreement with Université de Montréal]
Kaustubh Kulkarni [INRIA grant]
Antoine Letouzey [INRIA grant]
Ramya Narasimha [INRIA grant]
Régis Perrier [INRIA grant]
Benjamin Petit [INRIA grant]
Jordi Sanchez-riera [INRIA grant]
Avinash Sharma [INRIA grant]
Kiran Varanasi [INRIA grant]

Post-Doctoral Fellows

Jan Cech
Etienne Von Lavante [until September 2010]
Miles Hansard

Visiting Scientist

Kawasaki Hiroshi [Visiting Professor,Saitama University, Japan]

Administrative Assistant

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 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 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. The ANR project Morpho – Analysis of Human Shapes and Motions

MORPHO is aimed at designing new technologies for the measure and for the analysis of dynamic surface evolutions using visual data. 3 academic partners will collaborate on this project: the INRIA Rhône-Alpes with the Perception team and the Evasion team, the GIPSA-lab Grenoble and the INRIA-Lorraine with the Alice team.

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

We started a 12 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.

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 kinect project.

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 multi-camera platform dedicated to spatio-temporal modeling including immersive and interactive applications. It hosts a multiple-camera system connected to a PC cluster, as well as visualization facilities including head mounted displays. This platform 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.

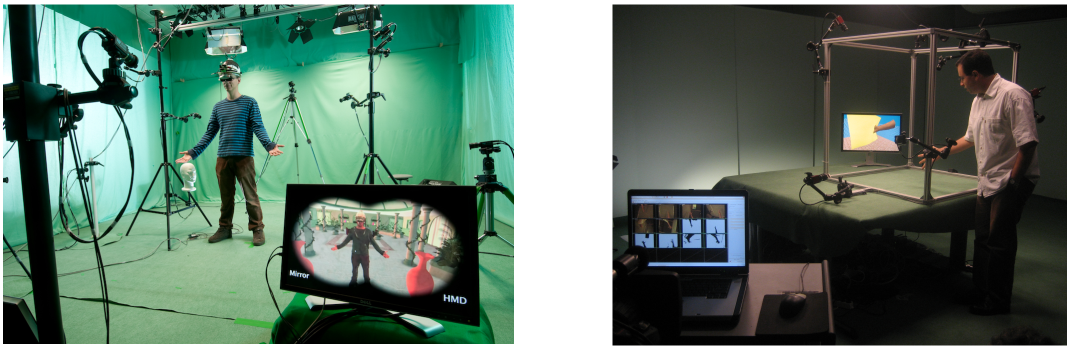

Figure 1. Left: The Grimage platform allows immersive/interactive applications such as this one. The real character is reconstruced in real-time and immersed in a virtual world, such that he/she can interact with virtual objects. Right: The mini-Grimage platform holds on a table top. It uses six cameras 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 Rhône-Alpes, and in collaboation with the company 4D View Solutions. The PERCEPTION team is leading this project.

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. LucyViewer

Lucy Viewer http://4drepository.inrialpes.fr/lucy_viewer/ is an interactive viewing software for 4D models, i.e, dynamic three-dimensional scenes that evolve over time. Each 4D model is a sequence of meshes with associated texture information, in terms of images captured from multiple cameras at each frame. Such data is available from various website ver the world including the 4D repository website hosted by INRIA Grenoble http://4drepository.inrialpes.fr/. The software was developed in the context of the European project iGlance, it is available as an open source software under the Gnu LGP Licence.

5.2.2. 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. 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/

5.2.3. 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/.

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

We continued 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 publications as well as in the PhD thesis of Diana Mateus (2009). The SpecMatch software package is publicly available as open source with GPL on http://open-specmatch.gforge.inria.fr/index.php.

5.2.5. 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.

5.2.6. Audio-visual localization of speakers

We completed the development of 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 was developed within the European project POP. New developments are foreseen in the European project HUMAVIPS.

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 and of diffusion geometry. Namely, we started by developing a shape registration method based on *spectral graph matching and heat-kernel embedding* and we applied this method for registering shapes described by meshes [41]. The method is based on the Laplacian and heat-kernel 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.

In parallel, we started to investigate the problem of shape segmentation using a combination of unsupervised and semi-supervised clustering techniques, namely we combined spectral clustering with with probabilistic label transfer based on Gaussian mixtures [42].

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. This work will be continued within the HUMAVIPS project.

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 [12], [35]

6.4. Analysis and Exploitation of Reflectance Properties and Lighting

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

Over the last few years, 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 [21].

6.4.2. Recovery of appearance information from community 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 community 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.

We have recently developed several approaches to simultaneously recover the illumination present in the scene, the surface appearance of an object, and the cameras' radiometric calibration, from unstructured photo collections [29]. The problem is difficult since these images have been acquired by different cameras, at a different dates/times, under different lighting and weather conditions. To make it tractable, we use several priors, on radiometric calibration (Grossberg and Nayar's result on the empirical space of camera response functions) and surface appearance (spatially varying albedo). As for the illumination, we so far investigated two different models, a first one consisting of a directional light source and ambient lighting and a second one where lighting is modeled *via* spherical harmonics. Even with these priors, the problem remains difficult and has many variable to estimate. We developed various closed-form solutions for initialization purposes, and a robust non-linear optimization approach for refining initial estimates. Our results are very promising – it was found to be possible to recover e.g. the illumination present in the scene even with the minimal prior knowledge we have at our disposal.

6.5. Omnidirectional vision

6.5.1. Geolocalization from omnidirectional images using building skylines

This is joint work with MERL that has been carried out in the last two years. 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 relatively reliably by segmenting 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 reasonably accurate, over image sequences corresponding to camera displacements of hundreds of meters [40].

6.5.2. Generic self-calibration approach

In [19], we have finalized and improved upon previous works on self-calibration for generic camera models. From 2004 on, we performed research on generic camera models, that are highly flexible and can accomodate practically any camera in routine use, be it based on a fish-eye, a catadioptric design, etc. Our work on self-calibration is one of the most general works in the literature: all that is assumed by our approaches is that a camera has a single effective viewpoint and that its imaging geometry is continuous. Otherwise, the camera's projection function may be entirely arbitrary. In [19], we propose a theoretical analysis of the self-calibration problem andseveral practical approaches, based on particular camera motions, such as pure translations and rotations. The results we have obtained are surprisingly accurate.

6.5.3. Plane-based calibration for linear cameras

Line scanners and pushbroom cameras (still a dominant imaging technology for space-borne applications) have a particular imaging geometry: a 1D sensor acquires images while translating, leading to 2D images that are not perspective. We have developed a practical approach for calibrating such systems [15]. Like the standard approach for regular cameras, we employ planar calibration grids, which are easy to manufacture and to handle.

6.6. Active 3D modelling approaches

6.6.1. 3D scanning with a projector–camera system

Some structured light type 3D acquisition systems are based on the combination of one or several cameras with one or several projectors. The latter project appropriate patterns onto the scene that allow to establish matches between cameras or between cameras and projectors, that enable 3D modelling, even if modelled surfaces are textureless. In [32], we propose an approach that works with a single camera and a single projector, even if their relative position is unknown. The motivation is to reach the flexibility to move the camera and/or projector around an object, to obtain a complete 3D model, and to carry out this motion without dedicated equipment, e.g. to use a handheld camera. This added flexibility comes with the price of a much more difficult matching problem. We developed a theoretical analysis of the matching problem in our case and a practical method that solves the matching problem in a combinatorial framework. The outcome of this is a so-called one-shot 3D acquisition system: from a single acquired image, one obtains a 3D object model, whereas traditional structured light systems require several images, corresponding to a varying projected pattern. This thus opens the way for the 3D acquisition of deformable objects with an extremely simple and cheap hardware setup. This work is a collaboration with the group by Prof. Kawasaki from Saitama/Kagoshima University.

6.6.2. 3D modeling from shadowgrams

A specific active approach for 3D modeling consists in illuminating an object by a single light source and capturing the shadow cast on a flat surface, with a camera. Images such acquired are called shadowgrams and have been analyzed previously e.g. by the group of Kanade at CMU. The 3D modeling is based on the silhouettes of the shadowgrams and is thus akin to traditional shape-from-silhouettes. If the position of the light source is known for each image, the problem is thus solved easily, by any known shape-from-silhouette, or visual hull, approach. We analyzed the case of unknown light source positions [28] – it turns out that in the general case, the 3D model is ambiguous and that the amount of ambiguity is characterized by the so-called bas-relief transformations, which also play a role in photometric stereo. We showed how to reduce this ambiguity by exploiting direct observations of the light sources in the camera, thus giving rise to a 3D modeling approach from “uncalibrated” shadowgrams.

6.7. Satellite imaging

6.7.1. Modeling, estimating and compensating vibrations

Within a collaboration with EADS 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 global methods 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. We also propose an approach to fuse image date with data from other sensors, such as so-called star trackers and inertial sensor readings. These works have been published in [36], [38], [48], [39] and a patent has been filed that covers the developed approaches.

6.8. Spatio Temporal Modeling

6.8.1. Temporally Consistent Segmentation

We have addressed the problem of segmenting consistently an evolving 3D scene reconstructed individually at different time-frames. The spatial reconstruction of 3D objects from multiple views has been studied extensively in the recent past. Various approaches exist to capture the performance of real actors into voxel-based or mesh-based representations. However, without any knowledge about the nature of the scene being observed, such reconstructions suffer from artifacts such as holes and topological inconsistencies. We explore the problem of segmenting such reconstructions in a temporally coherent manner, as a decomposition into rigidly moving parts. We work with mesh-based representations, though our method can be extended to other 3D representations as well. Unlike related works, our method is independent of the scene being observed, and can handle multiple actors interacting with each other. We also do not require an ‘a priori’ motion-estimate, which we compute simultaneously as we segment the scene. We individually segment each of the reconstructed scenes into approximately convex parts, and compute their reliability through rigid motion estimates over the sequence. We finally merge these various parts together into a holistic and consistent segmentation over the sequence. See [44] for more information.

6.8.2. Surface Tracking

Tracking arbitrary shapes that evolve over time is a fundamental issue in computer vision motivated by the ever growing number of applications that require consistent shape information along temporal sequences. We have proposed a framework that considers a temporal sequence of independently reconstructed surfaces and iteratively deforms a reference mesh to fit these observations. To effectively cope with outlying and missing geometry, we introduce a novel probabilistic mesh deformation framework. Using generic local rigidity priors and accounting for the uncertainty in the data acquisition process, this framework effectively handles missing data, relatively large reconstruction artefacts and multiple objects. Extensive experiments demonstrate the effectiveness and robustness of the method on various 4D datasets. See [23], [22], [24] for more details.

6.8.3. Occupancy Grids

We have investigated shape and motion retrieval in the context of multi-camera systems. We propose a new lowlevel analysis based on latent silhouette cues, particularly suited for low-texture and outdoor datasets. Our analysis does not rely on explicit surface representations, instead using an EM framework to simultaneously update a set of volumetric voxel occupancy probabilities and retrieve a best estimate of the dense 3D motion field from the last consecutively observed multi-view frame set. As the framework uses only latent, probabilistic silhouette information, the method yields a promising 3D scene analysis method robust to many sources of noise and arbitrary scene objects. It can be used as input for higher level shape modeling and structural inference tasks. We validated the approach and demonstrated its practical use for shape and motion analysis experimentally [30], [47].

6.9. Telepresence

Networked virtual environments like Second Life enable distant people to meet for leisure as well as work. But users are represented through avatars controlled by keyboards and mouses, leading to a low sense of presence especially regarding body language. Multi-camera real-time 3D modeling offers a way to ensure a significantly higher sense of presence. But producing quality geometries, well textured, and to enable distant user tele-presence in non trivial virtual environments is still a challenge today. In this paper we present a tele-immersive system based on multi-camera 3D modeling. Users from distant sites are immersed in a rich virtual environment served by a parallel terrain rendering engine. Distant users, present through their 3D model, can perform some local interactions while having a strong visual presence. We experimented our system between three large cities a few hundreds kilometers apart from each other. This work demonstrate the feasibility of a rich 3D multimedia environment ensuring users a strong sense of presence [18], [54].

6.10. Image Collections

We have considered large image collections and their organization into meaningful data structures upon which applications can be build (e.g. navigation or reconstruction). In contrast to structures that only reflect local relationships between pairs of images we propose to account for the information an image brings to a collection with respect to all other images. Our approach builds on abstracting from image domains and focusing on image regions, thereby reducing the influence of outliers and background clutter. We introduce a graph structure based on these regions which encodes the overlap between them. The contribution of an image to a collection is then related to the amount of overlap of its regions with the other images in the collection. We demonstrate our graph based structure with several applications: image set reduction, canonical view selection and image-based navigation. The data sets used in our experiments range from small examples to large image collections with thousands of images [34].

6.11. Other results

6.11.1. Video deblurring and super-resolution

In joint work with Prof. Kawasaki and his group, we developed an approach for the difficult problem of video deblurring and super-resolution, for the case of multiple independently moving objects [45]. It combines an optical flow based image segmentation with motion deblurring and super-resolution and is able to sucessfully process handheld video sequences.

6.11.2. Calibration of camera networks

In a collaboration with ENPC, we contributed methods for camera network calibration [25], [26]. The originality of these methods is that they combine local calibration information – for triplets of cameras – in an efficient manner and that they allow to handle loops in the network.

7. Contracts and Grants with Industry

7.1. Contract with SAMSUNG

We started a 12 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. The first version of the software was successfully installed in November 2010 in Korea.

7.2. Contract with EADS 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 satellite vibrations, using information contained in the images.

7.3. Contract with ESA and EADS Astrium

In september 2010, a one-year contract has started between PERCEPTION and EADS Astrium, financed by the European Space Agency (ESA). The objective of this contract is a feasibility study of image-based 3D modeling approaches for the 3D modeling of asteroids.

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 in 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 Frédéric Devernay.

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

8.1.3. ANR project Morpho – Analysis of Human Shapes and Motions

MORPHO is aimed at designing new technologies for the measure and for the analysis of dynamic surface evolutions using visual data. Optical systems and digital cameras provide a simple and non invasive mean to observe shapes that evolve and deform and we propose to study the associated computing tools that allow for the combined analyses of shapes and motions. Typical examples include the estimation of mean shapes given a set of 3D models or the identification of abnormal deformations of a shape given its typical evolutions. Therefore this does not only include static shape models but also the way they deform with respect to typical motions. It brings a new research area on how motions relate to shapes where the relationships can be represented through various models that include traditional underlying structures, such as parametric shape models, but are not limited to them. The interest arises in several application domains where temporal surface deformations need to be captured and analyzed. It includes human body analyses but also extends to other deforming objects, sails for instance. Potential applications with human bodies are anyway numerous and important, from the identification of pathologies to the design of new prostheses. The project focus is therefore on human body shapes and their motions and on how to characterize them through new biometric models for analysis purposes. 3 academic partners will collaborate on this project: the INRIA Rh ne-Alpes with the Perception team and the Evasion team, the GIPSA-lab Grenoble and the INRIA-Lorraine with the Alice team.

8.1.4. ADT Vgate

Following the ADT (Action de Developpement Technologique) GrimDev proposed in the context of the Grimage interactive and immersive platform, we have proposed the ADT Vgate. The objective of Vgate is to manage the evolution of the Grimage platform both on the hardware and software sides to ensure improvements, reusability and durations of the Grimage platform perception and immersion capabilities. Vgate was proposed in collaboration with the EPI Moais from the INRIA Grenoble Rhône-Alpes.

8.2. European Initiatives

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. FP7 ICT STREP project HUMAVIPS

We are coordinators of the HUMAVIPS project involving the MISTIS and the PERCEPTION INRIA groups, as well as 4 other partners: The Czech Technical University, University of Bielefeld, IDIAP Institute, and Aldebaran Robotics.

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 s 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.

Website: http://humavips.inrialpes.fr.

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.

8.3.2. PHC project Temporally Consistent 3D Reconstruction and Action Recognition with a Multiple-Camera System

This is a “Partenariat Hubert Curien” (PHC) between the Technical University of Munich, Germany and PERCEPTION (2010-11). The scientific objectives of this collaboration aim at the advancement of temporal aspects of the 3D reconstruction of dynamic scenes and the human action recognition in multiple-camera systems.

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 member of the Scientific Council of the Barcelona Media Technology Centre.
- Peter Sturm is member of the following editorial boards:
  - Image and Vision Computing Journal
  - Journal of Computer Science and Technology
  - Journal of Mathematical Imaging and Vision
  - International Journal on Intelligent Computing and Cybernetic
  - Transactions on Computer Vision and Applications – IPSJ (Information Processing Society of Japan)
- Peter Sturm has been Area Chair for ACCV 2010.
- Peter Sturm has has been a member of the Program Committees of:
  - DAGM – Symposium of the German Association for Pattern Recognition
  - PCV – Symposium on Photogrammetric Computer Vision
  - OMNIVIS – Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras
  - CIARP – Iberoamerican Congress on Pattern Recognition
  - OmniRoboVis – Workshop on Omnidirectional Robot Vision
  - ICVGIP – Indian Conference on Computer Vision, Graphics and Image Processing
  - VECTaR – Workshop on Video Event Categorization, Tagging and Retrieval
  - SITIS – International Conference on Signal-Image Technology and Internet-Based Systems
- Edmond Boyer is a member of the editorial board of the Image and Vision Computing journal.
- Edmond Boyer has been Area Chair for ECCV 2010.
- Edmond Boyer has been a member of the program committees of: cvpr2010, cvmp2010, 3DPVT2010, BMVC2010, ACCV2010, SGA2010.

9.2. Services to the Scientific Community

- Peter Sturm is Co-chairman of the Working Group “Imaging and Geometry” of the GdR ISIS (a French research network), since 2006.
- Peter Sturm was reviewer of one habilitation thesis and of seven PhD theses, and examiner of one PhD thesis.
- Peter Sturm has been president of the national INRIA working group on “Actions Incitatives" since 2007.
- Edmond Boyer was reviewer of two PhD theses and examiner of one PhD thesis.
- Edmond Boyer was a member of the recruiting committees of the ENS Paris and of the ENSIMAG Grenoble.

9.3. Teaching

- 3D Modelling from Images or Videos, m2r, UJF, 18h, 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
- Generic Camera Models and 3D Computer Vision, Post-graduate, Oulu University (Finland), 8h, P. Sturm

9.4. Tutorials and invited talks

- Peter Sturm has given invited talks at:
  - European Patent Office, Munich, Germany
  - Computer Vision Colloquium, Tokyo, Japan
- Edmond Boyer has given talks at:
  - AERES evaluation of the LJK, January 2010.
  - AERES evaluation of the INRIA Grenoble Rhône-Alpes, March 2010.
  - TU Munich, March 2010.
  - ECCV area chair colloquium, Paris, May 2010.
  - ESIEE Paris, December 2010.
- Radu Horaud gave a tutorial on "Shape Analysis Using Diffusion Geometry" at ECCV'10.

9.5. Thesis

- Jamil Draréni (international co-supervision agreement with Université de Montréal)
- Ramya Narasimha
- Kiran Varanasi

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", December 2000, vol. 22, n^{o} 12, 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", 2001, vol. 19, n^{o} 2-3, 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", IEEE, jun 2005, vol. II, 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", jun 2005, vol. 1, 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", November 1999, vol. 18, n^{o} 11, 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", Springer, May 2004, vol. 2, 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.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] J. Draréni. *Exploitation de contraintes photométriques et géométriques en vision. Application au suivi, au calibrage et à la reconstruction*, Institut National Polytechnique de Grenoble, October 2010, http://perception.inrialpes.fr/publication.php3?bibtex=Dra10.

[12] R. Narasimha. *Méthodes d'estimation de la profondeur par mise en correspondance stéréoscopique à l'aide de champs aléatoires couplés*, Université Joseph-Fourier - Grenoble I, September 2010, http://hal.inria.fr/tel-00543238/en.

[13] K. Varanasi. *Modélisation Spatio-Temporelle des scènes dynamiques 3D à partir de données visuelles*, Université de Grenoble, December 2010, http://hal.inria.fr/tel-00569147/en.

Articles in International Peer-Reviewed Journal

[14] P. Belhumeur, K. Ikeuchi, E. Prados, S. Soatto, P. Sturm. *Editorial for the Special Issue on Photometric Analysis for Computer Vision*, in "International Journal of Computer Vision", 2010, vol. 86, n^{o} 2-3, p. 125-126 [*DOI :* 10.1007/s11263-009-0292-3], http://springerlink.metapress.com/content/gr565g8821h22226/, http://hal.inria.fr/inria-00511432/en.

[15] J. Drareni, S. Roy, P. Sturm. *Plane-Based Calibration for Linear Cameras*, in "International Journal of Computer Vision", 2010 [*DOI :* 10.1007/s11263-010-0349-3], http://hal.inria.fr/inria-00523994/en.

[16] M. Hansard, R. Horaud. *Cyclorotation Models for Eyes and Cameras*, in "IEEE Transactions on Systems Man and Cybernetics B", 2010, http://hal.inria.fr/inria-00435549/en.

[17] W. Lee, W. Woo, E. Boyer. *Silhouette Segmentation in Multiple Views*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", October 2010, 14, http://hal.inria.fr/inria-00568915/en.

[18] B. Petit, J.-D. Lesage, M. Clément, J. Allard, J.-S. Franco, B. Raffin, E. Boyer, F. Faure. *Multicamera Real-Time 3D Modeling for Telepresence and Remote Collaboration*, in "International journal of digital multimedia broadcasting", 2010, Article ID 247108, 12 pages, http://hal.inria.fr/inria-00436467/en.

[19] S. Ramalingam, P. Sturm, S. Lodha. *Generic self-calibration of central cameras*, in "Computer Vision and Image Understanding", 2010, vol. 114, n^o 2, p. 210-219, http://hal.inria.fr/inria-00523989/en.

[20] D. Weinland, R. Ronfard, E. Boyer. *A survey of vision-based methods for action representation, segmentation and recognition*, in "Computer Vision and Image Understanding", 2010 [*DOI :* 10.1016/j.cviu.2010.10.002], http://hal.inria.fr/inria-00544635/en.

[21] 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)", 2010, vol. 86, n^o 2-3, p. 192-210 [*DOI :* 10.1007/s11263-009-0222-4], http://springerlink.metapress.com/content/m82880786g283vhr/, http://hal.inria.fr/inria-00266293/en.

International Peer-Reviewed Conference/Proceedings

[22] C. Cagniart, E. Boyer, S. Ilic. *Free-Form Mesh Tracking: a Patch-Based Approach*, in "IEEE Conference on Computer Vision and Pattern Recognition", San Francisco, États-Unis, 2010, -, http://hal.inria.fr/inria-00568909/en.

[23] C. Cagniart, E. Boyer, S. Ilic. *Iterative Deformable Surface Tracking in Multi-View Setups*, in "The Fifth International Symposium on 3D Data Processing, Visualization and Transmission", Paris, France, 2010, http://hal.inria.fr/inria-00568910/en.

[24] C. Cagniart, E. Boyer, S. Ilic. *Probabilistic Deformable Surface Tracking From Multiple Videos*, in "11th European Conference on Computer Vision", Heraklion, Grèce, 2010, http://hal.inria.fr/inria-00568912/en.

[25] J. Courchay, A. Dalalyan, R. Keriven, P. Sturm. *A global camera network calibration method with Linear Programming*, in "International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)", Paris, France, 2010, http://hal.inria.fr/inria-00523984/en.

[26] J. Courchay, A. Dalalyan, R. Keriven, P. Sturm. *Exploiting loops in the graph of trifocal tensors for calibrating a network of cameras*, in "European Conference on Computer Vision (ECCV)", Heraklion, Grèce, 2010, http://hal.inria.fr/inria-00523988/en.

[27] A. Delaunoy, E. Prados, P. Belhumeur. *Towards Full 3D Helmholtz Stereovision Algorithms*, in "Asian Conference on Computer Vision", Queenstown, Nouvelle-Zélande, Springer (editor), November 2010, http://hal.inria.fr/inria-00525867/en.

[28] J. DRARENI, S. ROY, P. STURM. *Bas-Relief Ambiguity Reduction in Shape from Shadowgrams*, in "International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)", Paris, France, 2010, http://hal.inria.fr/inria-00523985/en.

[29] M. DÍAZ, P. STURM. *Photometric Calibration using Multiples Images and a prior on Camera Response*, in "IEEE International Conference on Machine Vision", Hong Kong, Hong-Kong, 2010, http://hal.inria.fr/inria-00523993/en.

[30] L. GUAN, J.-S. FRANCO, E. BOYER, M. POLLEFEYS. *Probabilistic 3D Occupancy Flow with Latent Silhouette Cues*, in "IEEE Computer Vision and Pattern Recognition", San Francisco, États-Unis, June 2010, p. 1-8, http://hal.inria.fr/inria-00463031/en.

[31] Z. JANKO, A. DELAUNOY, E. PRADOS. *Colour Dynamic Photometric Stereo for Textured Surfaces*, in "Asian Conference on Computer Vision", Queenstown, Nouvelle-Zélande, SPRINGER (editor), November 2010, http://hal.inria.fr/inria-00525869/en.

[32] H. KAWASAKI, R. SAGAWA, Y. YAGI, R. FURUKAWA, N. ASADA, P. STURM. *One-shot scanning method using an uncalibrated projector and camera system*, in "IEEE International Workshop on Projector-Camera Systems", San Francisco, États-Unis, 2010, http://hal.inria.fr/inria-00523992/en.

[33] K. KULKARNI, E. BOYER, R. HORAUD. *An Unsupervised Framework for Action Recognition Using Actemes*, in "The Tenth Asian Conference on Computer Vision", Queenstown, Nouvelle-Zélande, 2010, http://hal.inria.fr/inria-00568906/en.

[34] A. LADIKOS, E. BOYER, N. NAVAB, S. ILIC. *Region Graphs for Organizing Image Collections*, in "Eccv 2010 Workshop on Reconstruction and Modeling of Large-Scale 3D Virtual Environments", Heraklion, Grèce, 2010, http://hal.inria.fr/inria-00568914/en.

[35] R. NARASIMHA, E. ARNAUD, F. FORBES, R. HORAUD. *Disparity and normal estimation through alternating maximization*, in "international conference on image processing (ICIP)", Honk Kong, Hong-Kong, 2010, http://hal.inria.fr/inria-00517864/en.

[36] R. PERRIER, E. ARNAUD, P. STURM, M. ORTNER. *Estimating satellite attitude from pushbroom sensors*, in "CVPR", San Francisco, États-Unis, June 2010, http://hal.inria.fr/inria-00514483/en.

[37] R. PERRIER, E. ARNAUD, P. STURM, M. ORTNER. *Image-Based Satellite Attitude Estimation*, in "IEEE International Geoscience and Remote Sensing Symposium", Honolulu, États-Unis, 2010, http://hal.inria.fr/inria-00523990/en.

[38] R. PERRIER, E. ARNAUD, P. STURM, M. ORTNER. *Satellite Image Registration for Attitude Estimation with a Constrained Polynomial Model*, in "ICIP", Hong Kong, France, September 2010, http://hal.inria.fr/inria-00514896/en.

[39] R. PERRIER, E. ARNAUD, P. STURM, M. ORTNER. *Sensor measurements and image registration fusion to retrieve variations of satellite attitude*, in "Asian Conference on Computer Vision", Queenstown, Nouvelle-Zélande, 2010, http://hal.inria.fr/inria-00523996/en.

[40] S. RAMALINGAM, S. BOUAZIZ, P. STURM, M. BRAND. *SKYLINE2GPS: Localization in Urban Canyons using Omni-Skylines*, in "IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)", Taipei, Taïwan, Province De Chine, 2010, http://hal.inria.fr/inria-00523997/en.

[41] A. SHARMA, R. HORAUD. *Shape Matching Based on Diffusion Embedding and on Mutual Isometric Consistency*, in "Workshop on Nonrigid Shape Analysis and Deformable Image Alignment (NORDIA)", San Francisco, États-Unis, IEEE, June 2010, http://hal.inria.fr/inria-00549406/en.

[42] A. SHARMA, E. VON LAVANTE, R. HORAUD. *Learning Shape Segmentation Using Constrained Spectral Clustering and Probabilistic Label Transfer*, in "European Conference on Computer Vision", Hiraklion, Grèce, Lecture Notes in Computer Science, September 2010, p. 743-756 [*DOI :* 10.1007/978-3-642-15555-0_54], http://hal.inria.fr/inria-00549401/en.

[43] C. UNGER, E. WAHL, P. STURM, S. ILIC. *Probabilistic Disparity Fusion for Real-Time Motion-Stereo*, in "Asian Conference on Computer Vision", Queenstown, Nouvelle-Zélande, 2010, http://hal.inria.fr/inria-00523995/en.

[44] K. VARANASI, E. BOYER. *Temporally Coherent Segmentation of 3D Reconstructions*, in "The Fifth International Symposium on 3D Data Processing, Visualization and Transmission", Paris, France, 2010, http://hal.inria.fr/inria-00568905/en.

[45] T. YAMAGUCHI, H. FUKUDA, R. FURUKAWA, H. KAWASAKI, P. STURM. *Video deblurring and super-resolution technique for multiple moving objects*, in "Asian Conference on Computer Vision", Queenstown, Nouvelle-Zélande, 2010, http://hal.inria.fr/inria-00523998/en.

National Peer-Reviewed Conference/Proceedings

[46] A. DELAUNOY, E. PRADOS, K. FUNDANA, A. HEYDEN. *Segmentation convexe multi-région de données sur les surfaces*, in "17ème Congrès de Reconnaissance des Formes et Intelligence Artificielle", Caen, France, January 2010, http://hal.inria.fr/inria-00526301/en.

[47] J.-S. FRANCO, L. GUAN, E. BOYER, M. POLLEFEYS. *Flot d'occupation 3D à partir de silhouettes latentes*, in "Reconnaissance de Forme et Intelligence Artificielle", Caen, France, January 2010, http://hal.inria.fr/inria-00463032/en.

[48] R. PERRIER, E. ARNAUD, P. STURM, M. ORTNER. *Estimation de l'attitude d'un satellite par recalage d'images*, in "COmpression et REprésentation des Signaux Audiovisuels (CORESA)", Lyon, France, 2010, http://hal.inria.fr/inria-00523987/en.

Workshops without Proceedings

[49] E. PRADOS, E. ARNAUD, P.-Y. LONGARETTI, F. MANCEBO. *Mathematical and numerical analyses of local integrated models*, in "Workshop on Decision Analysis and Sustainable Development", Montreal, Canada, September 2010, http://hal.inria.fr/inria-00526289/en.

[50] E. PRADOS, E. ARNAUD, P.-Y. LONGARETTI, F. MANCEBO. *Presentation of the SOCLE3 project and the STEEP lab.*, in "72nd meeting of the European Working Group "Multiple Criteria Decision Aiding"", Paris, France, October 2010, http://hal.inria.fr/inria-00526292/en.

Research Reports

[51] M. HANSARD, R. HORAUD. *Complex Cells and the Representation of Local Image-Structure*, INRIA, December 2010, n^{o} RR-7485, http://hal.inria.fr/inria-00546779/en.

[52] D. WEINLAND, R. RONFARD, E. BOYER. *A Survey of Vision-Based Methods for Action Representation, Segmentation and Recognition*, INRIA, February 2010, n^{o} RR-7212, http://hal.inria.fr/inria-00459653/en.

Scientific Popularization

[53] E. PRADOS, E. ARNAUD. *Modélisation numérique : Quel développement durable ?*, in "La Recherche. Les Cahiers de l'Inria", October 2010, n^{o} 445 octobre 2010, http://hal.inria.fr/inria-00537143/en.

Other Publications

[54] B. PETIT, T. DUPEUX, B. BOSSAVIT, J. LEGAUX, B. RAFFIN, E. MELIN, J.-S. FRANCO, I. ASSENMACHER, E. BOYER. *A 3D Data Intensive Tele-immersive Grid*, 2010, http://hal.inria.fr/hal-00514549/en.