INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team movi

Computational models for computer vision

Rhône-Alpes

THEME 3B

Activity Report

2003

Table of contents

1. Team

Head of project-team
Radu Horaud [INRIA]

Administrative assistant
Anne Pasteur

Research scientists
Frédéric Devernay [INRIA]
Rémi Ronfard [INRIA]
Peter Sturm [INRIA]

Faculty member
Edmond Boyer [Université Joseph Fourier]

PhD students
Ouideh Bentrah [in collaboration with IGN]
Adrien Bartoli
Thomas Bonfort
Guillaume Dewaele
Jean-Sébastien Franco
Pau Gargallo
Aude Jacquot
Frédérick Martin
Clément Menier
Marta Wilczkowiak

Technical staff
Matthieu Personnaz

Guest technical staff
Markus Michaelis [Plettac-Electronics GmBH, Germany]

2. Overall Objectives

The overall objective of the MOVI 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 – 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, tele-presence and immersive systems, image-based rendering and image-based animation;
- Autonomous robots and robot-based assistants;
- Embedded systems for car and driving technologies, portable devices, defense, space, etc.

3. Scientific Foundations

3.1. Image and video processing

The robust extraction of low-level features and intermediate-level primitives from images and video is a foundation for all higher-level tasks in computer vision. Important examples from a practical perspective are selection, tracking and correspondence (data association) of feature points over time or across views in multiple-view video; segmentation of motion into independent layers, background subtraction, accurate extraction of object silhouettes from video and the problem of efficient target detection, and efficient target recognition at unknown scales and orientations. From a computational perspective, it is useful to state all those related problems in a multi-resolution framework, but much work remains to be done in the domain of multi-resolution video. From a theoretical perspective, it is interesting to relate the selection of optimal features for a given task to recent developments in the statistical characterization of natural images, but this has yet to be extended to video, where optimal features have not yet been defined or employed.

3.2. 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 schemes for 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.

3.3. Surface reconstruction

Recovering surfaces from images is a fundamental task in computer vision. Applications are numerous and include, in particular, 3D modeling applications and mixed reality applications where reconstructed surfaces are mixed with virtual environments. The problem faced is to recover surface models, thus positions and differential properties such as surface normals, from image primitives such as points, segments and regions as well as photometric information. A tremendous research effort has been made in the past and several approaches have been proposed. These approaches can be grouped into categories depending on the primitives considered. Stereo-vision approaches match pixels in two images and recover local surface properties; silhouette based approaches separate foreground -the silhouettes- and background information in several images and approximate surface models by intersecting the viewing cones associated to the silhouettes; voxel approaches consider space discretizations into elementary 3D cells – the voxels – which are carved according to their geometric or photometric consistency when projected in several images. Independently of the methods, crucial issues still to be addressed in this field are: precision, robustness and rapidity since real time applications become feasible nowadays.

3.4. Statistical learning

In the last few years, there has been a tremendous effort to combine the methods of computer vision and information retrieval, so that images and videos may be indexed, searched and retrieved more efficiently. But it soon became clear that there is a semantic gap between the lower-level visual primitives that computer vision can recognize and the higher-level concepts that information retrieval can usefully index. A promising approach extends Bayesian, statistical methods such as Hidden Markov Models - which have been very successful in speech processing - to video. Thus, objects and events are represented as flexible structures of recognizable primitives. A key difficulty is the choice of primitives, both in space and in time, from a large number of possible low-level features. The analysis and description of human motion is one area where this issue can be resolved using built-in *geometric and kinematic models* of the human body. Yet, detecting people without prior knowledge remains an unresolved problem, which necessitates a combination of different low-level clues and high-level reasoning. Recently, a related, but much more general approach has emerged, where the recognition of objects in images is viewed as a *translation problem* between images and their natural language description. Thus, statistical translation models are learned from large collections of described images, as was done in the 1990's for collection of textual translations. Extending this approach to the recognition of events in video will be a challenging but promising new area of research.

3.5. Motion and gesture capture and analysis

Vision-based human motion capture, consists in recovering the motion parameters of a human body, such as limb position and joint angles, using camera images. It has a wide range of applications, and has gained much interest lately [53]. Recovering an *unconstrained* human motion is particularly challenging, mainly because of the high number of degrees of freedom, the small size of some part, and the multiple self-occlusions that may occur in images of the body of or part of it, such as the hand.

Once the motion parameters have been recovered, arises the problem of classifying and recognizing events, which are just objects evolving in time. Indeed, the categorization of events, especially human actions, raises difficult conceptual issues, e.g. what are the parts of an action, where do they begin and end, how do we build abstract representations of actions from examples, independently of their actors (objects or people). Yet, being able to perceive and recognize events, and to interpret them in terms of the actions of external agents (people, animals or even inanimate objects) is fundamental to many areas of visual cognition. An initial difficulty with this task is that there are of course many categories of actions. Actions may be divided into activities such as walking, each part of which is also a walking action; accomplishments such as jumping, which consists in a process and its culmination; or achievements such as winning a race which is instantaneous and results in a change of state. Therefore, the *temporal* segmentation of actions is a difficult problem, both for defining the classes and for recognizing their occurrences.

3.6. Distributed real-time systems

Recent advances in camera technologies have generalized real time acquisition of digital images, and nowadays, any modern PC can acquire such images at standard video rates. This allows complete acquisition systems to be built by simply connecting sets of digital cameras and PCs, without help from special-purpose material. The interest arises in various application domains where digital cameras are involved and where image information extracted in real time is required for interaction or control purposes. These domains include, for instance, scene virtualizations, video surveillances or human-machine interfaces. However, while much research has been devoted to algorithms that address the situation where a few cameras are connected to a single or few PCs, less efforts have been made toward larger sets of cameras and PCs which are, on the other hand, becoming standard. Moreover, most of the computer vision applications that make use of several cameras connected to several PCs do not take advantage of the available computing power and generally rely on a single PC for computations. These scalability and optimality issues are thus to be addressed, in particular by considering parallel strategies for 3D modeling computations.

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. Film and video indexing

Choosing terms for describing and indexing video content is a difficult and important problem. A very important source of high-quality video descriptions is the *continuity script* which describes very carefully the content of a movie shot by shot. We are developing new tools for segmenting and aligning individual shots in a movie with their script. We have now reached a stage where we maintain a large and quite unique set of 800 described shots. Using those shots as examples, our current work focuses on improving and refining the segmentation and alignment by learning statistical models of associations between the keywords in the script and the visual descriptors in the video. As a long-term research goal, we would like to generate such descriptions automatically by recognizing some of the places, actors and actions described in the script.

4.3. Human motion analysis

The analysis of human motion is difficult because of the deformable nature of the human body and of the large numbers of degrees of freedom. The scientific and technological challenge consists of using sophisticated body models and of matching these models with the visual data. The latter may be gathered with one or several static and/or moving cameras.

The capture and analysis of hand gestures is of great importance for understanding human behaviour and for interfacing computer systems with the physical and the biological world. Tracking the hands of a person in real-time is a difficult task. There is no way to predict the trajectories of the fingers and there are many self-occlusions.

4.4. Virtual videography

Virtual videography is the process of generating and assembling new views from existing video sequences. Live performances such as dance, theatre, music, and sports can be recorded from multiple views, and new views can be synthesized from angles different than the real ones, possibly with simulated camera motion - a camera choreography which would be impossible to plan or execute during the performance. This capability promises to be very useful for recording and celebrating cultural events, where it would have tremendous economical and social impacts.

The main difficulties to overcome are that the recording process must not interfere with the event being recorded - no body markers, no blue screens, etc. We are thus developing techniques for generating such synthetic views by combining geometric and kinematic models of human motion with statistical models of their spatial and temporal appearance in image sequences, using domain knowledge when applicable (sports, ballet).

4.5. Mixed realities

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.

4.6. The multiple-camera/PC-cluster experimental platform

An experimental platform for immersive applications has been set up recently at the INRIA Rhône-Alpes. This platform includes a set of cameras for model acquisitions, a set of projectors for the visualization and a PC cluster for computations. The challenging issue is to combine methods from different fields: computer vision, parallel processing and computer graphics, to get the complete process, from acquisition to visualization, working in real time. Computer vision modeling issues are therefore to be considered within this specific environment.

5. Software

5.1. Interactive scene modeling

We developed several algorithms for interactive scene modeling. These algorithms are based on silhouettes in images extracted using a background subtraction method. From these silhouettes and the cameras' geometry, we can compute the viewing cones which are the generalized cones of the viewing lines associated to every silhouette. The 3D intersection of these cones is a volume that necessarily encloses the object depicted by the silhouettes. Such volume is called the visual hull and represents an approximated model of the observed scene. Its precision depends naturally on the number of cameras considered. The algorithms we proposed implement new methods that allows for complex models which are exact with respect to the silhouettes, as well as real time modeling [38][39].

5.2. Point tracking in video sequences

A software for extracting and tracking interest points in video sequences has been developed. The software is based on standard computer vision techniques, and comes with a user-friendly GUI.

5.3. Dense stereo matching

The algorithms involved for a dense 3D reconstruction from two frames are highly dependent of the viewed scene. Furthermore, the user goals themselves should induce a specific choice. Indeed, the user focuses alternatively on the rendering, the running speed or the metrologic quality. We designed a software library [51] which allows to perform a dense matching between two images, by taking account both of the various user intents and world scenes. This leads first to the integration of several efficient and possibly complex dense matching algorithms. The multiplicity and the complexity of the algorithms should however not be at the expense of the ergonomy of the software system. A further goal is therefore to keep an easy and safe access to the library components. An iterative and object-oriented process is applied in order to cope with the design of a such library.

5.4. Camera calibration

We developed a system allowing to calibrate a camera from a special purpose calibration object. In particular, the focal length and the radial distortion parameters are estimated. Other functionalities such the acquisition of images, the stereo-calibration and 3D measures are also available. In detail, the software package includes the following features:

- Calibration: Non-linear estimation of the intrinsic parameters: focal length, optical center, skew parameter, radial distortion. Non-linear estimation of the relative position of the camera to the calibration object
- Stereo Calibration: Non-linear estimation of the parameters of a pair of cameras: intrinsic parameters and relative position of the two cameras. Calibration from several pairs of views of the calibration object.
- Metrology: 3D reconstruction by triangulation. Statistics on the 3D reconstruction accuracy.
- Images Acquisition: Acquisition from IC-RGB acquisition card. Acquisition from pgm image files.

The software package is downloadable at http://www.inrialpes.fr/movi/soft/calibration/index.html.

6. New Results

6.1. Motion panoramas

We developed a method for analysing video sequences and for representing them as mosaics or panoramas. Previous work on video mosaicking essentially concentrated on static scenes. We generalise these approaches to the case of a rotating camera observing both static and moving objects where the static portions of the scene are not necessarily dominant, as it has been often hypothesised in the past. We start by describing a robust technique for accurately aligning a large number of video frames under unknown camera rotations and camera settings. The alignment technique combines a feature-based method (initialisation and refinement) with rough motion segmentation followed by a colour-based direct method (final adjustment). This precise frame-to-frame alignment allows the dynamic building of a background representation as well as an efficient segmentation of each image such that moving regions of arbitrary shape and size are aligned with the static background. Thus a motion panorama visualises both dynamic and static scene elements in a geometrically consistent way. Extensive experiments applied to archived videos of track-and-field events validate the approach [49].

6.2. Hand tracking

A method was proposed to track the full hand motion from 3D points on the surface of the hand that were reconstructed and tracked using a stereoscopic set of cameras. This approach combines the advantages of previous methods that use 2D motion (e.g. optical flow), and those that use a 3D reconstruction at each time frame to capture the hand motion. Matching either contours or a 3D reconstruction against a 3D hand model is usually very difficult due to self-occlusions and the locally-cylindrical structure of each phalanx in the model, but our use of 3D point trajectories constrains the motion and overcomes these problems.

Our tracking procedure uses both the 3D point matches between two time frames and a smooth surface model of the hand. Realistic models are used in this procedure: the hand motion model uses animation techniques to represent faithfully the skin motion especially near joints, and the surface model is defined as an implicit surface. Robustness is obtained by using an EM version of the iterative closest point algorithm for matching points between consecutive frames, and the tracked points are then registered to the surface of the hand model. Results were obtained on a stereoscopic sequence of a moving hand, and were evaluated using a side view of the sequence.

6.3. Reconstruction of specular surfaces

We developed an approach for reconstructing specular surfaces using monocular image sequences [36][37]. The camera sees the reflection of an object in the specular surface, and the geometrical laws governing the combination of reflection and image projection allow to recover the shape of the specular surface. The approach uses a voxel-based representation of 3-space, and resembles voxel carving methods for opaque objects. Compared to previous approaches for specular surfaces, its main advantage is that no continuity assumptions are made.

6.4. Generalized camera model and calibration

We developed a generic calibration approach, allowing to calibrate all main types of cameras used in computer vision using a single algorithm [52]. The approach works for pinhole cameras as well as for cameras with arbitrary geometrical distortions (radial, tangential, etc.), catadioptric cameras, stereo systems, etc. We developed a theory for our approach and several algorithms, adapted to cameras with a single effective viewpoint ("central cameras") and to general non-central cameras (e.g. stereo systems). For both these cases, algorithms using 3D or planar calibration objects are available.

6.5. Camera self-calibration using planar objects

We have developed various practical methods for camera self-calibration using planar objects with otherwise unknown geometry or texture [41][42]. Closed-form solutions for several minimal scenarios as well as parameterizations for non-linear methods have been proposed. Our methods are especially suited to calibrate cameras from video sequences.

6.6. Scene modeling using geometrical constraints

Previously, we have proposed methods for camera calibration and 3D reconstruction that are based on geometrical constraints on the observed scene (parallelism, perpendicularity, co-planarity, etc.). This work has resulted in a complete system for interactive scene modeling, that allows to generate textured 3D models from a single or multiple images of a scene. Our system uses the available geometric constraints in an iterative procedure, automatically deciding at each step, which constraints to use to reconstruct additional primitives. Previous such systems are not robust to redundant constraints, which is why we developed an approach of detecting and handling this case [46][47], allowing the reconstruction procedure to converge more safely. The basic procedure does not compensate for errors that are accumulated during the sequential incorporation of constraint in the reconstruction process. Such an error compensation is ideally carried out by a bundle adjustment. In a cooperation with the COPRIN project-team of INRIA Sophia-Antipolis, we have developed methods for the automatic generation of minimal parameterizations of the scene model [45][48]. These incorporate all geometric constraints and allow thus an efficient bundle adjustment using standard unconstrained optimization techniques.

6.7. Scene modeling from image sequences

Various techniques for 3D scene modeling from image or video sequences were developed. A method for motion estimation has been proposed, that recovers all camera positions throughout a closed image sequence in one go, has been proposed [40]. Theoretical and practical considerations on the handling of gauge freedoms in bundle adjustment are given in [32]. Work on scene modeling from video sequences done by the consortium of the European project VISIRE (see below) is summarized in the overview paper [43].

6.8. Structure and motion recovery from line correspondences

Various methods for 3D reconstruction and motion estimation from line correspondences were developed [33][34]. They are based on theoretical results on the parameterization of motions of 3D lines that we established last year.

6.9. Aligning movies with their scenarios

In our previous work, we had obtained a structured representation of a movie script into meaningful entities such as scenes, shots, actions and dialogs. This year, we investigated automatic shot-to-shot alignment of video and script. With our new method, we obtained an incomplete, but roughly correct alignment, which was manually corrected and organized into a database of 800 shots from an entire movie [44]. We demonstrated the system with an *advanced DVD player* which connects to our remote server for browsing and searching the shots based on actors, places and actions, and plays the corresponding segments from a local DVD. This demo was shown at the IEEE Multimedia Conference and Expo in July in Baltimore, USA, and the “Fête de la Science” in October in Grenoble.

6.10. Real-time surface reconstruction

As mentioned earlier, we have proposed several new methods to compute 3D models -visual hulls- of objects from their extracted silhouettes in several images. Such methods overcome several limitations of previous methods which are: precision of the models and their computation time. In order to implement these methods in a real context, and in particular within the experimental platform GRIMAGE, we have collaborated with the APACHE INRIA team and worked on distributed implementations of these methods. This has resulted in real time implementations which, associated to distributed visualization algorithms, allow to visualize 3D models of scenes, typically persons, merged with virtual environments in real time. Our main contribution with respect to existing systems is a distributed and thus scalable system which acquires and displays real and virtual models in real-time.

6.11. Multiple-camera acquisition and synchronization

The simultaneous acquisition and processing of multiple synchronized video streams requires data transfer rates that are far above the specifications of a standard workstation. An alternative is to use a distributed system, where each node in a grid handles one or two cameras with different viewpoints. However, our methods require an accurate synchronization (typically, up to a few hundred microseconds) of each set of simultaneous images in the video streams. The classical solution consists in adding a secondary network which is dedicated to propagating external synchronization signals, but this solution doesn’t scale up easily when the grid is composed of dozens of workstations and cameras. Our solution consists in using the existing network to synchronize very accurately CPU clocks, and to use real-time system components based on Linux to generate the synchronization pulses at the right time. This solution is highly versatile and configurable, and can handle for example several subsets of synchronized cameras with different periods.

6.12. Video Surveillance

We contributed to the establishment of a computational model of visual attention useful for video surveillance applications. Attention models encapsulate both bottom-up and top-down processes such as: selection of a visual event of interest, detection of image features associated with this event, mechanisms for maintaining these features in the visual field of view, and recognition and interpretation. We developed a system which performs visual attention with two cameras: A static one and an active one, the latter being mounted onto a pan-tilt device. The static camera performs event detection in its associated low-resolution image. The active camera tracks the event over time and provides an accurate high-resolution description of the event. The method has been applied to the problem of traffic monitoring and of detecting and tracking pedestrians and cyclists.

8. Other Grants and Activities

8.1. National initiatives

8.1.1. GdR ISIS.

MOVI has a “Projet Jeunes Chercheurs” entitled CALIPSOO with IRIT (Institut de Recherche en Informatique, Toulouse), under the auspice of Gdr ISIS. The collaboration is about camera self-calibration from planar scenes.

8.1.2. ROBEA.

MOVI participates in the project PARKNAV, in the framework of the ROBEA programme. The other partners are CYBERMOVE and PRIMA (INRIA Rhône-Alpes), VISTA (IRISA Rennes) and RIA (LAAS, Toulouse). The project is about the interpretation of complex dynamic scenes and reactive motion planing in such scenes.

8.1.3. ACI.

MOVI collaborates with the Imagis project-team (now, ARTIS project-team) on the project CYBER (ACI Jeunes Chercheurs). Its main objective is the real-time immersion of a person, inside a virtual environment. Applications aimed at are related to TV, video games and theme parks. A research platform was set up, consisting of cameras, professional lighting and computers. Among others, work was done on calibration of the multi-camera system and real-time 3D reconstruction using the system.

Late 2003, the project CYBER-II started (ACI Masses de Données), with an extended consortium (ARTIS, MOVI, APACHE, LIRIS Lyon). Research on various topics will be carried out. Concerning MOVI, this will concern real-time 3D reconstruction, recovery of surface reflectance properties, and virtual relighting of scenes.

8.2. Projects funded by the European Commission

8.2.1. Visire.

The project IST-1999-10756 VISIRE (Virtual Image-Processing System for Intelligent Reconstruction of 3D Environments) was done in collaboration with Eptron S.A. (Spain, coordinator), Giunti Multimedia (Italy), Lund University (Sweden) and the Polytechnical University of Madrid (Spain). It started in may 2000 and finished in October 2003. The project concerned both fundamental research and the development of a prototype for 3D reconstruction from video sequences acquired by a hand-held camcorder. MOVI worked on camera self-calibration, interactive 3D modeling using scene constraints, 3D mesh generation and issues related to texture mapping.

8.2.2. Events.

It is an FP5-IST project which started in 2000 and ended in 2003. The objective of this project is to use computer vision techniques in order to simulate a virtual camera. Consider for example the following scenario. Two cameras film an event such as a football game from two different positions. These two camera may well be viewed as a stereo pair and therefore a disparity map can be obtained. One may also associated the image texture with each depth value. A new camera can be simulated such that its position is a virtual one, somewhere in between the two real cameras. Using the disparity map and the image textures, the challenge is to produce a realistic video sequence as viewed by the simulated camera. This project was done in partnership with EPTRON Ltd. and Via Digital Ltd. (Spain), Siemens Ltd. (Germany) as well as with the University of Oxford and the University of Paderborn (Germany).

9. Dissemination

9.1. Editorial boards and programme committees

- *International Journal of Robotics Research*: Radu Horaud is a member of the editorial board. He co-edited a special issue on “Visual Analysis of Human Motion”, June 2003.
- *Computer Vision and Image Understanding* (R. Horaud is *area editor*)
- Radu Horaud is a member of the SPECIF award committee for the period 2003-2005
- Radu Horaud and Peter Sturm are members of the programme committee of the IEEE Conference on Computer Vision and Pattern Recognition, Wisconsin, June 2003.
- Radu Horaud is member of the programme committees of the International Conference on Advanced Robotics (Coimbra, Portugal, June-July 2003) and of the IEEE International Conference on Robotics and Automation (Taiwan, September 2003).

- Peter Sturm is member of the programme committees of ICCV (International Conference on Computer Vision), ICIP (IEEE International Conference on Image Processing), OMNIVIS (Workshop on Omnidirectional Vision and Camera Networks), SACV (IEEE Workshop on Statistical Analysis in Computer Vision), AmbInt (Workshop on Ambient Intelligence) and ICASSP (International Conference on Acoustics, Speech and Signal Processing).

9.2. Teaching

- Optimisation, DEA IVR, INPG, 6h, P. Sturm.
- Vision 3D, DEA IVR, INPG, 12h, P. Sturm.
- Vision stéréoscopique, Mastère photogrammétrie Numérique, ENSG, 14h, F. Devernay.
- Synthèse d'images, MAGISTÈRE INFORMATIQUE, UNIV. JOSEPH FOURIER, RICM, ISTG, 100h, E. Boyer
- Analyse d'images, DESS INFORMATIQUE, UNIV. JOSEPH FOURIER, 30H, E. Boyer.
- Géométrie projective, DEA IVR, INPG, 6h, E. Boyer.

9.3. Invited talks

- Radu Horaud gave an invited talk at Siemens Corporate Research, USA, July 2003.
- Rémi Ronfard gave an invited talk to the Aristote Foundation at École Polytechnique, Palaiseau, January 2003.
- Peter Sturm gave an invited talk at Malmö Högskola, Sweden, April 2003.
- Peter Sturm gave an invited talk on "Calibration and Orientation for Omnidirectional Vision" at the Workshop on New Developments in Close Range Photogrammetry, Bonn, Germany, March 2003.
- Peter Sturm gave a tutorial on self-calibration at ORASIS (Journées Jeunes Chercheurs), Gérardmer, France, May 2003.

10. Bibliography

Major publications by the team in recent years

[1] S. CHRISTY, R. HORAUD. *Euclidean Shape and Motion from Multiple Perspective Views by Affine Iterations.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », number 11, volume 18, November, 1996, pages 1098-1104, ftp://ftp.inrialpes.fr/pub/movi/publications/rec-affiter-long.ps.gz.

[2] P. GROS, O. BOURNEZ, E. BOYER. *Using Local Planar Geometric Invariants to Match and Model Images of Line Segments.* in « Computer Vision and Image Understanding », number 2, volume 69, 1998, pages 135-155.

[3] R. HARTLEY, P. STURM. *Triangulation.* in « Computer Vision and Image Understanding », number 2, volume 68, 1997, pages 146-157.

[4] R. HORAUD, G. CSURKA, D. DEMIRDJIAN. *Stereo Calibration from Rigid Motions.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », number 12, volume 22, December, 2000, pages 1446-1452, ftp://ftp.inrialpes.fr/pub/movi/publications/HoraudCsurkaDemirdjian-pami2000.ps.gz.

[5] R. HORAUD, F. DORNAIKA, B. ESPIAU. *Visually Guided Object Grasping.* in « IEEE Transactions on Robotics and Automation », number 4, volume 14, August, 1998, pages 525-532.

[6] R. HORAUD, F. DORNAIKA. *Hand-Eye Calibration.* in « International Journal of Robotics Research », number 3, volume 14, June, 1995, pages 195-210.

[7] R. HORAUD, F. DORNAIKA, B. LAMIROY, S. CHRISTY. *Object Pose: The Link between Weak Perspective, Paraperspective, and Full Perspective.* in « International Journal of Computer Vision », number 2, volume 22, March, 1997, pages 173-189.

[8] R. HORAUD, O. MONGA. *Vision par ordinateur: outils fondamentaux.* Editions Hermès, Paris, 1995, *Deuxième édition revue et augmentée*.

[9] L. HÉRAULT, R. HORAUD. *Figure-ground discrimination: a combinatorial optimization approach.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », number 9, volume 15, September, 1993, pages 899-914.

[10] 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://www.inrialpes.fr/movi/publi/Publications/2001/LBP01.

[11] K. MIKOLAJCZYK, C. SCHMID. *Indexing based on scale invariant interest points.* in « Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada », pages 525-531, 2001, http://www.inrialpes.fr/movi/publi/Publications/2001/MS01.

[12] R. MOHR, B. BOUFAMA, P. BRAND. *Understanding Positioning from Multiple Images.* in « Artificial Intelligence », volume 78, 1995, pages 213-238.

[13] S. PETITJEAN, E. BOYER. *Regular and Non-Regular Point Sets: Properties and Reconstruction.* in « Computational Geometry - Theory and Application », number 2-3, volume 19, 2001, pages 101-126, http://www.inrialpes.fr/movi/publi/Publications/2001/PB01.

[14] L. QUAN, T. KANADE. *Affine Structure from Line Correspondences with Uncalibrated Affine Cameras.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », number 8, volume 19, août, 1997, pages 834-845.

[15] L. QUAN. *Invariants of Six Points and Projective Reconstruction from Three Uncalibrated Images.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », number 1, volume 17, janvier, 1995, pages 34-46.

[16] L. QUAN. *Conic Reconstruction and Correspondence from Two Views.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », number 2, volume 18, février, 1996, pages 151-160.

[17] L. QUAN. *Self-Calibration of An Affine Camera from Multiple Views.* in « International Journal of Computer Vision », number 1, volume 19, mai, 1996, pages 93-105.

[18] A. RUF, R. HORAUD. *Visual Servoing of Robot Manipulators, Part I: Projective Kinematics.* in « International Journal of Robotics Research », number 11, volume 18, November, 1999, pages 1101-1118, http://www.inria.fr/rrrt/rr-3670.html.

[19] C. SCHMID, R. MOHR, C. BAUCKHAGE. *Evaluation of Interest Point Detectors.* in « International Journal of Computer Vision », number 2, volume 37, 2000, pages 151-172, http://www.inrialpes.fr/movi/publi/Publications/2000/SMB00.

[20] C. SCHMID, R. MOHR. *Object Recognition Using Local Characterization and Semi-Local Constraints.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », number 5, volume 19, mai, 1997, pages 530-534.

[21] 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 », pages 432-437, June, 1999.

[22] 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 », pages 1100-1105, Juin, 1997.

[23] P. STURM, B. TRIGGS. *A Factorization Based Algorithm for Multi-Image Projective Structure and Motion.* in « Proceedings of the 4th European Conference on Computer Vision, Cambridge, England », pages 709-720, Avril, 1996.

[24] B. TRIGGS. *Matching Constraints and the Joint Image.* in « IEEE Int. Conf. Computer Vision », E. GRIMSON, editor, pages 338-43, Cambridge, MA, June, 1995.

[25] B. TRIGGS. *Autocalibration and the Absolute Quadric.* in « IEEE Conf. Computer Vision & Pattern Recognition », Puerto Rico, 1997.

Doctoral dissertations and “Habilitation” theses

[26] A. BARTOLI. *Reconstruction et alignement en vision 3D : points, droites, plans et caméras.* Ph. D. Thesis, INPG - Institut National Polytechnique de Grenoble, Grenoble, France, September, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/Bar03b.

Articles in referred journals and book chapters

[27] J. BARRETO, F. MARTIN, R. HORAUD. *Visual Servoing/Tracking Using Central Catadioptric Images.* series Springer Tracts in Advanced Robotics 5, Springer Verlag, 2003, chapter VI, pages 245–254, http://www.inrialpes.fr/movi/publi/Publications/2003/BMH03.

[28] A. BARTOLI, P. STURM. *Constrained Structure and Motion From Multiple Uncalibrated Views of a Piecewise Planar Scene.* in « IJCV - International Journal of Computer Vision », number 1, volume 52, April, 2003, pages 45–64, http://www.inrialpes.fr/movi/publi/Publications/2003/BS03c.

[29] Y. DUFOURNAUD, C. SCHMID, R. HORAUD. *Image Matching with Scale Adjustment.* in « Computer Vision and Image Understanding », 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/DSH03.

[30] G. SHECHTER, F. DEVERNAY, E. COSTE-MANIÈRE, A. QUYYUMI, E. MCVEIGH. *Three-Dimensional Motion Tracking of Coronary Arteries in Biplane Cineangiogram.* in « IEEE Transaction on Medical Imaging », number 4, volume 22, April, 2003, pages 493–503, http://www.inrialpes.fr/movi/publi/Publications/2003/SDCQM03.

[31] C. SMINCHISESCU, B. TRIGGS. *Estimating Articulated Human Motion with Covariance Scaled Sampling.* in « International Journal of Robotics Research », 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/ST03.

Publications in Conferences and Workshops

[32] A. BARTOLI. *Towards Gauge Invariant Bundle Adjustment: A Solution Based on Gauge Dependent Damping.* in « ICCV'03 - In Proceedings of the IEEE International Conference on Computer Vision, Nice, France », IEEE CSP, October, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/Bar03.

[33] A. BARTOLI, R. HARTLEY, F. KAHL. *Motion From 3D Line Correspondences: Linear and Non-Linear Solutions.* in « CVPR'03 - In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA », IEEE CSP, pages 477–484, June, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/BHK03.

[34] A. BARTOLI, P. STURM. *Multiple-View Structure and Motion From Line Correspondences.* in « ICCV'03 - In Proceedings of the IEEE International Conference on Computer Vision, Nice, France », IEEE CSP, October, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/BS03b.

[35] C. BLONDEL, G. MALANDAIN, R. VAILLANT, F. DEVERNAY, E. COSTE-MANIÈRE, N. AYACHE. *4D Tomographic Representation of Coronary Arteries From One Rotational X-ray Sequence.* in « Medical Image Computing and Computer-Assisted Intervention (MICCAI2003) », series Lecture Notes in Computer Sciences, Springer-Verlag, November, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/BMVDCA03.

[36] T. BONFORT, P. STURM. *Reconstruction de surfaces réflechissantes à partir d'images.* in « ORASIS », pages 59-65, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/BS03d, Gérardmer, France.

[37] T. BONFORT, P. STURM. *Voxel Carving for Specular Surfaces.* in « Proceedings of the 9th IEEE International Conference on Computer Vision », IEEE CSP, October, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/bS03.

[38] E. BOYER, J.-S. FRANCO. *A Hybrid Approach for Computing Visual Hulls of Complex Objects.* in « Computer Vision and Pattern Recognition », pages 695-701, June, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/BF03, Madison, Wisconsin, USA.

[39] J.-S. FRANCO, E. BOYER. *Exact Polyhedral Visual Hulls.* in « Fourteenth British Machine Vision Conference (BMVC) », pages 329–338, September, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/FB03, Norwich, UK.

[40] N. GUILBERT, A. BARTOLI. *Batch Recovery of Multiple Views with Missing Data using Direct Sparse*

Solvers. in « BMVC'03 - In Proceedings of the Fourteenth British Machine Vision Conference, Norwich, UK », September, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/GB03.

[41] P. GURDJOS, P. STURM. *Methods and Geometry for Plane-Based Self-Calibration.* in « IEEE Conference on Computer Vision and Pattern Recognition », IEEE, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/GS03, Madison, Wisconsin.

[42] E. NOIRFALISE, F. JURIE, P. STURM. *Autocalibration itérative de caméras à partir de structures planes.* in « ORASIS », pages 157-164, 2003, Gérardmer, France.

[43] T. RODRIGUEZ, P. STURM, M. WILCZKOWIAK, A. BARTOLI, M. PERSONNAZ, N. GUILBERT, F. KAHL, M. JOHANSSON, A. HEYDEN, J. MENENDEZ, RONDA, JAUREGUIZAR. *VISIRE - Photorealistic 3D Reconstruction from Video Sequences.* in « International Conference on Image Processing », IEEE, pages 705-708, September, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/RSWBPGKJHMRJ03, Barcelona, Spain.

[44] R. RONFARD, T. TRAN-THUONG. *A framework for aligning and indexing movies with their script.* in « IEEE International Conference on Multimedia and Expo », pages 21-24, July, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/RT03, Baltimore, USA.

[45] G. TROMBETTONI, M. WILCZKOWIAK. *Scene Reconstruction Based on Constraints: Details on the Equation System Decomposition.* in « Proceedings of the 9th International Conference on Principles and Practice of Constraint Programming (CP'03) », September, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/TW03, Kinsale, Ireland.

[46] M. WILCZKOWIAK, P. STURM, E. BOYER. *The Analysis of Ambiguous Solutions in Linear Systems and its Application to Computer Vision.* in « Proceedings of the 14th BMVC », volume 1, pages 53–62, September, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/WSB03, Norwich, UK.

[47] M. WILCZKOWIAK, P. STURM, E. BOYER. *Une approche pour l'analyse et exploitation des contraintes géométriques en modélisation 3D.* in « ORASIS », pages 77–86, May, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/WSB03b, Gérardmer, France.

[48] M. WILCZKOWIAK, G. TROMBETTONI, C. JERMANN, P. STURM, E. BOYER. *Scene Modeling Based on Constraint System Decomposition Techniques.* in « Proceedings of the 9th International Conference on Computer Vision », IEEE, IEEE, October, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/WTJSB03.

Internal Reports

[49] A. BARTOLI, N. DALAL, R. HORAUD. *Motion Panoramas.* Technical report, number RR-4771, INRIA, March, 2003, http://www.inria.fr/rrrt/rr-4771.html.

[50] O. CHUM, T. PAJDLA, P. STURM. *On Geometric Error for Homographies.* Technical report, number CTU-CMP-20, Czech Technical University, Prague, 2003, http://www.inrialpes.fr/movi/publi/Publications/2003/CPS03.

[51] M. PERSONNAZ, P. STURM, F. DEVERNAY. *Design of a Library for Dense Matching.* Technical report, number RT-0278, INRIA, April, 2003, http://www.inria.fr/rrrt/rt-0278.html.

[52] P. STURM, S. RAMALINGAM. *A Generic Calibration Concept – Theory and Algorithms.* Technical report, number 5058, INRIA, December, http://www.inria.fr/rrrt/rr-5058.html.

Bibliography in notes

[53] T. MOESLUND, E. GRANUM. *A survey of computer vision-based human motion capture.* in « Computer Vision and Image Understanding », number 3, volume 81, 2001, pages 231-268.