Project-Team movi

Computational models for computer vision

Rhône-Alpes

THEME COG

Activity

Report

2005

Table of contents

1. Team

Research director

Radu Horaud [Research Director (DR) INRIA]

Administrative assistant

Anne Pasteur

Research scientists

Frédéric Devernay [Research Associate (CR) INRIA]

Emmanuel Prados [Research Associate (CR) INRIA]

Rémi Ronfard [Research Associate (CR) INRIA]

Peter Sturm [Research Associate (CR) INRIA]

Faculty member

Edmond Boyer [Université Joseph Fourier]

Ph.D students

Ouideh Bentrah [in collaboration with IGN]

Thomas Bonfort [INRIA grant]

Guillaume Dewaele [ENS PhD student]

Jean-Sébastien Franco [MESR grant]

Pau Gargallo [MESR grant]

Frédéric Huguet [MESR grant]

Aude Jacquot [CIFRE funding with Thales]

David Knossow [INRIA grant]

Diana Mateus [Marie-Curie grant]

Clément Menier [ENS PhD student]

Julien Morat [CIFRE funding with Renault]

Shrikumar Ramalingam [Marie-Curie grant, co-supervision agreement with U California]

Daniel Weinland [Marie-Curie Grant]

Andrei Zaharescu [Marie-Curie grant]

Post-doctoral fellows

Dana Cobzaş [NSERC grant (Canada)]

Kenneth Sundaraj [Singapore]

Technical staff

Hervé Mathieu [IR, INRIA]

Florian Geffray

Bertrand Holveck

Marc Lapierre

Loïc Lefort

Philippe Martin

Elise Taillant

2. Overall Objectives

2.1. 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 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 car and driving technologies, portable devices, defense, space, 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. Multiple-camera acquisition of visual data

Modern computer vision techniques and applications require the deployment of a large number of cameras linked to a powerful multi-PC computing platform. Therefore, such a system must fulfill the following requirements: The cameras must be synchronized up to the millisecond, the bandwidth associated with

image transfer (from the sensor to the computer memory) must be large enough to allow the transmission of uncompressed images at video rates, and the computing units must be able to dynamically store the data and/to process them in real-time.

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

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

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 have 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 but 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 are matched against an articulated object model (based on cylindrical parts, implicit surfaces, conical parts, and so forth). The matching is carried out iteratively using various methods, such as ICP (iterative close point) or EM (expectation/maximization).

Challenging problems are the detection of motion and motion tracking. When a vision systems observes complex articulated motion, such as the motion of the hands, 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 estimated the motions associated with each individual degree of freedom.

3.5. Video Processing

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.

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.

Another important domain is the interaction with multi-media databases through advanced multimodal interfaces. In order to archive and manage multimedia visual material such as news, social and cultural events, movies, theater and music performances, etc., it is necessary to extract and store information concerning its content in addition to its mere recording. This implies that a system is able to perform automatic analysis of visual information available with video sequences. Generally speaking, modern audio-visual systems for understanding, classifying, archiving, and accessing multimedia databases must encapsulate the following features: (a) shot detection and classification, (b) recognition of individuals (actors, players, athletes, ...), (c) recognition of facial expressions, and (d) action and gesture recognition.

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 MOVI 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.

5. Software

5.1. Platforms

5.1.1. *The Grimage platform*

The Grimage platform is a room dedicated to video acquisition, 3D computations and visualisation. It is an experimental platform for several research teams of the INRIA Rhone-Alpes including MOVI, MOAIS, ARTIS and EVASION. Many specific software developments have been realized for this platform using National or European fundings.

5.1.2. *The MV platform*

The MV platform is a mobile platform composed of 4-6 video cameras and a small PC cluster including a laptop for visualization. It allows real-time visualization of dynamic textured models of real scenes. This platform has been developed by the MOVI team for the European Project Holonics.

5.2. Camera calibration

5.2.1. MVCamera: Multi-camera calibration

We developed software for calibrating large numbers of cameras with minimal user intervention. The calibration object is very simple, consisting of a rigid rod holding active or passive markers. The calibration process consists of three steps – acquiring a short sequence of the calibration object moving in the field of the cameras, detecting and tracking the markers automatically in all views, and simultaneously reconstructing the 3D trajectories of the markers and the camera parameters. The new software includes a graphical user interface (GUI) to allow even non-expert users to perform and validate a calibration. In the current version, user intervention is strictly limited to initializing the camera positions and orientations in a global reference coordinate system. The software will be used internally in the GRIMAGE room and also delivered to some of our industry partners.

5.2.2. Mono and stereo 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 as 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://perception.inrialpes.fr/Soft/calibration/index.html.

5.3. Blinky: Real-time image acquisition from multiple cameras

The Blinky software library aims at real-time acquisition of images for multiple cameras spread over a PC cluster or a computing grid. The library contains tools to develop two kinds of software components:

- A *frontend* is directly connected to the camera driver and is in charge of doing the image acquisition. Blinky makes the images available either to the local host by using shared memory or to other hosts by networking, allowing images to be captured transparently accross a local network. Actually, the frontend can be seen as an image server, which can also be used to change the camera parameters (shutter speed, aperture, zoom...). The frontend is also able to record the raw video stream to a file, which can later be used as if it were a live camera.

- A *backend* is the user application that captures the images and processes them. Each camera is designated by a device name which is valid accross the network, and acquiring an image is as simple as reading a file. Multiple backends can connect to a single frontend, allowing for many applications to run at the same time using the same cameras. A single backend can also connect to multiple cameras, and a single function call is necessary to acquire a set of synchronized images from multiple cameras spread over the network.

The blinky distribution also contains a set of sample frontends and backends:

- *blinkyf1394*, a frontend for IIDC-compliant IEEE1394 cameras on Linux.

- *blinkyflnrd*, a frontend for CameraLink cameras, using the Arvoo Leonardo frame grabber on Linux.
- *blinkyfdummy*, a dummy frontend serving predictable images.
- *blinkybenchmark*, a backend to test the acquisition rate (FPS) of a camera.
- *blinkysaveimages*, a backend that images aquired for one or several cameras to any standard image format.
- *blinkysdl*, a backend that visualizes the video stream using the SDL graphical library.

The library itself is fully POSIX-compliant, and was ported to several variants of Unix and to Windows 2000/XP. As soon as this software is out of its beta-stage, it will be distributed with an open-source license.

5.4. MVVideo: Multiple-camera video player and recorder

We developed a new software with a graphical user interface (GUI) to remotely control the acquisition, synchronization and display of video streams from multiple cameras (up to 16 cameras) arbitrarily distributed on a grid of workstations. The software can be used as a recorder, in which case it controls all camera parameters (e.g. gain, shutter speed, focus, white balance) through the IEEE 1394 interface. The software allows the hardware configuration to be dynamically modified, and the control of either hardware or software synchronization of all cameras. The software controls remote processes which are responsible for recording the video streams on disk with frame rates up to 30 frames per second and resolutions up to 800 x 600. The software displays the video streams from all cameras simultaneously in reduced resolution during recording.

The software can also be used as a video player, allowing the synchonized display of all recorded video streams simultaneously, as well as limited video editing functions (including transcoding, trimming, bookmarking and transfer over networks).

The software makes it possible even for non-experts to successfully record and play-back multiple-view video. It will be used internally in the GRIMAGE room and also delivered to some of our academic and industry partners. It will also serve as a framework for developing other real-time video processing modules.

5.5. 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.

5.6. Real-time shape acquisition and visualization

We developed a complete model acquisition chain, from camera acquisition to model visualization. Modeling algorithms are silhouette based and produce an approximated model of the scene, the visual hull, in real-time. The software was deployed on the GRIMAGE platform which is composed of up to 10 firewire cameras, 20 PCs and 16 projectors for visualization. Its implementation was made in collaboration with the MOAIS INRIA team and includes today: image acquisition (described before), background subtraction to extract silhouettes, 3D modeling on PC clusters to compute visual hulls, 3D rendering and 3D display using several projectors for high screen resolutions.

5.7. Stereo-based 3D reconstruction

A versatile software library for stereo reconstruction was developed. This library contains a number of state-of-the-art algorithms, and it helps in prototyping advanced stereo algorithms, such as those dealing with occlusions and low-textured areas. Besides, two real-time implementations of standard stereo algorithms were developed: A parallel version to be run on a PC cluster, and a version using the GPU (Graphics Processing Unit). All versions can run from live cameras using the Blinky acquisition library, making them usable for real-time applications.

5.8. Visual-Hull library (EPVH)

We have developed a library that allows visual hulls to be computed given a set of silhouettes and the associated camera calibration. Complexity is quasi-optimal allowing therefore real-time computation of 3D models. This library is the state-of-art in the field and has already been used by several research teams in the world.

5.9. Human-body modelling for motion capture and tracking

We have developed a human-body modeller that takes input from videos and which interactively allows an user to associate 3D points with rotational joints. Based on these point-to-joint assignments, a person is modelled with 19 body parts and 48 rotational joints. Each body part is modelled by a truncated elliptical cone.

5.10. Dynamic texture reconstruction and visualization

We have developed a software which displays sequences of 3D textured dynamic models. This software is dedicated to multiple video context. It takes as input a sequence of 3D models over time and the associated projective textures, or videos, from several viewpoints. The output is a textured 3D models that can be seen interactively from any viewpoint, and at any time in the sequence including continous playing. To our knowledge, this is the first public software for such input data.

5.11. Cyclope: A 6-DOF optical tracker based on a single camera

Cyclope is a 6 degrees of freedom tracker using a single camera, an infrared flash, and a tracking device equipped with retro-reflective markers. It is robust and portable, making it a good candidate for virtual reality applications. Cyclope was demonstrated at SIGGRAPH'2005.

6. New Results

6.1. Generic Camera Models

The motivation of the works in this section is to propose generic approaches for camera calibration and structure-from-motion, that would work for any type of camera. To do so, we first adopt a simple but general camera model that consists in representing one projection ray in 3D per image pixel. This model comprises most camera types used in computer vision, e.g. pinhole (with or without radial or tangential distortion), catadioptric cameras (central or non-central), pushbroom, non-central mosaics, multi-camera systems, etc.

6.1.1. Calibration of generic camera models

In [76] we proposed a generic calibration algorithm for our general camera model that requires three images of a calibration grid, taken from unknown positions. Four different algorithms are proposed, working for 3D or planar calibration grids and for fully non-central or central cameras.

This first work only allowed to calibrate regions of the image where the calibration grids' projections overlap. In [58], [69] we then showed how to calibrate the whole camera, by integrating multiple images of calibration grids.

In [60], we address the calibration of an intermediate case between central (single viewpoint) and fully non-central camera models: so-called axial cameras are non-central, but there exists a line in 3D that cuts all projection rays. This is the case for pushbroom cameras, mis-aligned catadioptric cameras and possibly for fisheye lenses.

6.1.2. Self-calibration of generic camera models

We recently proposed a first self-calibration approach for the general camera model introduced in the previous paragraph. This is a rather difficult problem and our first contribution thus considers the special case

of central cameras (but with otherwise unconstrained projection rays) and relies on specific camera motions (pure translation or rotations) [59].

6.1.3. Unification of structure-from-motion using generic camera models

In [62], [37] we propose a unified treatment of the structure-from-motion problem by giving generic algorithms for the basic tasks of pose estimation, motion estimation and 3D point triangulation. Like our calibration approach (see above), these algorithms are applicable to whatever camera type, the only distinction that may have to be made is between central and non-central cameras.

6.1.4. Multi-view geometry for generic camera models

In [63], [38] we present a multi-view geometry for our general camera model, embodied by multi-view tensors and matching constraints. Like with perspective cameras, constraints for up to 4 views at a time exist. The obtained matching tensors operate on Plücker coordinates of projection rays rather than homogeneous coordinates of image points as in the classical case, and are thus of dimension $6 \times 6 \times \cdots$.

6.1.5. Calibration of cameras with general radial distortion

This work addresses a camera model lying between the classical parametric models and the general non-parametric model used in the previous paragraphs: cameras with arbitrary radial distortion. This model is still sufficiently general for most typical cameras, including pinhole cameras with moderate distortions, most catadioptric cameras and fisheyes. In [64] we propose calibration algorithms that use a planar calibration grid (a flat screen is used in practice). The algorithms can already work with a single image and with only a few images, stable results were achieved. Examples of distortion corrections using the calibration results are shown in figure 1.
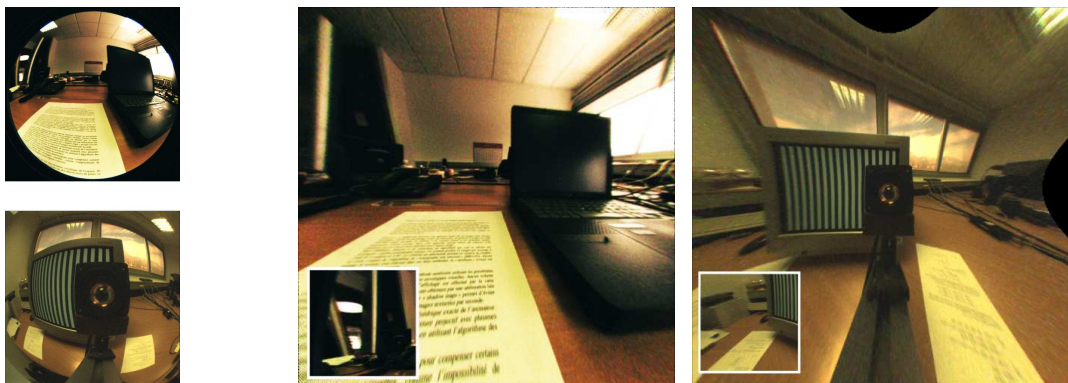
Figure 1. Image rectification examples obtained with calibration of general radial distortion. Left: Original images (fisheye lens and home-made catadioptric camera, built using a camera and a spherical christmas ornament). Middle: Rectified image for the fisheye. Right: Rectified image for the home-made catadioptric camera. Small inset images show rectification of the border regions.

6.2. Calibration

6.2.1. Calibration with linear objects

In [49], we propose closed-form solutions for camera calibration using images of a linear object, e.g. a stick with markers attached ("wand"). Further, we reveal the critical, or degenerate, motions that cause the calibration to fail and in the vicinity of which the calibration will be unstable.

6.2.2. Calibration of Multi-Camera Systems

We propose and evaluate several approaches for providing initial solutions for multi-camera calibration [57], [56], that can then be refined by bundle adjustment. Calibration is done using one, two or three rigidly coupled wands (stick with markers attached) and exploiting the available geometrical constraints in different manners.

6.2.3. Camera calibration from silhouettes

Calibrating a multiple camera environment using image silhouettes is an interesting alternative to usually fastidious standard calibration procedures which require specific patterns: LED or calibration sticks for instance. A person moving in the acquisition space is then sufficient to update camera parameters. We have studied why, and how, a set of image silhouettes give constraints on such parameters, and we have derived from that a practical approach for the calibration of sets of camera [44].

6.2.4. Camera cooperation through the coupling of epipolar geometry and kinematics

Many vision tasks need rotating and zooming cameras. We investigated the case of a fixed panoramic camera that is coupled to a pan-tilt-zoom camera and we revealed the equations governing the link between the epipolar geometry and the kinematics of the pan-tilt mechanisms. We developed a practical method that is well suited for visual surveillance systems that use camera cooperation. This work was done by Radu Horaud, David Knossow, and Markus Michaelis (Plettac Electronics) [31].

6.3. Representation, registration, and tracking of rigid and articulated shapes

6.3.1. Markerless human motion capture

One of our ambitious research projects is to solve the problem of human motion capture from several cameras and without using body markers. Whenever a 3D object is described with a set of smooth surfaces, it projects onto the image as an extremal contour. We developed a parameterization of such a contour point as a function of the kinematic description of an articulated chain. This has been applied to the problem of tracking human motion from its observed image contours. Cylindrical elliptic cones are smooth surfaces used to model body parts. They project onto images as straight lines, whose motion can be obtained from body part motions in closed-form. The Jacobian of this parameterization is used within an efficient and general tracker, figure 2 and [52]. The work is part of David Knossow's PhD thesis under the supervision of Remi Ronfard, and with the collaboration of Radu Horaud and Frederic Devernay.
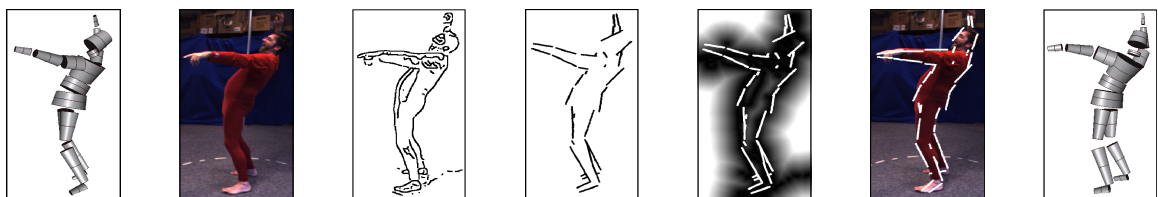
Figure 2. From left to right: The current model is matched against a new image. The contours extracted from this image are compared with the extremal contours predicted from the model using the chamfer-distance image. Finally, the newly estimated model is consistent with this image.

6.3.2. Matching 3D points with articulated and deformable shapes

We developed a method for matching 3D points (e.g. stereo data) with articulated shapes. The latter are represented by an articulated implicit surface that models local shape deformations such as bending. The method was applied to the problem of tracking the motions of a hand with 27 degrees of freedom. The tracker

can accomodate with bad data and with situations where the shape rapidly turns away from the viewer, figure 3. The work was carried out within the PhD of Guillaume Dewaele under the supervision of Radu Horaud and Frederic Devernay [72] [27].
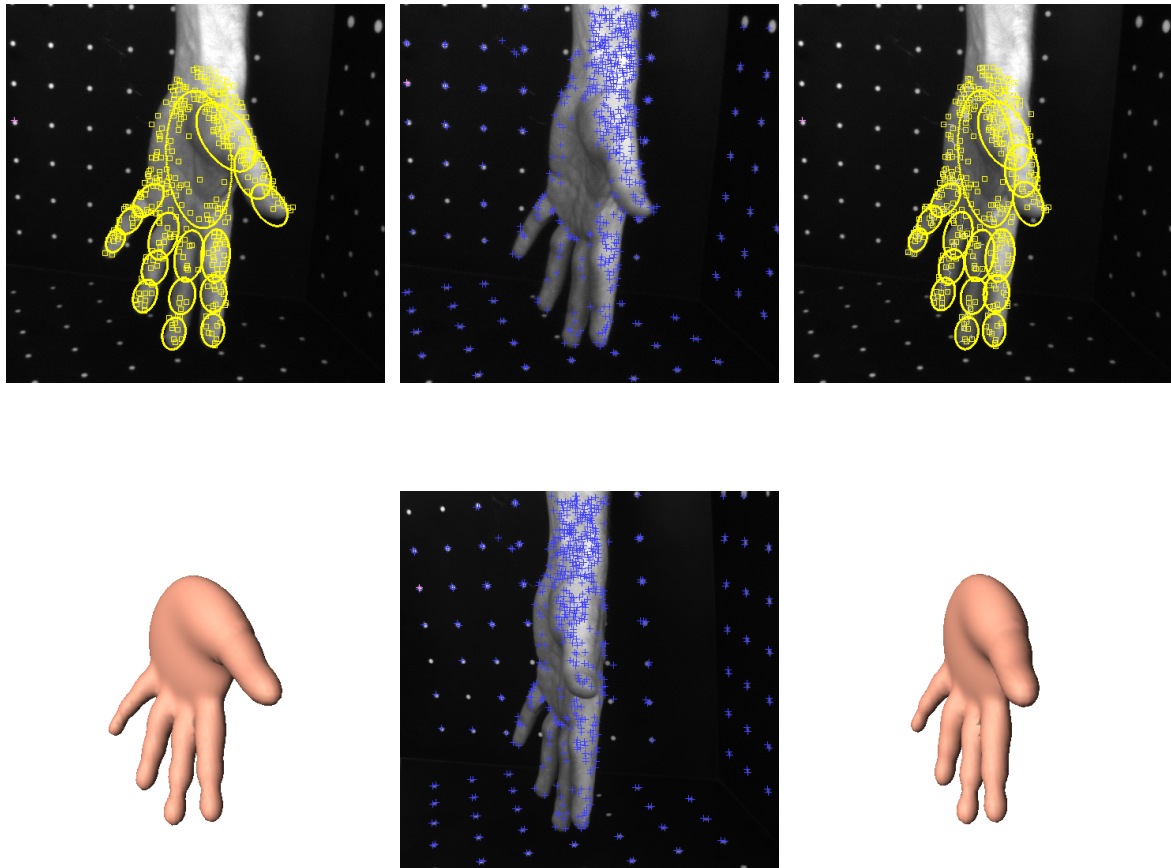
Figure 3. This figure illustrates the method that we developed for matching 3D points with articulated and deformable objects such as a hand. Left: The articulated model (texture and surface) shown in its previously estimated pose, Middle: a set of 3D data points obtained by stereo, and Right: the result of aligning the data (middle) with the model (left).

6.3.3. *Motion analysis of 3D deformable surfaces*

Motion analysis methods in Computer Vision usually deal with tracking rigid, articulated or deformable objects in a scene, but little or no work has been done on the analysis of 3D deformable surface, such as cloth, the surface of a liquid, or the surface of a plastic solid subject to various deformations. With the recent availability of high-resolution and high-dynamic cameras, it is now possible to perceive very subtle changes in the images, and bring up information about the 3D deformations of these surfaces, especially when several viewpoints of the surface are available. There are a few methods specialized in the reconstruction of continuous surfaces from several viewpoints, which take into account constraints on the regularity of the observed surface, such as PDE-based stereo methods. However, 3D reconstruction data of a surface at each time instant does not contain information about the 3D motion of each individual point on the surface. What we get is a 2D manifold in a 3D space (i.e. a surface reconstruction), whereas we also may want a 3D vector field (the motion

field) defined on that manifold. A testbed application for these new methods we are developping is the ACI Masse de données GEOLSTEREO.

6.3.4. *Robust matching of 3D point sets*

The problem of matching 3D point sets is a central one in computer vision. We developed a method that consider this problem as a point-classification problem: Data points are assigned to one or several model points when the latter are viewed as classes of points. In practice the mathing is carried out using a robust EM algorithm. The work was carried out by Guillaume Dewaele, Radu Horaud, Frederic Devernay, and Florence Forbes (MISTIS group), [27] and a paper was submitted to ECCV'06.

6.3.5. *Articulated motion capture from from 3D points and normals*

We modeled human bodies with articulated implicit surfaces and we developed a method for finding the pose parameters of this shape from 3D data. The latter are obtained from multiple silhouettes using visual-hull extraction and embeds both positions and orientations. We defined a new data-to-implicit-surface sum-of-square distance that is minimized over the pose parameters [55].

6.3.6. *Building shape/kinematic models for human-body tracking and motion capture*

One important ingredient of any motion capture and tracking method is the description of the human body. We developed a method and a software package that considers a kinematic representation with 19 body parts and 54 degrees of freedom, as well as a representation of each body-part as a 3D shape, i.e. a truncated elliptical cone. The fixed parameters of both the kinematic model and the 3D shapes are estimated using an interactive procedure as well as bundle adjustment techniques. The work is carried out by David Knossow, Loic Lefort, and Remi Ronfard.

6.3.7. *Action recognition*

Action recognition is a new research topic for MOVI. We are investigating the use of volumetric reconstructions from multiple cameras for the purpose of recognizing primitive actions such as crossing arms, waiving, standing up, sitting down, etc. To that end, we derived motion descriptors based on the cylindric Fourier decomposition of a new representation called the *Motion History Volume* [65]. The descriptors were shown to be useful for classifying and segmenting human action. This is a promising avenue of research which we will continue to explore in 2006.

6.4. 3D Modeling

6.4.1. *Bayesian 3D Modeling from Images using Multiple Depth Maps*

In [47], [48] we propose a Bayesian approach for dense 3D surface modelling from calibrated input images. The surface is represented by the union of multiple depth maps, one per image, and associated color maps for the estimated surface color. Occlusions are handled by hidden visibility variables, telling how likely it is that some 3D points are visible or not in some input image. A prior on 3D surfaces that allows strong discontinuities, is used, and the likelihood of the estimated model is based on photoconsistency. The MAP (maximum a posteriori) estimate of the 3D surface is achieved using an EM algorithm or direct gradient descent. Good results have been obtained for scenes with different degrees of difficulty.

6.4.2. *Multi-view reconstruction of specular surfaces*

We have developed two main approaches for reconstructing specular surfaces by acquiring images of reflections of a calibration object. In [70], it was required that both, the camera and calibration object, be placed at different known positions. A voxel coloring type approach was then used to reconstruct points on the specular surface and the associated normals.

In [43], [61], we considerably relaxed the requirements; in the developed approach the camera remains static and only the calibration object is put into different, *unknown* positions. We developed different approaches for recovering these positions; once they are known, points on the specular surface are determined by triangulation. In practice, a flat screen is used as calibration object and due to displaying series of appropriate
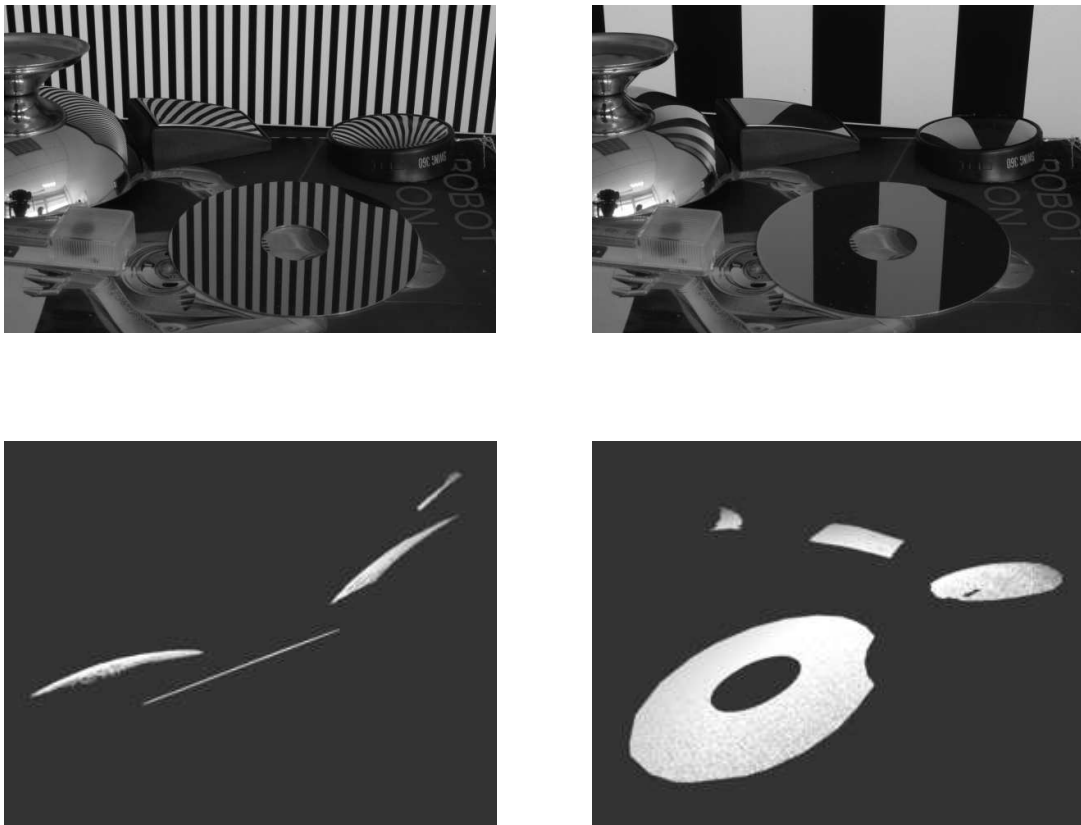
Figure 4. Example for reconstruction of specular surfaces. The top row shows two of the images used for the reconstruction. Notice the 3 curved mirrors (an ice-cream cup and two small wide-angle rear-view mirrors) and a planar hard drive platter. The second row shows the reconstruction viewed from two locations. The model contains over 525 000 independent points. Note the planarity of the reconstructed hard drive platter in the left image. 99.9% of the reconstructed points were less than 0.1 mm away from the best-fit plane, and 88% less than 0.05 mm.

gray codes, dense point correspondences and thus dense reconstructions of the specular surface, are achieved. An example is shown in figure 4.

6.5. Shape from multiple silhouettes

Assume we are given several image silhouettes of an object corresponding to different camera viewpoints. The visual hull is the maximal solid shape consistent with the object silhouettes. Such an approximation of the object captures all the geometric information available from the silhouettes. Visual hulls are often used to provide robust 3D shape models in various modeling applications. In the past, we have studied how to efficiently compute them and we have proposed quasi-exact and exact solutions allowing to model complex scenes, along with an optimal complexity in time and space [71], [73].

6.5.1. *Real-time and distributed algorithms for visual-hull computation*

Some applications such as tele-immersion require real-time modeling to allow for interactions between real and virtual worlds. We have developed real-time solutions for visual hull computations. These solutions are based on both parallelization and distribution schemes over a cluster of PCs, and are therefore scalable. Our current implementation allows models to be computed at 30 frames per second using 8-10 cameras [74], [75].

6.5.2. *Fusion of multiple-view silhouette cues*

Many computer vision applications are based on silhouettes extracted using *background subtraction* methods. These methods are usually monocular and do not account for possibly redundant information in several cameras. As a results, they are very sensitive to noise that pertubs the background model. One way to improve robustness of silhouette extraction in multiple camera environments is to integrate all the 2D image silhouette probabilities in a 3D probabilistic grid. Fusing the information from several images allows errors which are unconsistent over the image set to be removed. We have developed such a method in Bayesian framework [46].

7. Contracts and Grants with Industry

7.1. Thalès-Optronics SA

Target tracking with a single camera for air-ground missile guidance. 36 months (2003-2006). 30,000 euros and the salary of a PhD student (Aude Jacquot).

TOSA develops semi-automatic and automatic tools for air-to-ground missile guidance systems. Combat aircrafts are equiped with pan-tilt-yaw infrared cameras coupled with a laser beam, and the pilot has the tasks first, to designate a target onto this image and second, to keep this target within the field of view of the camera. The difficulties are associated with both finding a small target in a low-resolution image and in keeping track of this target in the presence of large aircraft motions (rotations, forward translation, etc.). When the aircraft flies at 300 meters/second, the target first appears as a blob and after only a few tenth of a second it changes its appearance to become a complex 3D structure.

The scientific and technological objectives are the following: to develop tools enabling (i) off-line modelling of complex 3D man-made objects from a collection of images (geometric modelling as well as the aspect of the objects, such as color and texture), (ii) automatic identification of the target, and automatic tracking of the target such that large changes are taken into account.

7.2. Renault SA

Detection and classification of objects which are ahead of a vehicle. 36 months (2004-2007). 50,000 euros and the salary of a PhD student (Julien Morat).

In June 2004 we started a 3 year collaboration with the French car manufacturer Renault SA (Direction de la Recherche). Within this collaboration Renault co-funds a PhD thesis with ANRT. The topic of the collaboration and of the thesis is the detection and classification of obstacles which are ahead of a vehicle. We

currently develop a prototype system based on *stereoscopic vision* with the following functionalities: low speed following, pre-crash, and pedestrian detection. In particular we study the robustness of the image processing algorithms with respect to camera/stereo calibration problems (the system should be able to self-detect such problems).

8. Other Grants and Activities

8.1. National initiatives

8.1.1. RNTL-OCETRE

OCETRE is a 2-year exploratory RNTL project granted by the French Ministry of Research. The project started in January 2004. The scientific goal of the project is to develop methods and techniques for recovering, in real-time, the geometry of a complex scene such as a scene composed of both static and dynamic objects (for example, people moving around). We will combine methods based on stereo with methods based on visual hull reconstruction from silhouettes. One original contribution of the project will be to combine dense depth data (gathered with stereo) with visual hulls.

We develop a camera setup composed of one color camera and two black-and-white camera. This three cameras are linked to a PC and they deliver synchronized videos at 30 frames per second. Moreover, several such setups will be deployed and synchronized using a PC cluster.

The industrial collaborators (Total-Immersion SA and Thales Training and Simulation SA) are interested to develop real-time augmented reality applications using our methods. Since the moving objects are reconstructed in real-time, it will be possible to treat them as graphical objects and therefore mix real and virtual objects in a realistic manner, i.e., in 3D space thus taking into account their interactions and mutual occlusions, and not in image space as is currently done by many augmented-reality systems.

8.1.2. RIAM–SEMOCAP

Along with Université de Rennes, MOVI is one of two scientific contributors to the SEMOCAP project funded by the CNC and Ministère de la Recherche et de l'Industrie as part of the RIAM network (Recherche et Innovation en Audiovisuel et Multimedia) which was started in January 2004 for two years. The goal of the project is to build a low-cost system for human motion capture without markers, using multi-view video analysis and biomechanical motion models. As part of the project, we have built a prototype motion capture system using the GRIMAGE infrastructure, which has been demonstrated for the first time in December in Rennes. Evaluation is being performed against ground truth data collected using a marker-based Vicon system.

8.1.3. ROBEA–ParkNav

MOVI participates in the project PARKNAV, in the framework of the ROBEA program. The other partners are eMOTION 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.

8.1.4. ACI–Cyber II

MOVI is part of the project CYBER-II (ACI Masses de Données), with an extended consortium (ARTIS, MOVI, APACHE, LIRIS Lyon). Started late in 2003 CYBER-II follows the project CYBER (ACI Jeunes Chercheurs) initiated in 2000. 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.

8.1.5. ACI–GEOLSTEREO

In September 2004 we started a 3 year collaboration with the Géosciences Azur laboratory (UMR 6526). This colloboration received funding from the French Minister of research through the *ACI Masses de données* program (Action Concertée Incitative).

Mathematical modelling as well as simulation and visualization tools are widely used in order to understand, predict, and manage geological phenomena. These simulation tools cannot fully take into account the complexitiy of the natural catastrophes such as surface earthquakes occuring at level of the sedimentary layer, land slidings, etc. Within this project we plan to study and develop measurement methods based on computer vision techniques. The physical model consists in a mock-up of the geological object to be studied. Existing techniques allow to reproduce, at the mock-up scale, the influence of several hundreds years of the Earth gravitational field. We plan to observe such a mock-up with a high-resolution stereoscopic camera pair and to apply dense stereo reconstruction techniques in order to study the 3D deformations over time. In particular, the expected accuracy of the planned measurements is of the order of 10μm which corresponds to an actual amplitude of a few centimeters.

8.1.6. ARC–Georep

This ARC is concerned with the representation of 3D objects which plays a central role in various domains such as computer graphics or computer vision. Different disciplines use different representations and conversions between these representations appears to be a challenging issue with an impact over a wide class of disciplines. To reach this goal, this ARC connects participants having skills in various disciplines (3D acquisition, 3D reconstruction, Digital Geometry Processing, Numerical Analysis and Computer Graphics). The MOVI team is concerned with the acquisition and reconstruction part of the project.

8.1.7. SEE-NUSH

In December 2005, Remi Ronfard signed an agreement with Grenoble Alpes Incubation on a spin-off project, nicknamed SEE-NUSH, with plans to create a startup company in 2006-07 to industrialize and commercialize the multiple-camera software developed by MOVI into products implementing virtual cameras and virtual stages. The targeted applications are in 3D cinematography for sports, performing arts, medecine, games and motion pictures.

8.2. Projects funded by the European Commission

8.2.1. FP6-IST STREP project Holonics

Holonics is a European 3-year project which started on September 1, 2004. We have three industrial partners: EPTRON, coordinator (Spain), Holografika (Hungary), and Total-Immersion (France). The general scientific and technological challenge of the project is to achieve realistic virtual representations of humans through two complimentary technologies: (i) multi-camera based acquisition of human data and of human actions and gestures, and (ii) visualization of these complex representations using modern 3D holographic display devices.

Our team will develop a real-time multi-camera and multi-PC system. The developments will be based on 3D reconstruction methods based on silhouettes and on visual hulls as well as on human-motion capture methods and action and gesture recognition.

8.2.2. FP6/Marie-Curie EST Visitor

Visitor is a 4 year European project (2004-2008) under the Marie-Curie actions for young researcher mobility – Early State Training or EST. Within these actions, the GRAVIR laboratory has been selected to host PhD students granted by the European commission. The MOVI team, which is part of the GRAVIR laboratory, actively participated in the project elaboration. The MOVI team is currently coordinating this project and hosts two PhD students from this program.

8.2.3. FP6/Marie-Curie RTN VISIONTRAIN

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

VISIONTRAIN addresses the problem of understanding vision from both computational and cognitive points of view. The research approach will be based on formal mathematical models and on the thorough

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

8.3. Associate Team

In January, we started an associate team with Brown University (Providence, Rhode Island, USA) on the theme of 3D cinematography. 3D cinematography, also sometimes called free-viewpoint video, is the process of building 3D models of dynamic scenes from multiple video camera inputs. Our teams at INRIA and Brown University have separately developed methods and tools for 3D photography, but moving to 3D cinematography raises difficult technological and scientific challenges, where we want to collaborate more actively. In this new application, the processing pipeline must be fully automated, so that 3D reconstruction can be performed at video frame rates. Implementing fast and robust vision algorithms using inputs from multiple streams is a difficult subject in itself. Generating new viewpoints from a limited number of cameras is another issue. Most existing methods necessitate the storage of the video streams from all cameras in full resolution, because they interpolate the new viewpoint from all available viewpoints at the same time frame. This is not a realistic scenario, because high quality can only be obtained by increasing the number of cameras. If we can reconstruct deforming surfaces with consistent texture coordinates over time, we can instead generate viewpoints by interpolating over multiple time frames. The goal of the associate team is to contribute to that effort by joint work in two crucial areas – real-time video processing for large networks of camera systems, using distributed/parallel and embedded architectures and real-time mesh processing for reconstruction, deformation, texturing and view interpolation of 3D objects from multiple video, using multi-resolution models amenable to efficient, scalable compression.

We had one week-long start-up meeting in Grenoble with Gabriel Taubin and Chad Jenkins in Grenoble in June, and several follow-up meetings in Providence over the summer and fall of 2005 to set up multiple collaborations between our researchers and PhD students. We have many active joint projects for 2006. In particular, Gabriel Taubin and Remi Ronfard will co-organize an international workshop on 3D cinematography in New York City on June 22, 2006, which will be attended by all associate team members.

9. Dissemination

9.1. Editorial boards and program committees

- Radu Horaud is a member of the editorial boards of the *International Journal of Robotics Research* and of the *International Journal of Computer Vision*, as well as *area editor* of *Computer Vision and Image Understanding*

- Peter Sturm has co-organized the ISPRS/IEEE Workshop Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images (BenCOS), held in conjunction with ICCV, Beijing, China, October 2005.

- Peter Sturm was member of the Program Committees of:

  – ICCV'05 (International Conference on Computer Vision, Beijing),

- CVPR'05 (IEEE International Conference on Computer Vision and Pattern Recognition, San Diego, USA),
- ICIP'05 (IEEE International Conference on Image Processing, Genova, Italy),
- BMVC'05 (British Machine Vision Conference, Oxford, UK),
- DV'05 (Workshop on Dynamical Models for Computer Vision, held in conjunction with ICCV, Beijing, China),
- OMNIVIS'05 (Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, held in conjunction with ICCV, Beijing, China),
- ORASIS'05 (Journées Jeunes Chercheurs, Fournol, France)
- and a reviewer for SIGGRAPH'05.

- Edmond Boyer was Area Chair of the British Machine Vision Conference conference.
- Edmond Boyer was member of the Program Committees of: ICCV'05, BMVC'05, ICIP'05, ICASSP'05 and ORASIS'05.
- Remi Ronfard was co-organizer of the IEEE workshop on Modeling People and Human Interaction (PHI'05), Beijing, October 15, 2005.
- Remi Ronfard was reviewer for the Siggraph 2005 and ICIP 2005 conferences.

9.2. Services to the Scientific Community

- Radu Horaud is in charge of European coordination at INRIA Rhône-Alpes.
- Radu Horaud is a member of the SPECIF award committee for the period 2003-2005.
- Peter Sturm is Co-Chairman of the Working Group "Image Orientation" of the ISPRS (International Society for Photogrammetry and Remote Sensing), for the period 2004-2008.
- Peter Sturm is a member of the INRIA Committee on "Actions Incitatives" (part of COST – Conseil d'Orientation Scientifique et Technologique).
- Remi Ronfard is member of the "Commission de Specialistes" for recruitments at the University Joseph Fourier of Grenoble.
- Edmond Boyer is member of the IMAG (Intitut d'Informatique et de Mathématiques Appliquées de Grenoble) Scientific Committe.
- Edmond Boyer is member of the "Commission de specialistes" for recruitments at the University Joseph Fourier of Grenoble and at the Institut National Polytechnique de Grenoble.
- Edmond Boyer is coordinator of the Marie-Curie Visitor Project and member of the Visitor Scientific Committe.
- Radu Horaud is the coordinator of the Visiontrain Marie Curie Research Training Network.

9.3. Teaching

- 3D Computer Vision, postgraduate course, University of Zaragoza, Spain, 20h, P. Sturm.
- Analyse d'images, DESS INFORMATIQUE, UNIV. JOSEPH FOURIER, 30H, R. Ronfard.
- Optimisation, M2R IVR, INPG, 6h, P. Sturm.
- Vision 3D, M2R IVR, INPG, 12h, P. Sturm.
- Géométrie projective, M2R IVR, INPG, 6h, E. Boyer.
- Stereoscopic Perception, Mastère Photogrammétrie Numérique, ENSG, Marne-la-Vallée, 17h, F. Devernay.
- Computer Vision, Mastère 2 Pro IICAO, Université Joseph Fourier, Grenoble, 24h, F. Devernay.

9.4. Invited talks

- Peter Sturm gave invited talks at:

  – CREA Lab, Amiens, France, November 2005.
  – National Lab for Pattern Recognition, Beijing, China, October 2005.
  – National Technical University Athens, Greece, April 2005.

- Remi Ronfard gave an invited talk at Brown University, July 2005, on *Human motion capture and action recognition for 3D cinematography*.
- Remi Ronfard, Herve Mathieu and Florian Geffray presented a live demonstration of 3D cinematography to 30 students of the Ecole Nationale Superieure des Arts du Theatre (Lyon, December 1).

9.5. Thesis

- Jean-Sébastien Franco and Guillaume Dewaele defend their PhD theses in December 2005.
- Peter Sturm acted as reviewer for the following PhD theses:

  – John Mallon, Dublin City University, Ireland.
  – Diego Ortín Trasobares, University of Zaragoza, Spain.

- Peter Sturm acted as examiner for the following theses:

  – Srikumar Ramalingam, University of California, Santa Cruz (qualifying examination).
  – Neil Birkbeck, University of Alberta, Canada, (MSc examination).
  – Sio-Hoï Ieng, Université Pierre et Marie Curie, Paris (PhD).

- The following PhD students did internships with Peter Sturm:

  – Kiyoung Kim, PhD student at Gwangju Institute of Science and Technology (South Korea), 2005/06 (6 months).
  – Lazaros Grammatikopoulos, PhD student at National Technical University Athens (Greece), 2005 (10 days).
  – Diego Aguilera, PhD student at Universidad de Salamanca (Spain), 2005 (3 months).
  – Jean-Philippe Tardif, PhD student at Université de Montréal, 2005

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", vol. 18, nº 11, November 1996, p. 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", vol. 69, nº 2, 1998, p. 135-155.

[3] R. HARTLEY, P. STURM. *Triangulation*, in "Computer Vision and Image Understanding", vol. 68, nº 2, 1997, p. 146-157.

[4] R. HORAUD, G. CSURKA, D. DEMIRDJIAN. *Stereo Calibration from Rigid Motions*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 22, nº 12, December 2000, p. 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", vol. 14, nº 4, August 1998, p. 525-532.

[6] R. HORAUD, F. DORNAIKA. *Hand-Eye Calibration*, in "International Journal of Robotics Research", vol. 14, nº 3, June 1995, p. 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", vol. 22, nº 2, March 1997, p. 173-189.

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

[9] L. HÉRAULT, R. HORAUD. *Figure-ground discrimination: a combinatorial optimization approach*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 15, nº 9, September 1993, p. 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://perception.inrialpes.fr/publication.php3?bibtex=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", 2001, p. 525-531, http://perception.inrialpes.fr/publication.php3?bibtex=MS01a.

[12] R. MOHR, B. BOUFAMA, P. BRAND. *Understanding Positioning from Multiple Images*, in "Artificial Intelligence", vol. 78, 1995, p. 213-238.

[13] S. PETITJEAN, E. BOYER. *Regular and Non-Regular Point Sets: Properties and Reconstruction*, in "Computational Geometry - Theory and Application", vol. 19, nº 2-3, 2001, p. 101-126, http://perception.inrialpes.fr/publication.php3?bibtex=PB01.

[14] L. QUAN, T. KANADE. *Affine Structure from Line Correspondences with Uncalibrated Affine Cameras*, in "ieee Transactions on Pattern Analysis and Machine Intelligence", vol. 19, nº 8, août 1997, p. 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", vol. 17, nº 1, janvier 1995, p. 34-46.

[16] L. QUAN. *Conic Reconstruction and Correspondence from Two Views*, in "ieee Transactions on Pattern Analysis and Machine Intelligence", vol. 18, nº 2, février 1996, p. 151-160.

[17] L. QUAN. *Self-Calibration of An Affine Camera from Multiple Views*, in "International Journal of Computer Vision", vol. 19, nº 1, mai 1996, p. 93-105.

[18] A. RUF, R. HORAUD. *Visual Servoing of Robot Manipulators, Part I: Projective Kinematics*, in "International Journal of Robotics Research", vol. 18, nº 11, November 1999, p. 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", vol. 37, nº 2, 2000, p. 151-172, http://perception.inrialpes.fr/publication.php3?bibtex=SMB00.

[20] C. SCHMID, R. MOHR. *Object Recognition Using Local Characterization and Semi-Local Constraints*, in "ieee Transactions on Pattern Analysis and Machine Intelligence", vol. 19, nº 5, mai 1997, p. 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", June 1999, p. 432-437.

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

[23] 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.

[24] 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", Avril 1996, p. 709-720.

[25] B. TRIGGS. *Matching Constraints and the Joint Image*, in "IEEE Int. Conf. Computer Vision, Cambridge, MA", E. GRIMSON (editor). , June 1995, p. 338-43.

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

Doctoral dissertations and Habilitation theses

[27] G. DEWAELE. *Modelisation, suivi et simulation d'objets articules et deformables. Application au modelage reel d'une argile virtuelle*, Ph. D. Thesis, Institut National Polytechnique de Grenoble, Grenoble, France, December 2005, http://perception.inrialpes.fr/Publications/2005/Dew05.

[28] J.-S. FRANCO. *Modélisation 3D à partir de silhouettes images*, Ph. D. Thesis, Institut National Polytechnique de Grenoble, Grenoble, France, December 2005, http://perception.inrialpes.fr/Publications/2005/Fra05.

Articles in refereed journals and book chapters

[29] A. BARTOLI, P. STURM. *Structure From Motion Using Lines: Representation, Triangulation and Bundle Adjustment*, in "Computer Vision and Image Understanding", 2005, http://perception.inrialpes.fr/Publications/2005/BS05.

[30] O. CHUM, T. PAJDLA, P. STURM. *The Geometric Error for Homographies*, in "Computer Vision and Image Understanding", vol. 97, n° 1, January 2005, p. 86-102, http://perception.inrialpes.fr/Publications/2005/CPS05.

[31] R. HORAUD, D. KNOSSOW, M. MICHAELIS. *Camera Cooperation for Achieving Visual Attention*, in "Machine Vision and Applications", 2005, http://perception.inrialpes.fr/Publications/2005/HKM05.

[32] I. MARTIN, R. RONFARD, F. BERNARDINI. *Detail-Preserving Variational Surface Design With Multiresolution Constraints*, in "Journal of Computing and Information Science in Engineering", vol. 5, n° 2, June 2005, p. 104–110, http://perception.inrialpes.fr/Publications/2005/MRB05.

[33] K. MASSAMUNE, A. DEGUET, H. MATHIEU, R. H. TAYLOR, E. BALOGH, J. ZINREICH, L. FAYAD, G. FISCHER, G. FICHTINGER. *Image Overlay Guidance for Needle Insertion in CT Scanner*, in "IEEE Transaction on Biomedical Engineering", vol. 52, n° 8, August 2005, p. 1415–1424, http://perception.inrialpes.fr/Publications/2005/MDMTBZFFF05.

[34] T. RODRIGUEZ, I. REID, R. HORAUD, N. DALAL, M. GOETZ. *Image Interpolation for Virtual Sports Scenarios*, in "Machine Vision and Applications", vol. 16, n° 4, 2005, http://perception.inrialpes.fr/Publications/2005/RRHDG05.

[35] T. RODRIGUEZ, P. STURM, P. GARGALLO, N. GUILBERT, A. HEYDEN, J. M. MENENDEZ, J. I. RONDA. *Photorealistic 3D Reconstruction from Handheld Cameras*, in "Machine Vision and Applications", vol. 16, n° 4, 2005.

[36] P. STURM, Z. CHENG, P. C. Y. CHEN, A. N. POO. *Focal Length Calibration from Two Views: Method and Analysis of Singular Cases*, in "Computer Vision and Image Understanding", vol. 99, n° 1, jul 2005, p. 58-95, http://perception.inrialpes.fr/Publications/2005/SCCP05.

[37] P. STURM, S. RAMALINGAM, S. LODHA. *On Calibration, Structure from Motion and Multi-View Geometry for Generic Camera Models*, in "Imaging Beyound the Pinhole Camera", K. DANI-

ILIDIS, R. KLETTE, A. LEONARDIS (editors). , to appear, Kluwer Academic Publishers, 2005, http://perception.inrialpes.fr/Publications/2005/SRL05a.

[38] P. STURM, S. RAMALINGAM. *Géométrie d'images multiples pour des modèles de caméra généraux*, in "Traitement du Signal", vol. 22, nᵒ 5, oct 2005, http://perception.inrialpes.fr/Publications/2005/SR05.

[39] M. WILCZKOWIAK, P. STURM, E. BOYER. *Using Geometric Constraints Through Parallelepipeds for Calibration and 3D Modelling*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 27, nᵒ 2, feb 2005, p. 194-207, http://perception.inrialpes.fr/Publications/2005/WSB05.

Publications in Conferences and Workshops

[40] J. ALLARD, J.-S. FRANCO, C. MÉNIER, E. BOYER, B. RAFFIN. *The GrImage Platform: A Mixed Reality Environment for Interactions*, in "Proceedings of the fourth IEEE International Conference on Computer Vision Systems, New York (USA)", to appear, 2006.

[41] J. ALLARD, C. MÉNIER, E. BOYER, B. RAFFIN. *Running Large VR Applications on a PC Cluster: the FlowVR Experience*, in "Immersive Projection Technology", October 2005, http://perception.inrialpes.fr/Publications/2005/AMBR05.

[42] F. BERTAILS, C. MÉNIER, M.-P. CANI. *A Practical Self-Shadowing Algorithm for Interactive Hair Animations*, in "Graphics Interface", Best student paper award, May 2005, http://perception.inrialpes.fr/Publications/2005/BMC05.

[43] T. BONFORT, P. STURM, P. GARGALLO. *General Specular Surface Triangulation*, in "Proceedings of the Asian Conference on Computer Vision, Hyderabad, India", to appear, January 2006.

[44] E. BOYER. *On Using Silhouettes for Camera Calibration*, in "Proceedings of the Seventh Asian Conference on Computer Vision, Hyderabad (India)", to appear, 2006.

[45] D. COBZAS, P. STURM. *3D SSD tracking with estimated 3D planes*, in "Proceedings of the Second Canadian Conference on Computer and Robot Vision, Victoria, Canada", may 2005, http://perception.inrialpes.fr/Publications/2005/CS05.

[46] J.-S. FRANCO, E. BOYER. *Fusion of Multi-View Silhouette Cues Using a Space Occupancy Grid*, in "Proceedings of the 10th International Conference on Computer Vision", oct 2005, http://perception.inrialpes.fr/Publications/2005/FB05a.

[47] P. GARGALLO, P. STURM. *Bayesian 3D Modeling from Images using Multiple Depth Maps*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California", vol. 2, jun 2005, p. 885-891, http://perception.inrialpes.fr/Publications/2005/GS05.

[48] P. GARGALLO, P. STURM. *Modélisation 3D à partir d'images en utilisant plusieurs cartes de profondeur*, in "Actes des Journées ORASIS, Fournol, France", may 2005, http://perception.inrialpes.fr/Publications/2005/GS05a.

[49] P. HAMMARSTEDT, P. STURM, A. HEYDEN. *Closed-form Solutions and Degenerate Cases for Camera Calibration with One-Dimensional Objects*, in "Proceedings of the International Conference on Computer

Vision, Beijing, China", oct 2005, http://perception.inrialpes.fr/Publications/2005/HSH05.

[50] A. JACQUOT, P. STURM, O. RUCH. *Adaptative Tracking of Non Rigid Objects Based on Color Histograms and Automatic Parameter Selection*, in "Proceedings of IEEE Workshop on Motion and Video Computing, Breckenridge, Colorado", IEEE Computer Society, jan 2005, p. 103-109, http://perception.inrialpes.fr/Publications/2005/JSR05.

[51] A. JACQUOT, P. STURM, O. RUCH. *Suivi Adaptatif d'Objets Non Rigides Basé sur des Histogrammes de Couleur et une Sélection Automatique de Paramètres*, in "Actes des Journées ORASIS", 2005, http://perception.inrialpes.fr/Publications/2005/JSR05a.

[52] D. KNOSSOW, R. RONFARD, R. HORAUD, F. DEVERNAY. *Tracking with the Kinematics of Apparent Contours*, in "Proceedings of the Asian Conference on Computer Vision, Hyderabad, India", to appear, 2006, http://perception.inrialpes.fr/Publications/2006/KRHD06.

[53] H. MATHIEU. *The Cyclope: A 6 DOF Optical Tracker Based on a Single Camera*, in "2nd INTUITION International Workshop, Paris, France", November 2005, http://perception.inrialpes.fr/Publications/2005/Mat05.

[54] C. MÉNIER, J.-S. FRANCO, E. BOYER, B. RAFFIN. *Modélisation Tri-dimensionnelle Temps-réel et Distribuée*, in "Actes des Journées ORASIS", May 2005, http://perception.inrialpes.fr/Publications/2005/MFBR05.

[55] M. NISKANEN, E. BOYER, R. HORAUD. *Articulated Motion Capture from 3-D Points and Normals*, in "British Machine Vision Conference, Oxford, UK", T. CLOCKSIN (editor). , vol. 1, British Machine Vision Association, BMVA, September 2005, p. 439–448, http://perception.inrialpes.fr/Publications/2005/NBH05.

[56] T. PRIBANIC, P. STURM, M. CIFREK. *A Practical Way to Initialize Camera Parameters using the Absolute Conic*, in "Proceedings of the Fifth IASTED International Conference on Visualization, Imaging, and Image Processing, Benidorm, Spain", sep 2005, http://perception.inrialpes.fr/Publications/2005/PSC05a.

[57] T. PRIBANIC, P. STURM, M. CIFREK. *Camera Parameter Initialization for 3D Kinematic Systems*, in "Proceedings of the International Symposium on Image and Signal Processing and Analysis, Zagreb, Croatia", sep 2005, http://perception.inrialpes.fr/Publications/2005/PSC05.

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

[59] S. RAMALINGAM, P. STURM, S. LODHA. *Towards Generic Self-Calibration of Central Cameras*, in "Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China", oct 2005, http://perception.inrialpes.fr/Publications/2005/RSL05b.

[60] S. RAMALINGAM, P. STURM, S. LODHA. *Theory and Calibration Algorithms for Axial Cameras*, in "Proceedings of the Asian Conference on Computer Vision, Hyderabad, India", to appear, January 2006.

[61] P. STURM, T. BONFORT. *How to Compute the Pose of an Object without a Direct View?*, in "Proceedings of the Asian Conference on Computer Vision, Hyderabad, India", to appear, January 2006.

[62] P. STURM, S. RAMALINGAM, S. LODHA. *On Calibration, Structure-From-Motion and Multi-View Geometry for General Camera Models*, in "Proceedings of the 2nd ISPRS Panoramic Photogrammetry Workshop, Berlin, Germany", R. REULKE, U. KNAUER (editors). , Published in the Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI-5/W8, ISPRS, feb 2005, http://perception.inrialpes.fr/Publications/2005/SRL05.

[63] P. STURM. *Multi-View Geometry for General Camera Models*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California", vol. 1, jun 2005, p. 206-212, http://perception.inrialpes.fr/Publications/2005/Stu05.

[64] J.-P. TARDIF, P. STURM. *Calibration of Cameras with Radially Symmetric Distortion*, in "Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China", oct 2005, http://perception.inrialpes.fr/Publications/2005/TS05.

[65] D. WEINLAND, R. RONFARD, E. BOYER. *Motion History Volumes for Free Viewpoint Action Recognition*, in "IEEE International Workshop on modeling People and Human Interaction (PHI'05)", 2005, http://perception.inrialpes.fr/Publications/2005/WRB05.

Internal Reports

[66] F. BERTAILS, C. MÉNIER, M.-P. CANI. *A Practical Self-Shadowing Algorithm for Interactive Hair Animation*, Technical report, nᵒ 5465, INRIA, January 2005, http://www.inria.fr/rrrt/rr-5465.html.

[67] E. BOYER. *Camera Calibration Using Silhouettes*, Technical report, nᵒ 5559, INRIA, April 2005, http://www.inria.fr/rrrt/rr-5559.html.

[68] J.-S. FRANCO, E. BOYER. *Fusion of Multi-View Silhouette Cues Using A Space Occupancy Grid*, Technical report, nᵒ 5551, INRIA, April 2005, http://www.inria.fr/rrrt/rr-5551.html.

[69] S. RAMALINGAM, P. STURM, S. LODHA. *Theory and Experiments towards Complete Generic Calibration*, Rapport de Recherche, nᵒ 5562, INRIA, apr 2005, http://www.inria.fr/rrrt/rr-5562.html.

Bibliography in notes

[70] 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://perception.inrialpes.fr/Publications/2003/BS03c.

[71] E. BOYER, J.-S. FRANCO. *A Hybrid Approach for Computing Visual Hulls of Complex Objects*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA", IEEE Computer Society Press, June 2003, p. 695-701, http://perception.inrialpes.fr/Publications/2003/BF03.

[72] G. DEWAELE, F. DEVERNAY, R. HORAUD. *Hand Motion from 3D Point Trajectories and a Smooth Surface Model*, in "Proceedings of the 8th European Conference on Computer Vision", T. PAJDLA, J. MATAS (editors). , vol. I, May 2004, p. 495-507, http://perception.inrialpes.fr/Publications/2004/DDH04.

[73] J.-S. FRANCO, E. BOYER. *Exact Polyhedral Visual Hulls*, in "Proceedings of the Fourteenth British Machine Vision Conference", Norwich, UK, September 2003, p. 329–338,

http://perception.inrialpes.fr/Publications/2003/FB03.

[74] J.-S. FRANCO, C. MÉNIER, E. BOYER, B. RAFFIN. *A Distributed Approach for Real-Time 3D Modeling*, in "CVPR Workshop on Real-Time 3D Sensors and their Applications", 2004, http://perception.inrialpes.fr/Publications/2004/FMBR04.

[75] C. MÉNIER, J. ALLARD, J.-S. FRANCO, B. RAFFIN, E. BOYER. *Marker-less Real Time 3D Modeling for Virtual Reality*, in "Immersive Projection Technology", 2004, http://perception.inrialpes.fr/Publications/2004/MAFRB04a.

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