INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *Project-Team movi*

# *Computational models for computer vision*

*Rhône-Alpes*

THEME COG

Activity Report

2004

# Table of contents


**1. Team** **1**
**2. Overall Objectives** **1**
**3. Scientific Foundations** **2**
3.1. The geometry of multiple images 2
3.2. Multiple-camera acquisition of visual data 2
3.3. Shape Acquisition 3
3.4. Motion Analysis 3
3.5. Video Processing 3
**4. Application Domains** **4**
4.1. 3D modeling and rendering 4
4.2. Mixed Reality 4
4.3. Human Motion Capture and Analysis 4
4.4. Augmented realities 4
4.5. Multi-media and interactive applications 4
4.6. Car driving technologies 5
4.7. Defense technologies 5
4.8. Geographic information systems 5
4.9. Video based surveillance and monitoring 5
**5. Software** **6**
5.1. Camera calibration 6
5.1.1. Multi-camera calibration 6
5.1.2. Mono and stereo calibration 6
5.2. Blinky: Real-time image acquisition from multiple cameras 8
5.3. Multiple-camera player/recorder 8
5.4. Point tracking in video sequences 9
5.5. Real-time shape acquisition and visualization 9
5.6. Real-time stereo reconstruction 9
**6. New Results** **9**
6.1. Multiple-camera synchronization 9
6.2. Calibration 11
6.2.1. Generalized camera model and calibration 11
6.2.2. Camera calibration using 1D objects 12
6.2.3. Camera self-calibration using planar objects 12
6.3. Tracking 12
6.3.1. Body tracking from silhouettes 12
6.3.2. Tracking of surface elements on multiple synchronized cameras 12
6.3.3. Hand tracking from stereo data 12
6.3.4. Tracking of non-rigid objects 14
6.3.5. Achieving visual attention with two cameras 15
6.4. 3D Modeling 15
6.4.1. Bayesian multi-view reconstruction 15
6.4.2. Reconstruction of specular surfaces 15
6.4.3. Scene modeling using geometrical constraints 16
6.4.4. Structure-from-motion with composite features 16
6.4.5. Real-time surface reconstruction 17
**7. Contracts and Grants with Industry** **17**
7.1. Thalès-Optronics SA 17

7.2. Renault SA 18
**8. Other Grants and Activities** **18**
8.1. National initiatives 18
8.1.1. RNTL-OCETRE 18
8.1.2. RIAM–SEMOCAP 18
8.1.3. ROBEA–ParkNav 19
8.1.4. ACI–Cyber II 19
8.1.5. ACI–GEOLSTEREO 19
8.2. Projects funded by the European Commission 19
8.2.1. FP6-IST STREP project Holonics 19
8.2.2. FP6/Marie-Curie EST Visitor 19
8.2.3. FP6/Marie-Curie RTN VISIONTRAIN 19
**9. Dissemination** **20**
9.1. Editorial boards and program committees 20
9.2. Services to the Scientific Community 21
9.3. Teaching 21
9.4. Invited talks 21
9.5. Thesis 22
**10. Bibliography** **23**

# 1. Team

**Research director**
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]

**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]
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]
Daniel Weinland [Marie-Curie Grant]
Marta Wilczkowiak [INRIA grant]

**Post-doctoral fellows**
Dana Cobzas [Canada]
Kenneth Sundaraj [Singapour]

**Technical staff**
Hervé Mathieu [IR, INRIA]
Florian Geffray
Bertrand Holveck
Loïc Lefort

# 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, 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 $640\times480$ grey-level/color images but in the near future $1600\times1200$ 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 prerecquisite 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 3-D 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 2-D data is the ambiguity of associating a multiple degree of freedom 3-D model with image contours, texture and optical flow. Methods using 3-D data are more relevant with respect to our recent research investigations. 3-D data are produced using stereo or a multiple-camera setup. These data are mathed 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 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.

## 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. Augmented realities

Augmented and mixed 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 and MR 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.

## 4.5. 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.

## 4.6. 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 maneouvering 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.

## 4.7. 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 3-D reconstruction and pose recovery from cameras on-board of the missile, as well as a method for tracking a target in the presence of large scale changes

## 4.8. Geographic information systems

In 2002 we started a three year collaboration with Institut Géographique National (IGN). IGN is interested in building geographic databases of city buildings. The technology that has been selected by IGN for that purpose is based on high-resolution images and computer vision techniques. The basic idea is to gather stereoscopic image pairs of buildings from the ground level and to transform this pictorial information into 3-D information.

This is an interesting application domain for the research that has recently been completed within the MOVI group – 3-D reconstruction of polyhedral structures.

## 4.9. Video based surveillance and monitoring

There is an increasing need for automatic human activity recognition based on video analysis. This demand is due to the huge amount of videos to analyse with respect to the small number of human operators in charge of this work. The application domains concern the security domain (for instance airport, museum, and public-building surveillance), the safety domain to prevent dangerous situations (i.e. agression in public areas) and to protect people from accidents (like old people falling down at home), as well as the monitoring domain (for instance, building statistics for museum visits, commercial shopping centers, attraction parks, etc.).

There are thousands of CCTV cameras in the subways, on highways, on stadiums, and so forth but only very few security operators which have the tedious task of selecting in real-time the video cameras to display on screens and to watch them. Such statistical data as the trajectories of people visiting a museum, the actual time spent in front of eachone of the exhibited artcrafts are simply not available.

In medical applications constant monitoring and surveillance of disabled people and/or people with special needs by authorized professional is not always feasible. There are not enough staff locally (at home or in the hospital) to visit regularly each patient. Remote surveillance is needed to alert the medical staff.

In order to protect the private life of people under monitoring, it is crucial to communicate verbal descriptions of images and not the images themselves. Therefore, the development of cognitive systems for these applications is of major importance. Automatic annotation of these videos in terms of meta-data describing the physical objects or persons and their activities is still a research challenge.

# 5. Software

## 5.1. Camera calibration

### *5.1.1. Multi-camera calibration*

We developed also new 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.

### *5.1.2. 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://www.inrialpes.fr/movi/soft/calibration/index.html.

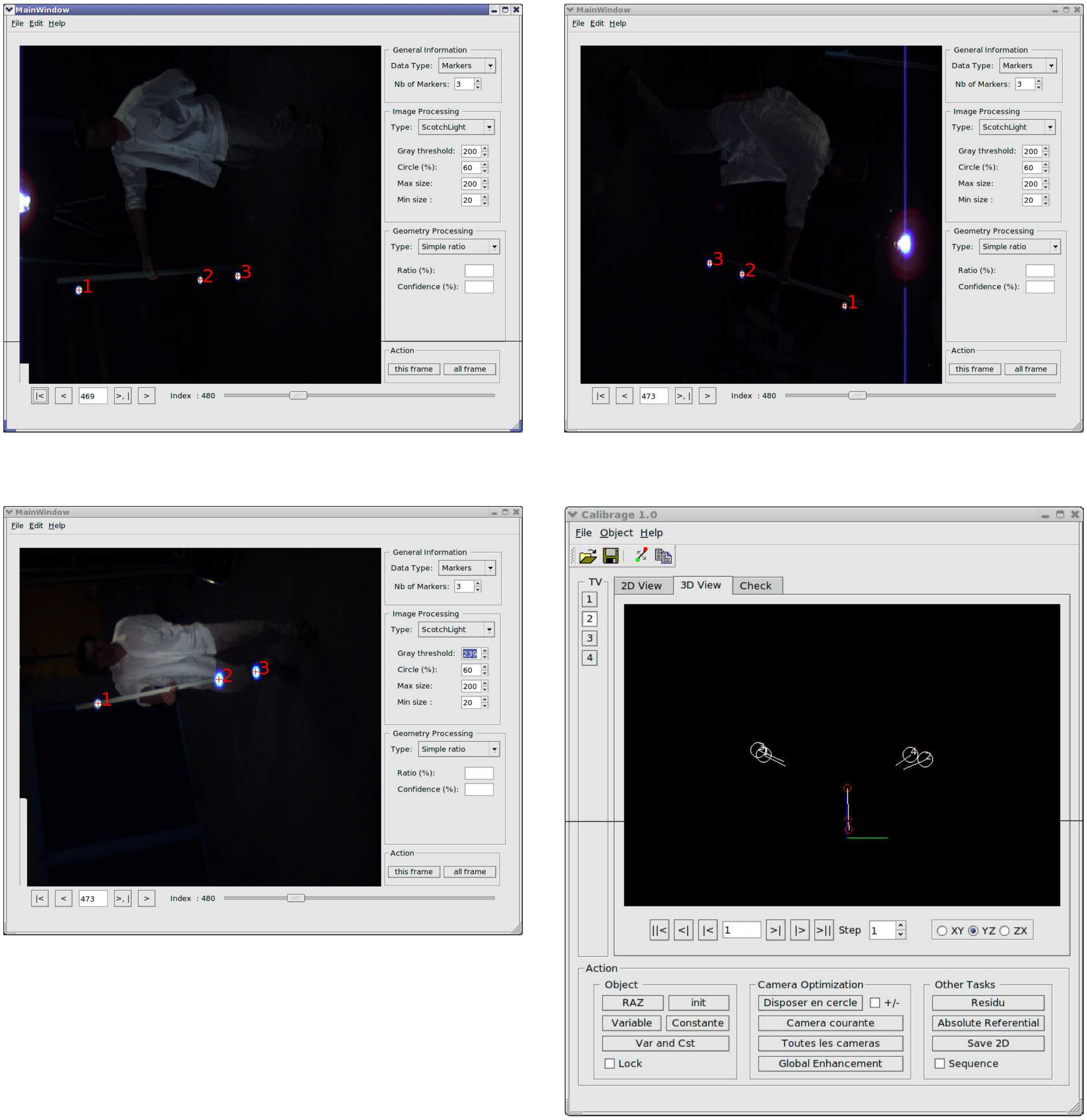


*Figure 1. This figure shows the output of the multi-camera calibration system which is under development. A 3-ball object is freely moved in the cameras' field of view. This object is identified in each image and at each time instant. Calibration data are thus gathered over a large number of frames and for all the cameras. All these data are then used to calibrate both the intrinsic and extrinsic parameters of the cameras.*

## 5.2. 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.

Figure 2 shows an subset of the camera setup that we developped. The camera setup is composed of standard FireWire cameras linked to the PC-cluster.

## 5.3. Multiple-camera player/recorder

We developped a new software with a graphical user interface (GUI) – Figure 3, 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 frames 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, and as well 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 developping other real-time video processing modules.


*Figure 2. These three cameras are linked to the PC cluster through the IEEE 1394 bus protocol and they are synchronized with all the other cameras forming the acquisition setup.*

## 5.4. 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.

## 5.5. 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, Figure 4 . Its implementation was made in collaboration with the Apache 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. A short video is downloadable at http://www.inrialpes.fr/movi/people/Boyer/Mpeg/model-divx.avi.

## 5.6. Real-time stereo reconstruction

A parallel implementation of correlation-based stereoscopy was implemented using PVM (Parallel Virtual Machine), making it usable on multi-processor machines or on a PC cluster. This algorithm is directly connected to cameras using the Blinky acquisition library described above, making it usable for real-time stereoscopic reconstruction from two video cameras.

# 6. New Results

## 6.1. Multiple-camera synchronization

The simultaneous (synchronized) acquisition and processing of multiple video sequences requires data transfer rates that are far above the specifications of a standard workstation [48]. An alternative is to use a

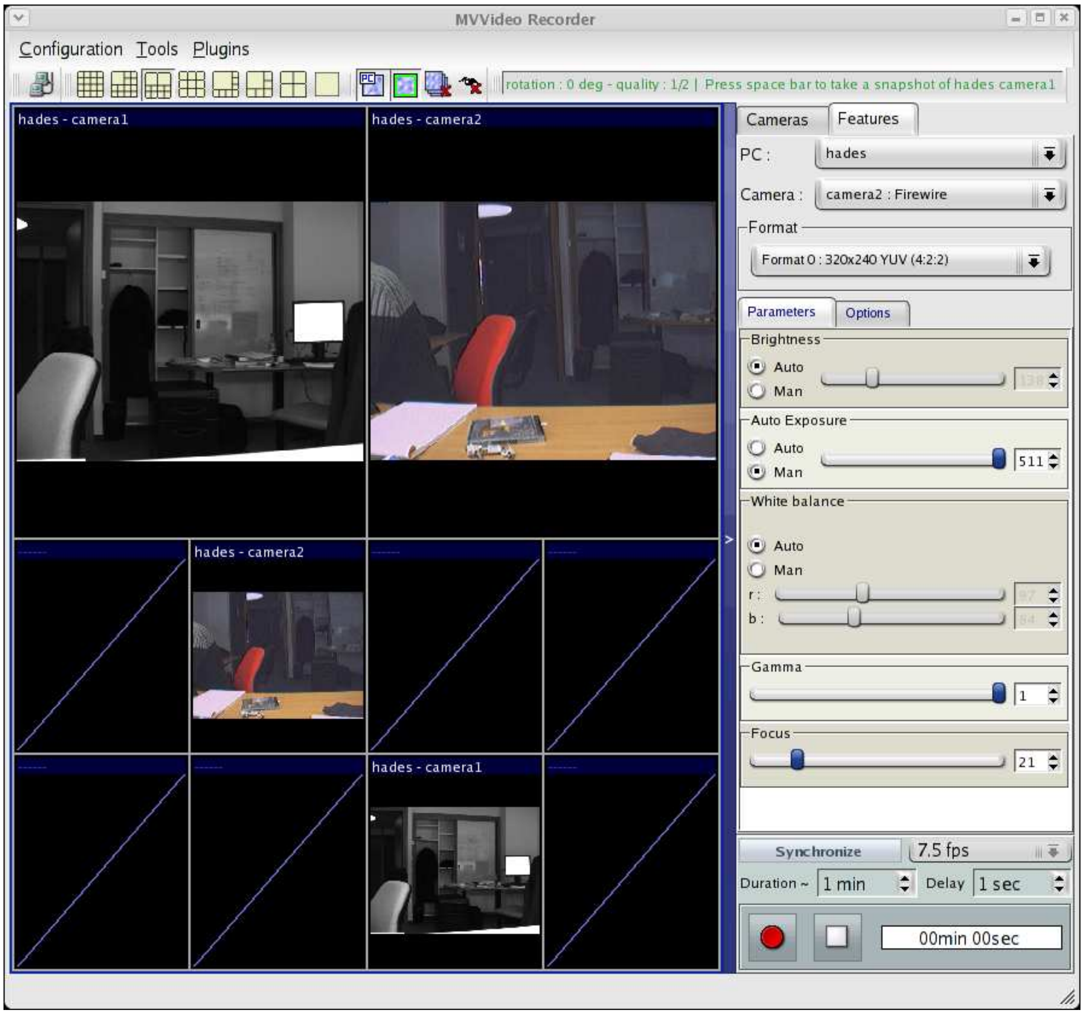


*Figure 3. Screen shot of a multiple-camera recording session. Left: live feed of all recording cameras. Right: Real-time control of camera parameters.*

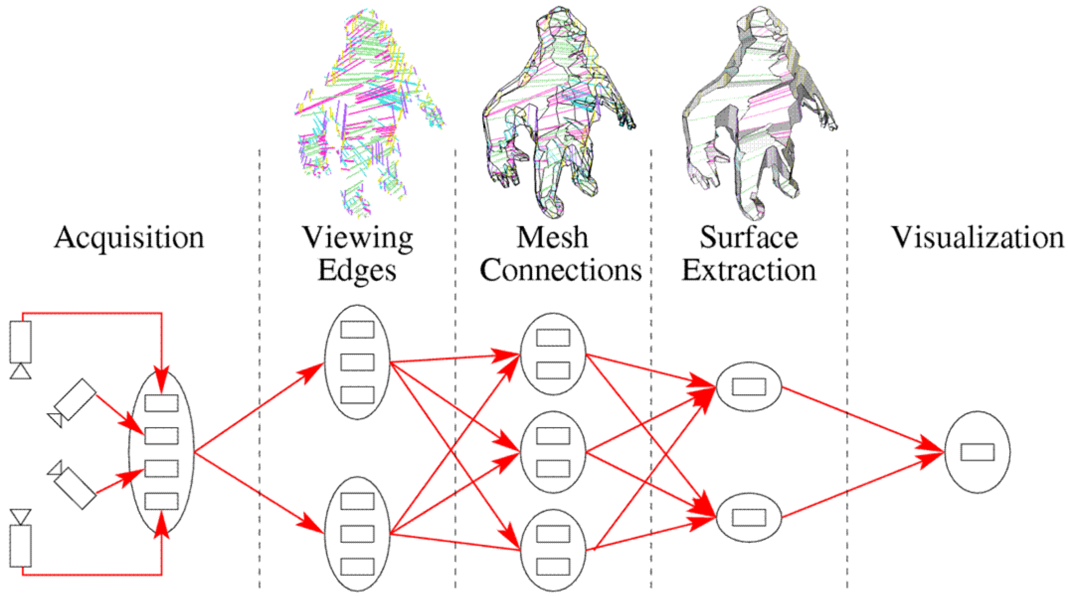


*Figure 4. This figure shows how the computation of the visual hull is carried out using a PC-cluster.*

distributed PC system, where each PC unit handles one or two cameras with different viewpoints. Therefore, arises the problem of multiple-camera synchronization over a PC network. We developed two synchronization solutions.

The first solution is somehow classical and it consists in adding a secondary network which is dedicated to propagating external synchronization signals; This solution properly synchronizes the cameras themselves but does not allow a user to identify the set of images available at some time instant. The second 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.

## 6.2. Calibration

### *6.2.1. Generalized camera model and calibration*

We developed a generic calibration approach, allowing to calibrate all main types of cameras used in computer vision using one framework [46]. It is based on a simple general camera model: a camera is considered as a set of pixels plus one projection ray per pixel, that models the line of sight for that pixel. This model encompasses pinhole cameras as well as cameras with any geometrical distortions (radial, tangential, etc.), catadioptric cameras, stereo systems, etc. We developed a theory for our calibration 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. Besides calibration, we also formulated other structure-from-motion problems for the general camera model, e.g. pose and motion estimation, 3D point triangulation and bundle adjustment [43]. Work on establishing a complete multi-view geometry for the general camera model, is ongoing.

### 6.2.2. Camera calibration using 1D objects

Recently, an approach for calibrating a camera using several images of a linear object (typically, a stick with a few markers on it) was proposed in the literature. Like many other calibration or self-calibration approaches, this one is subject to degeneracies, i.e. if the calibration object moves in specific ways, calibration will not be successful. We have completely analyzed these critical motions (an article has been submitted).

### 6.2.3. 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][45]. 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.

## 6.3. Tracking

### 6.3.1. Body tracking from silhouettes

Two new algorithms were proposed for tracking full body human motion seen by multiple cameras. The first algorithm is based on a careful mathematical analysis of the apparent motion of the occluding contours of body part primitives. More specifically, we use deformed cylinders such as truncated elliptical cones to model human body parts, project their occluding contours in all camera views, match them to the observed image contours within the actor's silhouette using a distance transform, and find the best-matching pose in the least-squares sense.

The second algorithm developed a different strategy. It relies first on building 3-D observations (surface patches) from image silhouettes using a feed-forward method and second on fitting an articulated object model to these observations through minimization in 3D. The objective function measuring the discrepancy between model and data takes into account both the *scaled algebraic distance* from data points to model surface and the *difference in orientation* between observed surface patches and normals to the model's surface – see Figure 5.

Both algorithms were successfully tested on synthetic and real sequences of simple movements. Current works investigate faster and more difficult movements, and methods to automatically calibrate models to new actors.

### 6.3.2. Tracking of surface elements on multiple synchronized cameras

Tracking of points or surface elements (surfels or surface patches) is usually done using a single camera and a method such as the Lucas-Kanade point tracker. However, points tracked this way are lost as soon as their appearance changes too much (e.g. because of surface slant) or as they become occluded. Besides, when using multiple cameras, nothing ensures that the point position remains consistent across all cameras.

To overcome these problems and to take advantage of multiple-camera acquisition, we developped a method that tracks the position and orientation. Each surfel is described by a texture patch, its 3-D position, and its orientation (3 angles). This method progressively updates the appearance of the surfel, and automatically detects when the surfel is not visible in one camera because of its orientation or because it's occluded. This becomes very useful when the object motion is completely free, e.g. when a dancer is performing in a multiple-camera studio: this method will give point tracks over a long period of time, even though the points disappear on some cameras and reappear on others.

### 6.3.3. Hand tracking from stereo data

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.

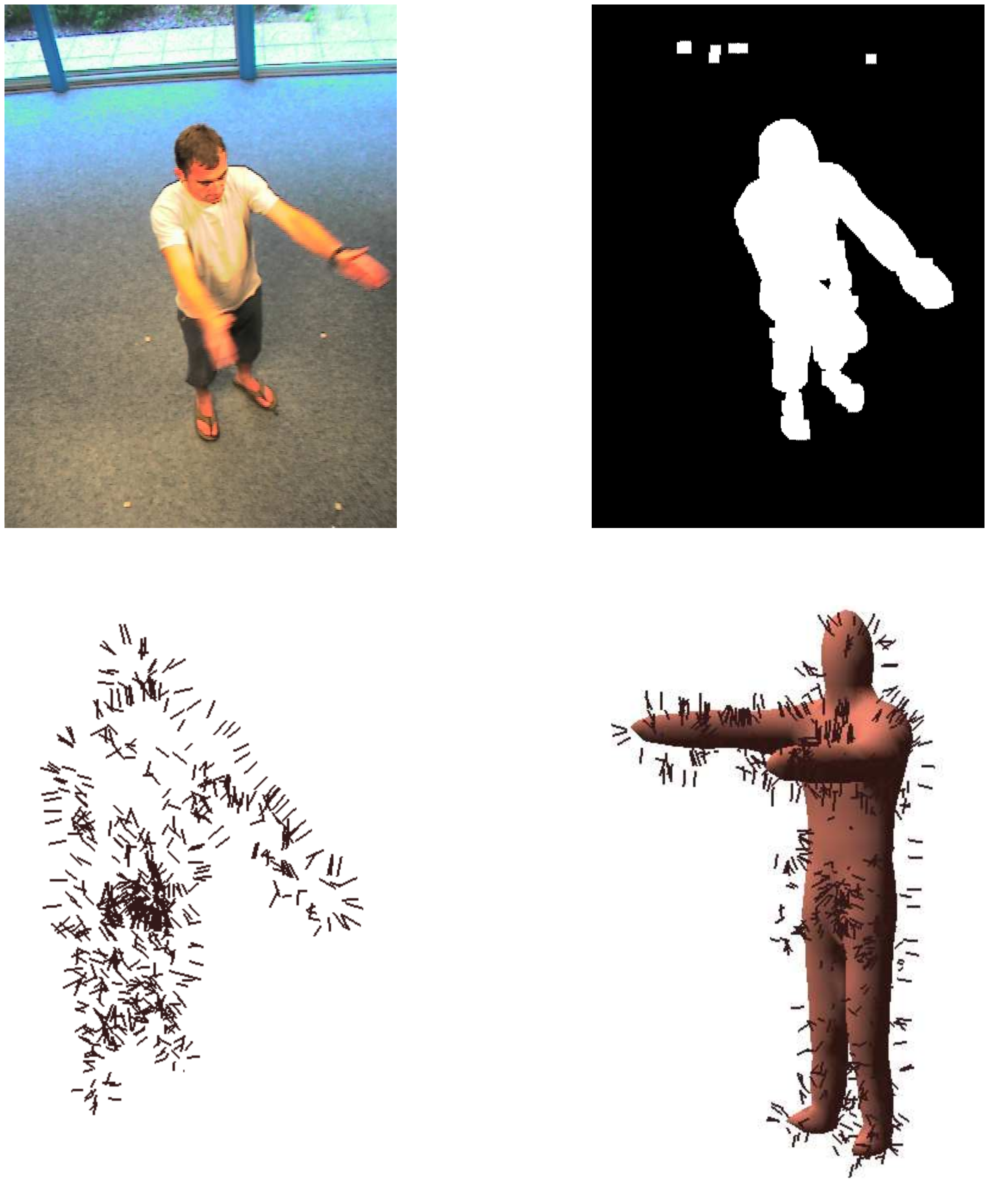

*Figure 5. This figure shows the result of the tracker applied to 6 video sequences gathered with synchronized and calibrated cameras. There were 150 frames in a sequence. A silhouette is extracted from the raw image and a collection of 3-D surface patches are estimated at each time instant. The corresponding pose of the model which has been fitted to these data is shown with the data superimposed on it. One should notice the quality of the pose in spite of an obvious discrepancy between the data and the model's surface.*

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 [38]. Results were obtained on a stereoscopic sequence of a moving hand, and were evaluated using a side view of the sequence – see Figure 6.

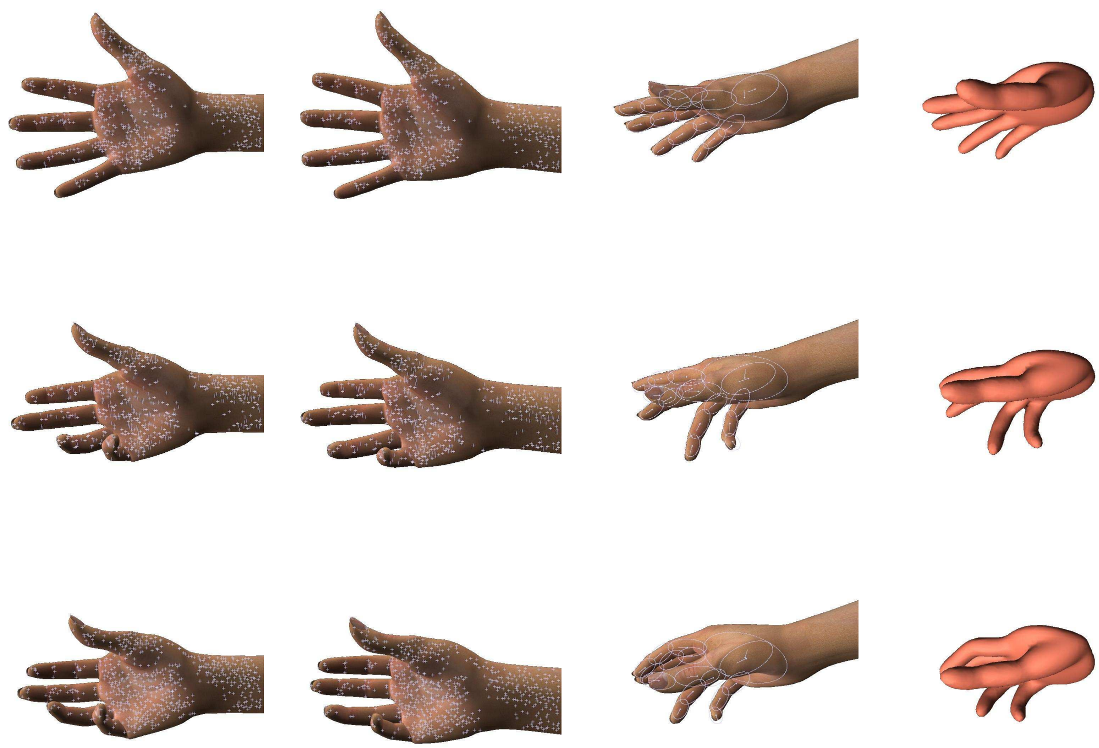

*Figure 6. Points of interests are matched and tracked in a binocular (stereo) image sequence. A virtual model of a hand composed of a kinematic model, a set of metaballs, and an implicit surface performs the same gesture as the real hand.*

### 6.3.4. Tracking of non-rigid objects

We extended recent approaches in the literature on color-based probabilistic object tracking. One main aim is to not only estimate a single object state for each frame, but rather to represent the probability distribution over state space. This is achieved via particle filtering. The object model used is based on color histograms of an image region to be tracked, and may be updated during the tracking, to accommodate to appearance changes (due to e.g. lighting changes or non-rigid object motion). Our recent work [40] concerns several aspects of this approach: (i) application of a criterion from the statistical literature to select the appropriate bin number for the model histogram (too many will lead to weak robustness to appearance changes, too few do not allow a good discrimination of the target). (ii) finding a compromise between losing all spatial information (color-histograms) and keeping all of it (template matching). Currently, we do this by representing an object by several histograms, for several sub-regions. (iii) strategy for the model update.

#### *6.3.5. Achieving visual attention with two cameras*

Within the context of visual attention, the problem to be addressed is to maintain a visual event within the field of view of a moving camera. We established a computational model for visual attention using cooperation between a wide field-of-view static camera and a rotating and zooming camera. The model is based on the coupling between geometry and kinematics. We derived an algebraic formulation for this coupling and we specified the practical conditions yielding a unique solution. A set of outdoor experiments validates the approach [49].

### 6.4. 3D Modeling

#### *6.4.1. Bayesian multi-view reconstruction*

We developed a probabilistic approach for dense 3D modeling from multiple images. In order to completely model the input data, we propose to represent the scene as a set of colored depth maps, one per input image. We formulate the problem as a Bayesian MAP maximization which leads to an energy minimization method. Hidden visibility variables are used to deal with occlusion, reflections and outliers. Main contributions of our work are: a prior for the visibility variables that treats the geometric occlusions; and a prior for the multiple depth maps model that smoothes and merges the depth maps while enabling discontinuities – see Figure 7. An article on this work has been submitted.

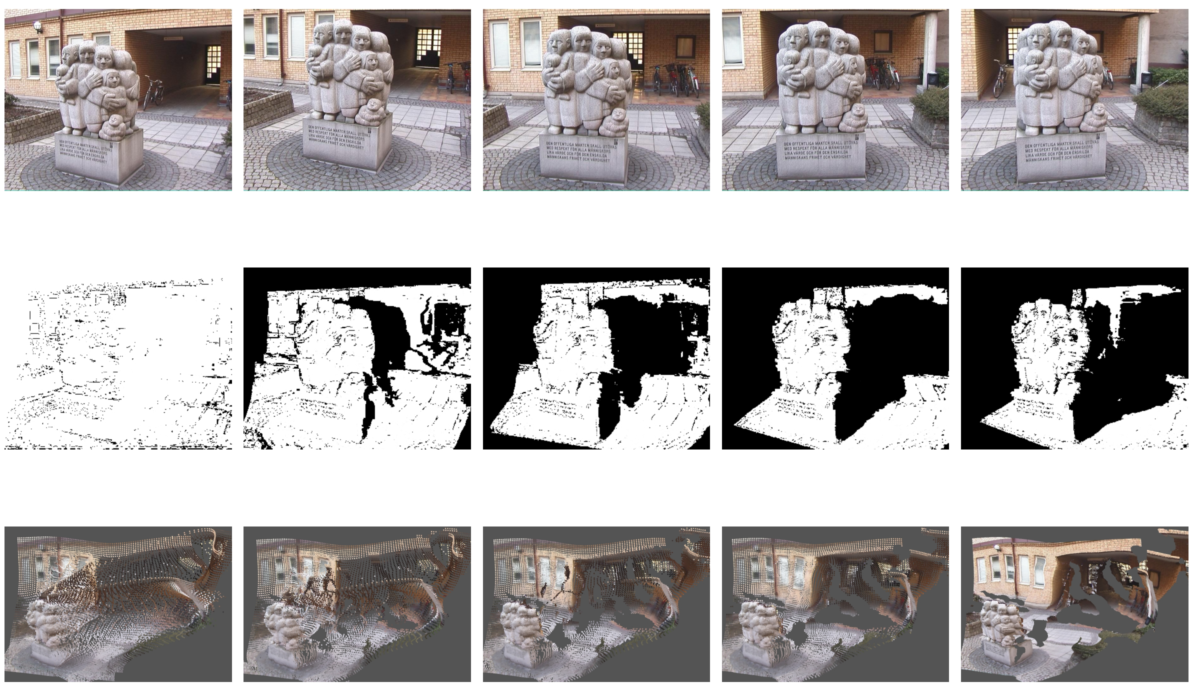

*Figure 7. The first row presents the five input images. The second one, the visibility maps of the first image. That is, the estimated probabilities of the 3D points instantiated by the first depth map of being visible in the each of the other images. The last row shows a point rendering of the set of all the depth maps at the same time during the evolution of the algorithm, from a very coarse initialization on the left, to the final model on the right.*

#### *6.4.2. Reconstruction of specular surfaces*

We continue our work on reconstructing specular surfaces from image sequences. Our previous work required taking images of known objects, from known positions. Recently, we have developed an approach that does not need to know positions, which makes it much more practical, and more general than most previously known methods. We have completed the theoretical developments and the formulation of a practical algorithm, and are currently performing practical experiments.

### *6.4.3. Scene modeling using geometrical constraints*

We have proposed methods for camera calibration, pose estimation and 3D reconstruction that are based on geometrical constraints on the observed scene (parallelism, perpendicularity, co-planarity, etc.) [27][47][35], Figure 8. 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. 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. These incorporate all geometric constraints and allow thus an efficient bundle adjustment using standard unconstrained optimization techniques.

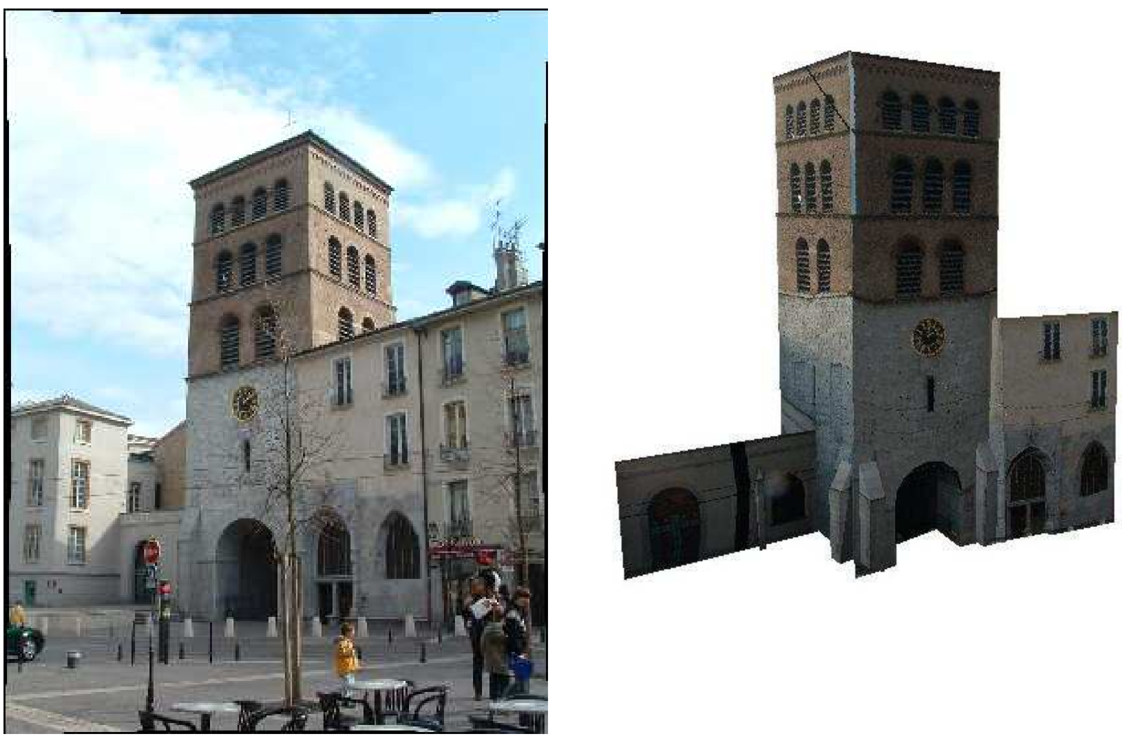

*Figure 8. Constraint based reconstruction example: on the left, one of five photographs used for the reconstruction; on the right, a screenshot of the reconstructed model.*

### *6.4.4. Structure-from-motion with composite features*

Structure-from-motion is usually considered for point or line features. An interesting composite feature type is what we call a “pencil of points”, i.e. a set of collinear points. These may be considered especially in man-made environments. They are interesting in several respects: a single feature already carries geometric matching constraints with it (relative position of points on their support line) and the fundamental matrix and other multi-view relations can be computed from relatively few matches. In our work, we deal with the detection and matching of these composite features, and propose algorithms for the estimation of epipolar geometry, triangulation, and bundle adjustment [37].

### *6.4.5. Real-time surface reconstruction*

We have proposed 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 [39][36], Figure 9.

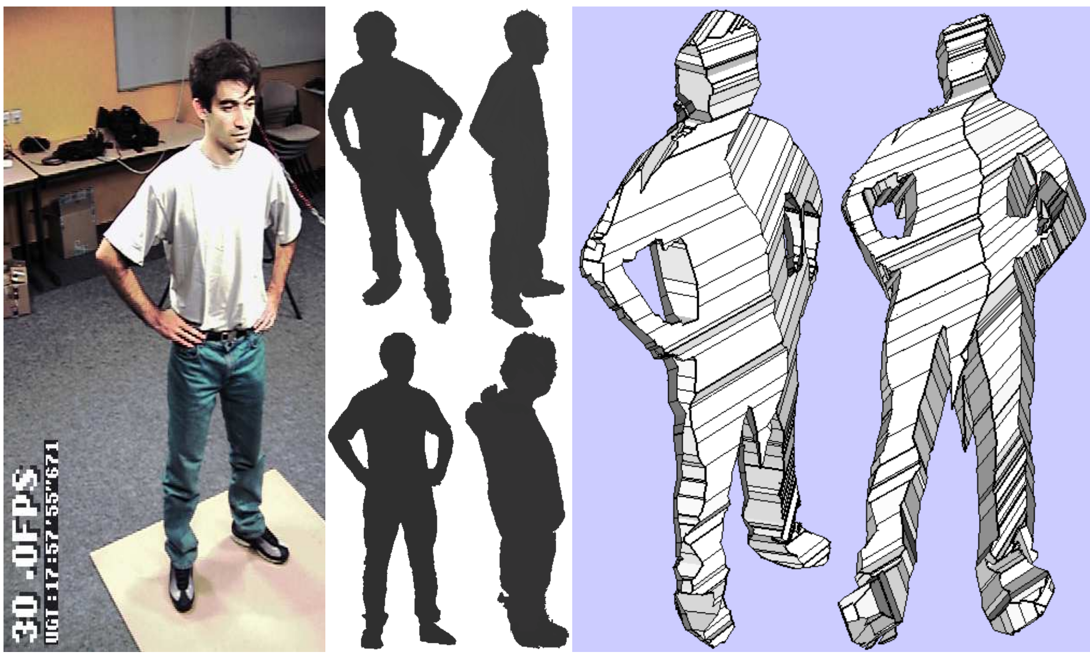


*Figure 9. Real time reconstruction of a person with 4 cameras and the exact visual-hull approach.*

# 7. Contracts and Grants with Industry

## 7.1. Thalès-Optronics SA

In october 2003 we started a 3-year collaboration with Thales-Optronics SA (TOSA). Moreover, Aude Jacquot was granted with an industrial PhD felloship from ANRT.

TOSA devolops 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 3-D structure.

The scientific and technological objectives are the following: to develop tools enabling (i) off-line modelling of complex 3-D 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

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 will develop a prototype system based on stereoscopic vision with the following functionalities: *low speed following*, *pre-crash*, and *pedestrian detection*. In particular we will study the robustness of the image processing algorithms with respect to weather conditions (the system should be able to alert the driver that it cannot forward reliable information, in such cases as fog, rain, darkness, etc.) and to camera/stereo calibration problems (the system should be able to self-detect such problems).

The prototype system that we will develop should perform at a speed of 10 frames per second on a standard PC architecture. Real-time and hardware-embedding issues will not be addressed within this collaboration.

# 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 3-D 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 are currently building a prototype using the GRIMAGE infrastructure, which will be demonstrated, evaluated and ported to a lightweight platform by our industrial partners ASICA and ARTEFACTO. The targetted applications are games and 3D animation.

### 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* programme (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 3-D deformations over time. In particular, the expected accuracy of the planned measurements is of the order of $10\mu$m which corresponds to an actual amplitude of a few centimeters.

## 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 3-D holographic display devices.

Our team will develop a real-time multi-camera and multi-PC system. The developments will be based on 3-D 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 commissio. 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

. In November 2003 we coordinated and submitted the VISIONTRAIN proposal for a Marie Curie Research Training Network, or RTN. 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 programme.

VISIONTRAIN will address 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 targetted research topics: computational theories and methods for low-level vision, motion understanding from image sequences, learning and recogntion 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 workprogramme will include participation to proof-of-concept achievements, annual thematic schools, industrial meetings, attendance of conferences, etc.

**The VISIONTRAIN consortium was invited for contract negociations in October 2004. All the contract documents were submitted and accepted by the EC representative in December 2004. The contract will start in 2005 for a duration of 4 years.**

# 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, with Seth Teller (MIT) and Tomáš Svoboda (Prague) the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, held in conjunction with ECCV in Prague, Czech Republic, May 2004 [26].
- Peter Sturm was member of the Advisory Board of the International Workshop on Vision Techniques Applied to the Rehabilitation of City Centres, held in Lisbon, Portugal, October 2004.
- Peter Sturm was member of the Program Committees of:
  - ECCV'04 (European Conference on Computer Vision, Prague),
  - ACCV'04 (Asian Conference on Computer Vision, Jeju Island, Korea),
  - ICIP'04 (IEEE International Conference on Image Processing, Singapore),
  - BMVC'04 (British Machine Vision Conference, Kingston, UK),
  - RFIA'04 (Congrès de Reconnaissance des Formes et Intelligence Artificielle, Toulouse),
  - ICASSP'04 (Intl. Conference on Acoustics, Speech and Signal Processing, Montreal).
- Edmond Boyer was member of the Program Committees of:
  - ECCV'04 (European Conference on Computer Vision, Prague),
  - ICIP'04 (IEEE International Conference on Image Processing, Singapore),
  - ICASSP'04 (Intl. Conference on Acoustics, Speech and Signal Processing, Montreal).

## 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 has been appointed Co-Chairman of the Working Group “Image Orientation” of the ISPRS (International Society for Photogrammetry and Remote Sensing), for the period 2004-2008.
- Remi Ronfard is member of the "Commission de Specialistes" for recruitments at the University Joseph Fourier of Grenoble.
- Remi Ronfard organized a weekly Reading Group on Human Motion Capture in 2003-2004 which was followed by researchers and students in MOVI and EVASION.
- 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

- Analyse d’images, DESS INFORMATIQUE, UNIV. JOSEPH FOURIER, 30H, R. Ronfard.
- Optimisation, DEA IVR, INPG, 6h, P. Sturm.
- Vision 3D, DEA IVR, INPG, 12h, P. Sturm.
- Géométrie projective, DEA IVR, INPG, 6h, E. Boyer.

## 9.4. Invited talks

- Rémi Ronfard gave an invited talk on “Detecting people in images and video” in May 2004 at CNRT TIM Séminaire Vision par Ordinateur pour les Télécoms, at France Telecom, Rennes,
- Peter Sturm gave invited talks at:
  - 12th Seminar on Theoretical Foundations of Computer Vision (“Imaging Beyond the Pinhole Camera”), Schloß Dagstuhl, Germany, June 2004. Title: A Generic Framework for Structure&Motion and Camera Calibration.
  - Sogang University, Dept. of Media Technology, Seoul, South Korea, August 2004,
  - Sogang University, Dept. of Mathematics, Seoul, South Korea, August 2004,
  - Seoul National University, South Korea, September 2004,
  - GIST (Gwangju Institute of Science and Technology), Gwangju, South Korea, Sept. 2004,
  - KAIST (Korea Advanced Institute of Science and Technology), Daejeon, South Korea, Sept. 2004,
  - Czech Technical University, Prague, Czech Republic, November 2004.

## 9.5. Thesis

- Marta Wilczkowiak defended her PhD thesis in April 2004.
- Peter Sturm acted as a committee member for the following PhD theses:
  - Branislav Mičušik (Czech Technical University, Prague),
  - Sébastien Cornou (Université Blaise Pascal, Clermont-Ferrand),
  - Carlos Hernández Esteban (Ecole Nationale Supérieure des Télécommunications).
  - Omar Tahri (Université de Rennes 1),
  - Jonathan Fabrizio (Université Pierre et Marie Curie, Paris 6).
- Radu Horaud acted as a committee member for the following PhD theses:
  - Il-Kyun Jung (Institut National Polytechnique de Toulouse),
  - Éric Marchand (HDR of Université de Rennes I).
- The following PhD students from universities abroad did internships with MOVI:
  - Srikumar Ramalingam (University of California at Santa Cruz, 6 months),
  - Tomislav Pribanić (Zagreb University, Croatia, 3 months),
  - Pär Hammarstedt (Malmö University, Sweden, 2 months),
  - Ferran Espuny (UPC Barcelona, Spain, 2 months).

# 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://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", 2001, p. 525-531, http://www.inrialpes.fr/movi/publi/Publications/2001/MS01.

[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, n° 2-3, 2001, p. 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", vol. 19, n° 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, n° 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, n° 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, n° 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, n° 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, n° 2, 2000, p. 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", vol. 19, n° 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. *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.

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

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

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

## Books and Monographs

[26] P. Sturm, T. Svoboda, S. Teller (editors). *Proceedings of the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Came ras, held in conjunction with ECCV, Prague, Czech Republic*, May 2004.

## Doctoral dissertations and Habilitation theses

[27] M. Wilczkowiak. *3D Modelling From Images Using Geometric Constraints*, Thèse de Doctorat, Institut National Polytechnique de Grenoble, April 2004.

## Articles in referred journals and book chapters

[28] A. Bartoli, N. Dalal, R. Horaud. *Motion Panoramas*, in "Computer Animation and Virtual Worlds", vol. 15, n° 6, November 2004, p. 501–517, http://www.inria.fr/rrrt/rr-4771.html.

[29] A. Bartoli, P. Sturm. *Non-Linear Estimation of the Fundamental Matrix With Minimal Parameters*, in "ieee Transactions on Pattern Analysis and Machine Intelligence", vol. 26, n° 4, 2004, p. 426–432.

[30] A. Bartoli, P. Sturm. *The 3D Line Motion Matrix and Alignement of Line Reconstructions*, in "International Journal of Computer Vision", vol. 57, n° 3, 2004, p. 159–178.

[31] O. Chum, T. Pajdla, P. Sturm. *The Geometric Error for Homographies*, in "Computer Vision and Image Understanding", to appear, 2004.

[32] Y. Dufournaud, C. Schmid, R. Horaud. *Image Matching with Scale Adjustment*, in "Computer Vision and Image Understanding", vol. 93, n° 2, February 2004, p. 175–194, http://www.inria.fr/rrrt/rr-4458.html.

[33] R. Ronfard. *MPEG, une norme pour la compression, la structuration et la description du son*, Informatique et Systemes d'Information, sous la direction de Francois Pachet et Jean-Pierre Briot, chap. 12, Hermes, 11 rue Lavoisier, 75 008, Paris, 2004, p. 403–422, http://www.inrialpes.fr/movi/publi/Publications/2004/Ron04.

[34] P. Sturm, Z. Cheng, P. Chen, A. Poo. *Focal Length Calibration from Two Views: Method and Analysis of Singular Cases*, in "Computer Vision and Image Understanding", to appear, 2004.

[35] 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", to appear, 2004.

## Publications in Conferences and Workshops

[36] J. Allard, E. Boyer, J. Franco, C. Menier, B. Raffin. *Marker-less Real Time 3D Modeling for Virtual Reality*, in "Immersive Projection Technology, Iowa State University", May 2004.

[37] A. Bartoli, M. Coquerelle, P. Sturm. *A Framework For Pencil-of-Points Structure-From-Motion*, in "Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic", T. Pajdla, J. Matas (editors)., Lecture Notes in Computer Science, vol. 3022, May 2004, p. 28–40.

[38] G. DEWAELE, F. DEVERNAY, R. HORAUD. *Hand Motion from 3D Point Trajectories and a Smooth Surface Model*, in "8th European Conference on Computer Vision", T. PAJDLA, J. MATAS (editors)., LNCS 3021, vol. I, Springer, May 2004, p. 495-507, http://www.inrialpes.fr/movi/publi/Publications/2004/DDH04.

[39] J. FRANCO, C. MENIER, E. BOYER, B. RAFFIN. *A Distrbuted Approach for Real-Time 3D Modeling*, in "IEEE CVPR Workshop on Real-Time 3D Sensors and their Applications, Washington DC", July 2004.

[40] A. JACQUOT, P. STURM, R. HORAUD, O. RUCH. *Adaptive Tracking of Non-Rigid Objects Based on Color Histograms and Automatic Parameter Selection*, in "Proceedings of the ieee Workshop on Motion and Video Computing, Breckenridge, Colorado, USA", to appear, January 2004.

[41] F. JURIE, E. NOIRFALISE, P. STURM. *Calibrage itératif de caméras à partir de scènes planes*, in "14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, T oulouse, France", vol. 1, January 2004, p. 261–269.

[42] I. MARTIN, R. RONFARD, F. BERNARDINI. *Detail-Preserving Variational Surface Design with Multiresolution Constraints*, in "International Conference on Shape Modeling and Applications", Genova, 2004, http://www.inrialpes.fr/movi/publi/Publications/2004/MRB04.

[43] S. RAMALINGAM, S. LODHA, P. STURM. *A Generic Structure-from-Motion Algorithm for Cross-Camera Scenarios*, in "Proceedings of the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical C ameras, Prague, Czech Republic", May 2004, p. 175–186.

[44] R. RONFARD. *Reading Movies - An Integrated DVD Player for Browsing Movies and their Scripts*, in "ACM Conference on Multimedia", New York, New York, 2004, http://www.inrialpes.fr/movi/publi/Publications/2004/Ron04b.

[45] P. STURM, P. GURDJOS. *Méthodes et fondements géométriques pour l'autocalibrage plan*, in "14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, T oulouse, France", vol. 1, January 2004, p. 351–360.

[46] P. STURM, S. RAMALINGAM. *A Generic Concept for Camera Calibration*, in "Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic", T. PAJDLA, J. MATAS (editors)., Lecture Notes in Computer Science, vol. 3022, Springer-Verlag, May 2004, p. 1–13.

[47] M. WILCZKOWIAK, P. STURM, E. BOYER. *Using Geometric Constraints for Camera Calibration and Positioning and 3D Scene Modelling*, in "Proceedings of the CIPA International Workshop on Vision Techniques Applied to the Rehabilitat ion of City Centres, Lisbon, Portugal", October 2004.

## Internal Reports

[48] B. HOLVECK, H. MATHIEU. *Infrastructure of the GrImage Experimental Platform: The Video Acquisition Part*, Technical report, n° RT-0301, INRIA, INRIA Rhône-Alpes, Montbonnot, November 2004, http://www.inria.fr/rrrt/rt-0301.html.

[49] R. HORAUD, D. KNOSSOW, M. MICHAELIS. *Camera cooperation for achieving visual attention*, submitted to Machine Vision and Applications, Technical report, n° RR-5216, INRIA, INRIA Rhône-Alpes, Montbonnot, June 2004, http://www.inria.fr/rrrt/rr-5216.html.