



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

*Project-Team perception*

*Interpretation and Modelling of Images  
and Videos*

*Grenoble - Rhône-Alpes*

THEME COG

*Activity*  
*R* *eport*

2007



## Table of contents

<b>1. Team</b>	<b>1</b>
<b>2. Overall Objectives</b>	<b>1</b>
2.1. Introduction	1
2.2. Highlights of the year	3
2.2.1. 4D View Solutions SAS	3
2.2.2. Markerless 3D interactions	3
<b>3. Scientific Foundations</b>	<b>4</b>
3.1. The geometry of multiple images	4
3.2. The photometry component	4
3.3. Shape Acquisition	4
3.4. Motion Analysis	4
3.5. Multiple-camera acquisition of visual data	5
<b>4. Application Domains</b>	<b>5</b>
4.1. 3D modeling and rendering	5
4.2. Mixed and Augmented Reality	5
4.3. Human Motion Capture and Analysis	6
4.4. Multi-media and interactive applications	6
4.5. Car driving technologies	6
4.6. Defense technologies	6
<b>5. Software</b>	<b>7</b>
5.1. Platforms	7
5.1.1. The Grimage platform	7
5.1.2. The mini-Grimage platform	7
5.2. 3-D surface reconstruction based on mesh evolution	7
5.3. Dense voxel registration	7
5.4. Real-time shape acquisition and visualization	8
<b>6. New Results</b>	<b>8</b>
6.1. Computational modelling of binocular vision	8
6.2. Audio-visual perception	9
6.2.1. Multi-speaker localization	9
6.2.2. An audiovisual robot head	10
6.2.3. Synchronized stereoscopic and binaural datasets with head movements	10
6.3. Dense stereo and Markov random fields	11
6.4. Human-body tracking and human-motion capture	13
6.4.1. Human-body tracking using an implicit surface, 3D points, and surface normals.	13
6.4.2. Human-body tracking using the kinematics of extremal contours.	13
6.4.3. Sequential Monte Carlo Inverse Kinematics.	13
6.5. Multiple camera reconstruction	17
6.5.1. Point-based reconstruction using robust factorization	17
6.5.2. Surface reconstruction based on mesh evolution	17
6.5.3. Multi-view stereo	17
6.6. Image-based modeling of reflectance properties	21
6.7. Motion Segmentation	21
6.7.1. Spectral clustering of motion trajectories	21
6.7.2. Unsupervised articulated motion segmentation	21
6.8. Articulated shape matching	22
6.9. Temporal surface tracking	22
6.10. Action representation and recognition	22
6.11. Omnidirectional vision	24

---

6.12. Other results	24
6.12.1. Conic fitting	24
6.12.2. Automatically identifying foreground in multiple images	25
<b>7. Contracts and Grants with Industry</b>	<b>25</b>
<b>8. Other Grants and Activities</b>	<b>26</b>
8.1. National initiatives	26
8.1.1. ANR project CAVIAR	26
8.1.2. ANR project DALIA	26
8.1.3. ANR project FLAMENCO	26
8.1.4. ARC-Georep	27
8.1.5. ARC-FANTASTIK	27
8.2. Projects funded by the European Commission	27
8.2.1. FP6-IST STREP project Holonics	27
8.2.2. FP6/Marie-Curie EST Visitor	27
8.2.3. FP6/Marie-Curie RTN VISIONTRAIN	27
8.2.4. FP6 IST STREP project POP	28
8.2.5. FP6-IST STREP project INTERACT	28
<b>9. Dissemination</b>	<b>29</b>
9.1. Editorial boards and program committees	29
9.2. Services to the Scientific Community	29
9.3. Teaching	30
9.4. Tutorials and invited talks	30
9.5. Thesis	30
<b>10. Bibliography</b>	<b>30</b>

# 1. Team

## Head of the team

Radu Horaud [ Research Director (DR), HdR ]

## Administrative assistant

Anne Pasteur [ Secretary (SAR) Inria ]

## Personnel Inria

Emmanuel Prados [ Research Associate (CR) ]

Peter Sturm [ Research Director (DR), HdR ]

## Research scientists (external)

Elise Arnaud [ Université Joseph Fourier Grenoble ]

Edmond Boyer [ Université Joseph Fourier Grenoble, HdR ]

## Ph.D students

Amaël Delaunoy [ INRIA grant ]

Mauricio Diaz [ Alban-EU grant ]

Pau Gargallo [ MESR grant ]

Diana Mateus [ Marie-Curie grant ]

Julien Morat [ CIFRE funding with Renault ]

Ramya Narasimha [ INRIA grant ]

Régis Perrier [ INRIA grant ]

Benjamin Petit [ INRIA grant ]

Kiran Varanasi [ INRIA grant ]

Daniel Weinland [ Marie-Curie Grant ]

Andrei Zaharescu [ Marie-Curie grant ]

## Post-doctoral fellows

Fabio Cuzzolin

Simone Gasparini

Miles Hansard

Clément Ménier

Kuk-Jin Yoon

## Technical staff

Hervé Mathieu [ Research Engineer (IR) ]

Florian Geffray [ Until August 2007 ]

Bertrand Holveck [ Development Engineer ]

David Knossow [ Development Engineer ]

## Visitors

Wonwoo Lee [ Visiting PhD student ]

Olivier Koch [ Visiting MIT PhD student ]

Josechu Guerrero [ University of Zaragoza, Spain ]

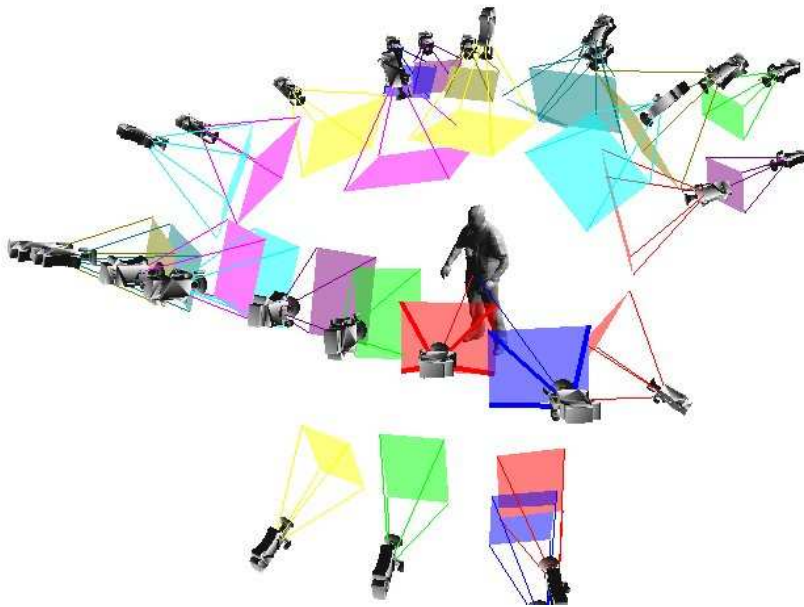
Carlos Torre Ferrero [ University of Cantabria, Spain ]

Dana Cobzas [ University of Alberta, Canada ]

Martin Jägersand [ University of Alberta, Canada ]

# 2. Overall Objectives

## 2.1. Introduction



*Figure 1. One of our main goals is to develop methods for building realistic 3-D visual content from multiple images and multiple videos. A set of 30 cameras that are precisely calibrated and finely synchronized (top) gather images and videos which are subsequently processed to produce highly realistic visual content, such as the 3-D texture mapped onto a surface (a mesh).*

The overall objective of the PERCEPTION research team is to develop theories, models, methods, and systems in order to allow computers to see and to understand what they see. A major difference between classical computer systems and computer vision systems is that while the former are guided by sets of mathematical and logical rules, the latter are governed by the laws of nature. It turns out that formalizing interactions between an artificial system and the physical world is a tremendously difficult task.

A first objective is to be able to gather images and videos with one or several cameras, to calibrate them, and to extract 2D and 3D geometric information from these images and videos, Figure 1. This is an extremely difficult task because the cameras receive light stimuli and these stimuli are affected by the complexity of the objects (shape, surface, color, texture, material) composing the real world. The interpretation of light in terms of geometry is also affected by the fact that the three dimensional world projects onto two dimensional images and this projection alters the Euclidean nature of the observed scene.

A second objective is to analyse articulated and moving objects. The real world is composed of rigid, deformable, and articulated objects. Solutions for finding the motion fields associated with deformable and articulated objects (such as humans) remain to be found. It is necessary to introduce prior models that encapsulate physical and mechanical features as well as shape, aspect, and behaviour. The ambition is to describe complex motion as “events” at both the physical level and at the semantic level.

A third objective is to describe and interpret images and videos in terms of objects, object categories, and events. In the past it has been shown that it is possible to recognize a single occurrence of an object from a single image. A more ambitious goal is to recognize object classes such as people, cars, trees, chairs, etc., as well as events or *objects evolving in time*. In addition to the usual difficulties that affect images of a single object there is also the additional issue of the variability within a class. The notion of statistical shape must be introduced and hence statistical learning should be used. More generally, learning should play a crucial role and the system must be designed such that it is able to learn from a small training set of samples. Another goal is to investigate how an object recognition system can take advantage from the introduction of non-visual input such as semantic and verbal descriptions. The relationship between images and meaning is a great challenge.

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

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

## 2.2. Highlights of the year

### 2.2.1. 4D View Solutions SAS

Five members of PERCEPTION (Florian Geffray, Clément Menier, Edmond Boyer, Hervé Mathieu, and Radu Horaud) co-founded the start-up company 4D View Solutions SAS (<http://www.4dviews.com>) whose CEO is Richard Broadbridge. The company will commercialize the multiple-camera technology developed by several team members.

### 2.2.2. Markerless 3D interactions

In collaboration with two other teams (MOAIS and EVASION) we developed a multi-camera multi-PC platform that combines computer vision with physical simulation and distributed computing to move one step towards the next generation of virtual reality applications. The platform allows a new form of immersive experience: *Put any object into the interaction space, and it is instantaneously modeled in 3D and injected into a virtual world populated with solid and soft objects. Push them, catch them, and squeeze them....* This platform was presented at SIGGRAPH'07 (Exhibition on emerging technologies) and at INRIA's 40th anniversary celebration at Lille - Palais des Congres.

## 3. Scientific Foundations

### 3.1. The geometry of multiple images

Computer vision requires models that describe the image creation process. An important part (besides e.g. radiometric effects), concerns the geometrical relations between the scene, cameras and the captured images, commonly subsumed under the term “multi-view geometry”. This describes how a scene is projected onto an image, and how different images of the same scene are related to one another. Many concepts are developed and expressed using the tool of projective geometry. As for numerical estimation, e.g. structure and motion calculations, geometric concepts are expressed algebraically. Geometric relations between different views can for example be represented by so-called matching tensors (fundamental matrix, trifocal tensors, ...). These tools and others allow to devise the theory and algorithms for the general task of computing scene structure and camera motion, and especially how to perform this task using various kinds of geometrical information: matches of geometrical primitives in different images, constraints on the structure of the scene or on the intrinsic characteristics or the motion of cameras, etc.

### 3.2. The photometry component

In addition to the geometry (of scene and cameras), the way an image looks like depends on many factors, including illumination, and reflectance properties of objects. The reflectance, or "appearance", is the set of laws and properties which govern the radiance of the surfaces. This last component makes the connections between the others. Often, the “appearance” of objects is modeled in image space, e.g. by fitting statistical models, texture models, deformable appearance models (...) to a set of images, or by simply adopting images as texture maps.

Image-based modelling of 3D shape, appearance, and illumination is based on prior information and measures for the coherence between acquired images (data), and acquired images and those predicted by the estimated model. This may also include the aspect of temporal coherence, which becomes important if scenes with deformable or articulated objects are considered.

Taking into account changes in image appearance of objects is important for many computer vision tasks since they significantly affect the performances of the algorithms. In particular, this is crucial for feature extraction, feature matching/tracking, object tracking, 3D modelling, object recognition etc.

### 3.3. Shape Acquisition

Recovering shapes from images is a fundamental task in computer vision. Applications are numerous and include, in particular, 3D modeling applications and mixed reality applications where real shapes are mixed with virtual environments. The problem faced here is to recover shape information such as surfaces, point positions, or differential properties from image information. A tremendous research effort has been made in the past to solve this problem and a number of partial solutions had been proposed. However, a fundamental issue still to be addressed is the recovery of full shape information over time sequences. The main difficulties are precision, robustness of computed shapes as well as consistency of these shapes over time. An additional difficulty raised by real-time applications is complexity. Such applications are today feasible but often require powerful computation units such as PC clusters. Thus, significant efforts must also be devoted to switch from traditional single-PC units to modern computation architectures.

### 3.4. Motion Analysis

The perception of motion is one of the major goals in computer vision with a wide range of promising applications. A prerequisite for motion analysis is motion modelling. Motion models span from rigid motion to complex articulated and/or deformable motion. Deformable objects form an interesting case because the models are closely related to the underlying physical phenomena. In the recent past, robust methods were



developed for analysing rigid motion. This can be done either in image space or in 3D space. Image-space analysis is appealing and it requires sophisticated non-linear minimization methods and a probabilistic framework. An intrinsic difficulty with methods based on 2D data is the ambiguity of associating a multiple degree of freedom 3D model with image contours, texture and optical flow. Methods using 3D data are more relevant with respect to our recent research investigations. 3D data are produced using stereo or a multiple-camera setup. These data (surface patches, meshes, voxels, etc.) are matched against an articulated object model (based on cylindrical parts, implicit surfaces, conical parts, and so forth). The matching is carried out within a probabilistic framework (pair-wise registration, unsupervised learning, maximum likelihood with missing data).

Challenging problems are the detection and segmentation of multiple moving objects and of complex articulated objects, such as human-body motion, body-part motion, etc. It is crucial to be able to detect motion cues and to interpret them in terms of moving parts, independently of a prior model. Another difficult problem is to track articulated motion over time and to estimate the motions associated with each individual degree of freedom.

### **3.5. Multiple-camera acquisition of visual data**

Modern computer vision techniques and applications require the deployment of a large number of cameras linked to a powerful multi-PC computing platform. Therefore, such a system must fulfill the following requirements: The cameras must be synchronized up to the millisecond, the bandwidth associated with image transfer (from the sensor to the computer memory) must be large enough to allow the transmission of uncompressed images at video rates, and the computing units must be able to dynamically store the data and/to process them in real-time.

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

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

## **4. Application Domains**

### **4.1. 3D modeling and rendering**

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

### **4.2. Mixed and Augmented Reality**

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

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

### **4.3. Human Motion Capture and Analysis**

We are particularly interested in the capture and analysis of human motion, which consists in recovering the motion parameters of the human body and/or human body parts, such as the hand. In the past researchers have concentrated on recovering constrained motions such as human walking and running. We are interested in recovering unconstrained motion. The problem is difficult because of the large number of degrees of freedom, the small size of some body parts, the ambiguity of some motions, the self-occlusions, etc. Human motion capture methods have a wide range of applications: human monitoring, surveillance, gesture analysis, motion recognition, computer animation, etc.

### **4.4. Multi-media and interactive applications**

The employment of advanced computer vision techniques for media applications is a dynamic area that will benefit from scientific findings and developments. There is a huge potential in the spheres of TV and film productions, interactive TV, multimedia database retrieval, and so forth.

Vision research provides solutions for real-time recovery of studio models (3D scene, people and their movements, etc.) in realistic conditions compatible with artistic production (several moving people in changing lighting conditions, partial occlusions). In particular, the recognition of people and their motions will offer a whole new range of possibilities for creating dynamic situations and for immersive/interactive interfaces and platforms in TV productions. These new and not yet available technologies involve integration of action and gesture recognition techniques for new forms of interaction between, for example, a TV moderator and virtual characters and objects, two remote groups of people, real and virtual actors, etc.

### **4.5. Car driving technologies**

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

### **4.6. Defense technologies**

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

## 5. Software

### 5.1. Platforms

#### 5.1.1. *The Grimage platform*

The Grimage platform is an experimental laboratory dedicated to multi-media applications of computer vision. It hosts a multiple-camera system connected to a PC cluster, as well as to a multi-video projection system. This laboratory is shared by several research groups, most prominently PERCEPTION and MOAIS. In particular, Grimage allows challenging real-time immersive applications based on computer vision and interactions between real and virtual objects, Figure 2.



Figure 2. The Grimage platform allows immersive/interactive applications such as this one. The real character is reconstructed in real-time and immersed in a virtual world, such that he/she can interact with virtual objects.

#### 5.1.2. *The mini-Grimage platform*

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

### 5.2. 3-D surface reconstruction based on mesh evolution

During this year we started to develop TransforMesh, a software that deals with the problem of 3-D reconstruction. More specifically, TransforMesh deals with the problem of topology changes during mesh evolution. This software is a continuation of our efforts to provide an end-to-end 3-D reconstruction chain from multiple cameras.

### 5.3. Dense voxel registration

This software package deals with the problem of registering two sets of voxels. Therefore, it takes as input two graphs describing the two sets of voxels and produces as output a one-to-one correspondence between the nodes (voxels) of these two graphs. The software is associated with our shape registration method.

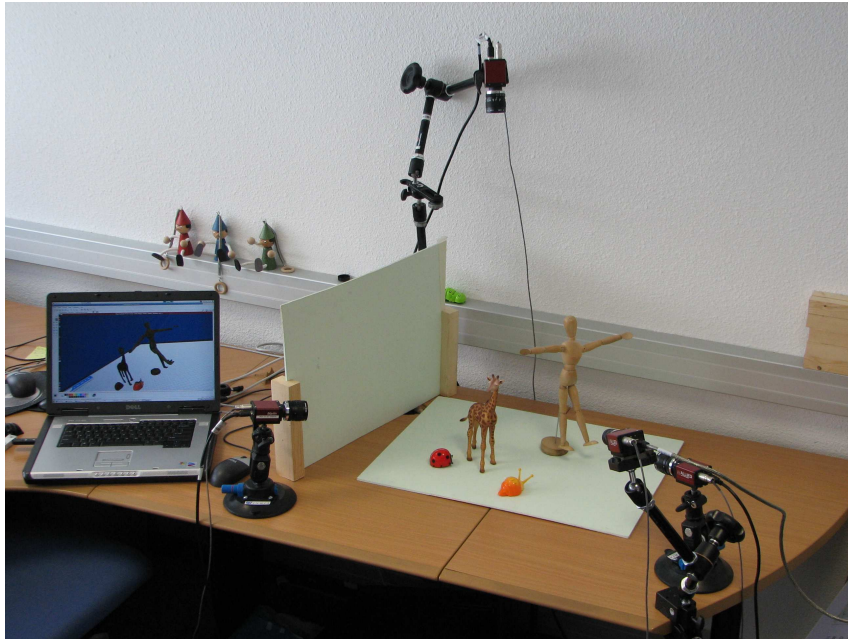


Figure 3. The mini-Grimage platform holds on a table top. It uses six cameras connected to six mini-PCs and to a laptop.

## 5.4. Real-time shape acquisition and visualization

We continued to develop a complete software package, from multiple-camera acquisition of video sequences to 3-D shape reconstruction and realistic visualization. We maintain two version of this software: an on-line real-time version that runs on a PC-cluster, and an off-line version that runs on a standard workstation/PC. In particular the software is used both by the Grimage and mini-Grimage platforms. A typical Grimage configuration is composed of 10-20 1Mpixel cameras and 10-30 PCs. The software packages include: frame-rate, uncompressed multi-camera image acquisition, image segmentation (silhouette extraction), 3-D shape reconstruction, 3-D rendering, and display using several projectors for high screen resolutions.

## 6. New Results

### 6.1. Computational modelling of binocular vision

We investigate computational stereopsis from the point of view of biological plausibility. So far we concentrated onto two topics: the control of eye movements for achieving binocular gaze and the relationship between gaze control, epipolar geometry, and binocular correspondence [44], [45].

Binocular image-pairs contain information about the three-dimensional structure of the visible scene, which can be recovered by the identification of corresponding points. However, the resulting disparity field also depends on the orientation of the eyes. If it is assumed that the exact eye-positions cannot be obtained from oculomotor feedback, then the gaze parameters must also be recovered from the images, in order to properly interpret the retinal disparity field.

Existing models of biological stereopsis have addressed this issue independently of the binocular-correspondence problem. It has been correctly assumed that *if* the correspondence problem can be solved, then the disparity field can be decomposed into gaze and structure components, as described above. In this work we take a different approach; we emphasize that although the complete point-wise disparity field is sufficient for gaze estimation, it is not in fact *necessary*. We show that the gaze parameters can be recovered directly from the images, independently of the point-wise correspondences.

The relationship between binocular vergence and the resulting epipolar geometry is derived. Our algorithm is then based on the simultaneous representation of all epipolar geometries that are feasible with respect to a fixating oculomotor system. This is done in an essentially two-dimensional space, parameterized by azimuth and viewing-distance. We define a cost function that measures the compatibility of each geometry with respect to the observed images. The true gaze parameters are estimated by a simple voting-scheme, which runs in parallel over the parameter space. We describe an implementation of the algorithm, and show results obtained from real images.

Our algorithm requires binocular units with large receptive-fields, such as those found in area MT. The model is also consistent with the finding that depth-judgments can be biased by microstimulation in MT; if the artificial signal generates an ‘incorrect’ set of gaze parameters, then we would expect the subsequent interpretation of the disparity field to be biased. Our model could be tested using binocular stimuli based on the *patterns* of disparity that we describe. We note that these patterns are geometrically analogous to parametric motion fields. It has already been shown that such flow-fields are effective stimuli for motion-sensitive cells in area MST; we predict an analogous binocular ‘gaze-tuning’ in the extrastriate cortex.

## 6.2. Audio-visual perception

This work takes place in the context of the POP European project and includes further collaborations with researchers from University of Sheffield, UK. The context is that of multi-modal sensory signal integration. We focus on audio-visual integration. Fusing information from audio and video sources has resulted in improved performance in applications such as tracking. However, crossmodal integration is not trivial and requires some cognitive modelling because at a lower level, there is no obvious way to associate depth and sound sources. Combining our expertise with expertise both from project-team MISTIS and from the University of Sheffield’s Speech and Hearing Group, we address the difficult problems of integrating spatial and temporal audio-visual stimuli using a geometrical and probabilistic framework and attack the problem of associating sensorial descriptions with representation of prior knowledge.

### 6.2.1. Multi-speaker localization

First, we address the problem of speaker localization within an unsupervised model-based clustering framework. Both auditory and visual observations are available. We gather observations over a time interval  $[t_1, t_2]$ . We assume that within this time interval the speakers are static so that each speaker can be described by its 3-D location in space. A cluster is associated with each speaker. In practice we consider  $N + 1$  possible clusters corresponding to the addition of an extra outlier category to the  $N$  speakers.

We then consider a set of  $M$  visual observations. Each such observation corresponds to a binocular disparity. Note that such a binocular disparity corresponds to the location of a physical object that is visible in both the left and right images of the stereo pair. We define a function  $v : \mathcal{R}^3 \rightarrow \mathcal{R}^3$  such that  $v(\mathbf{s}_n)$  represents the binocular disparity of speaker  $n$  when his/her location is given by  $\mathbf{s}_n$ .

Similarly, let us consider a set of  $K$  auditory observations. Each such observation corresponds to an auditory disparity, namely the *interaural time difference*, or ITD. We define a function  $u : \mathcal{R}^3 \rightarrow \mathcal{R}$  such that  $u(\mathbf{s}_n)$  evaluates the ITD of speaker  $n$  given his coordinates in the 3-D space.

We then show that recovering speakers localizations can be seen as a parameter estimation issue in a missing data framework. The parameters to be estimated are the speaker locations, and the missing variables are the assignment variables associating each individual observation to one of the  $N$  speakers or to the outlier class. We are currently investigating the use of the EM algorithm to provide these parameter estimates.

### 6.2.2. An audiovisual robot head

In collaboration with the University of Coimbra we developed an audio-visual robot head. This head is displayed on Figure 4. The head is equipped with two cameras and two microphones. It can gather binocular/binaural audio-visual data which are then processed by our algorithms. In particular, the camera's vergence control is consistent with our theoretical work on binocular vision.

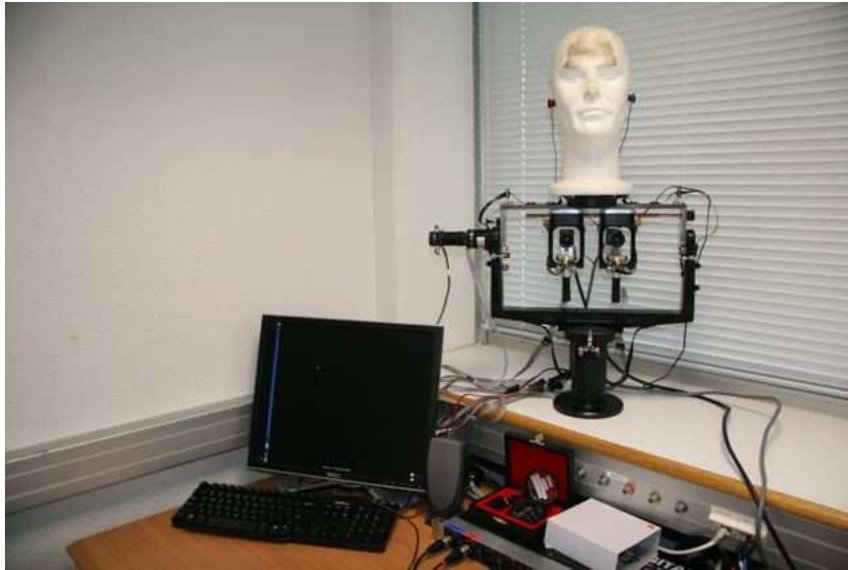


Figure 4. The audio-visual head illustrated on this figure has been developed in collaboration with the Coimbra and Sheffield universities within the POP project. The head has two rotational degrees of freedom (pand and tilt) and there is a rotational degree of freedom associated with each camera, thus allowing the control of binocular vergence.

### 6.2.3. Synchronized stereoscopic and binaural datasets with head movements

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

To achieve the acquisition of such a data collection, the following setup has been developed, Figure 5 (let us note that this setup is designed to be easily plugged with the audio-visual robot head). The audio-visual perceiver is either a person or the dummy head/torso wearing earbud microphones. The perceiver is also fitted with a helmet on which is mounted a pair of stereo cameras. On top of the head, a 4 point tracking device is attached. This has to be viewable from the tracking camera, which is to be placed above; either suspended from the ceiling or similar. The three cameras (stereo pair and tracking) are controlled with a software package and the raw image sequences are recorded on to a PC. The audio is recorded on to a laptop or PC. The three

cameras are synchronized with the audio signal using NTP network. The calibrated data collection will be freely accessible for research purposes.



Figure 5. Recently we gathered an audio-visual set of data that consists in stereo image pairs and binaural recordings. An audio-visual sensor consisting in a pair of cameras and a pair of microphones can be mounted on a person's head or on a mannequin head. In addition we recorded the position and orientation of the head.

### 6.3. Dense stereo and Markov random fields

Current approaches to dense stereo matching estimate the disparity by maximizing its a posteriori probability, given the images and the prior probability distribution of the disparity function. This is done within a Markov random field model that makes tractable the computation of the joint probability of the disparity field. In practice the problem is analogous to minimizing the energy of an interacting spin system plunged into an external magnetic field. Statistical thermodynamics provide the proper theoretical framework to model such a problem and to solve it using stochastic optimization techniques. However the latter are very slow. Alternative deterministic methods were recently used, such as deterministic annealing, mean-field approximation (see figure 6), graph cuts, and belief propagation. Basic assumptions of all these approaches are that the two images are properly rectified (such that the epipolar lines coincide with the image rows, that the illumination is homogeneous and the surfaces are lambertian (such that corresponding pixels have identical intensity values), and that there are not too many occluded or half-occluded surfaces.

We started to investigate the link between intensity-based stereo and contour-based stereo. In particular, we want to properly describe surface-discontinuity contours for both piecewise planar objects and objects with smooth surfaces, and to inject these contours into the probabilistic framework and the associated minimization methods described above.

In particular, we carry out cooperatively both disparity and object boundary estimations by setting the two tasks in a unified Markovian framework. We define an original joint probabilistic model that allows to estimate disparities through a Markov random field model. Boundary estimation is then not reduced to a second independent step but cooperates with disparity estimation to gradually and jointly improve accuracy. The feedback from boundary estimation to disparity estimation is made through the use of an additional auxiliary field referred to as a displacement field. This field suggests the corrections that need to be applied at disparity discontinuities in order that they align with object boundaries. The joint model reduces to a Markov random field model when considering disparities while it reduces to a Markov chain when focusing on the displacement field. The performance of our approach is illustrated on real stereo images sets, demonstrating the power of this cooperative framework.

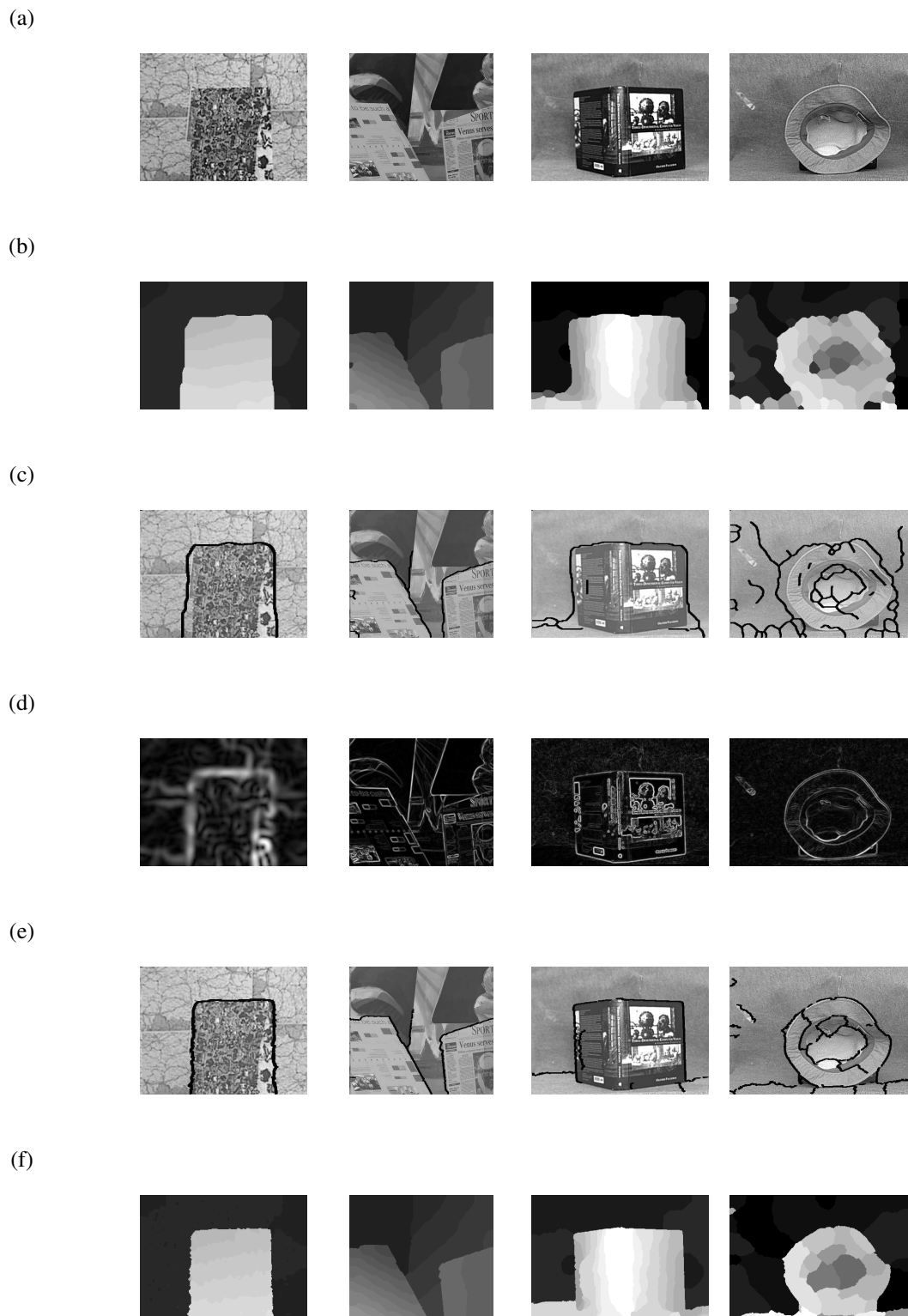


Figure 6. (a) Original Images; (b) Disparity maps using Variational Mean-field only; (c) Disparity discontinuities corresponding to (b); (d) Image Gradient maps; (e) Corrected disparity discontinuities using our approach; (f) Corresponding disparity maps.



## 6.4. Human-body tracking and human-motion capture

We address the problem of articulated object tracking using either 2-D features or 3-D features. In both cases we use a multiple-camera setup along the lines described above.

### 6.4.1. *Human-body tracking using an implicit surface, 3D points, and surface normals.*

We developed a new method for tracking human motion based on fitting an articulated implicit surface to 3-D points and normals, Figure 7. There are two important contributions of this work to the state of the art. First, we introduce a new distance between an observation (a point and a normal) and an ellipsoid. We show that this can be used to define an implicit surface as a blending over a set of ellipsoids which are linked together to form a kinematic chain. Second, we exploit the analogy between the distance from a set of observations to the implicit surface and the negative log-likelihood of a mixture of Gaussian distributions. This allows us to cast the problem of implicit surface fitting into the problem of maximum likelihood estimation with missing variables. We argue that outliers are best described by a uniform component that is added to the mixture, and we formally derive the associated EM algorithm.

Casting the data-to-model association problem into unsupervised clustering has already been addressed in the past within the framework of point registration. We appear to be the first to apply EM clustering to the problem of fitting a blending of ellipsoids to a set of 3-D observations and to explicitly model outliers within this context.

### 6.4.2. *Human-body tracking using the kinematics of extremal contours.*

We also address the problem of human motion tracking from 2-D features available with image sequences [46], [35], [31]. The human body is described by an articulated mechanical chain and human body-parts are described by volumetric primitives with curved surfaces.

An extremal contour appears in an image whenever a curved surface turns smoothly away from the viewer. We have developed a method that relies on a kinematic parameterization of such extremal contours. The apparent motion of these contours in the image plane is a function of both the rigid motion of the surface and the relative position and orientation of the viewer with respect to the curved surface. The method relies onto the following key features: A parameterization of an extremal-contour point, and its associated image velocity, as a function of the motion parameters of the kinematic chain associated with the human body; The zero-reference kinematic model and its usefulness for human-motion modelling; The chamfer-distance used to measure the discrepancy between predicted extremal contours and observed image contours; Moreover the chamfer distance is used as a differentiable multi-valued function and the tracker based on this distance is cast in an optimization framework. We have implemented a practical human-body tracker that may use an arbitrary number of cameras. One great methodological and practical advantage of our method is that it relies neither on model-to-image, nor on image-to-image point matches. In practice we model people with 5 kinematic chains, 19 volumetric primitives, and 54 degrees of freedom; We observe silhouettes in images gathered with several synchronized and calibrated cameras. An output of the method and a comparison with the wellknown VICON system can be seen on Figure 8.

### 6.4.3. *Sequential Monte Carlo Inverse Kinematics.*

We proposed a new and original approach to solve the inverse kinematics problem. Our approach has the advantages to avoid the classical pitfalls of numerical inversion methods such as singularities and to accept arbitrary types of constraints. As shown fig 9 – where we compared the average time per iteration of two numerical IK solutions (the Jacobian transpose method and the damped pseudo-inverse methods) and our method – our approach exhibits a linear complexity with respect to degrees of freedom which makes it far more efficient for articulated figures with a high number of degrees of freedom. Our framework is based on Sequential Monte Carlo Methods that were initially designed to filter highly non-linear, non-Gaussian dynamic systems. They are used here in an online motion control algorithm that allows to integrate motion priors. The effectiveness of our method is shown fig 10 for a human figure animation application and fig 11 for an exemple of hand animation. Future work will consist in integrating measurements from image sequences to constraint the algorithm.

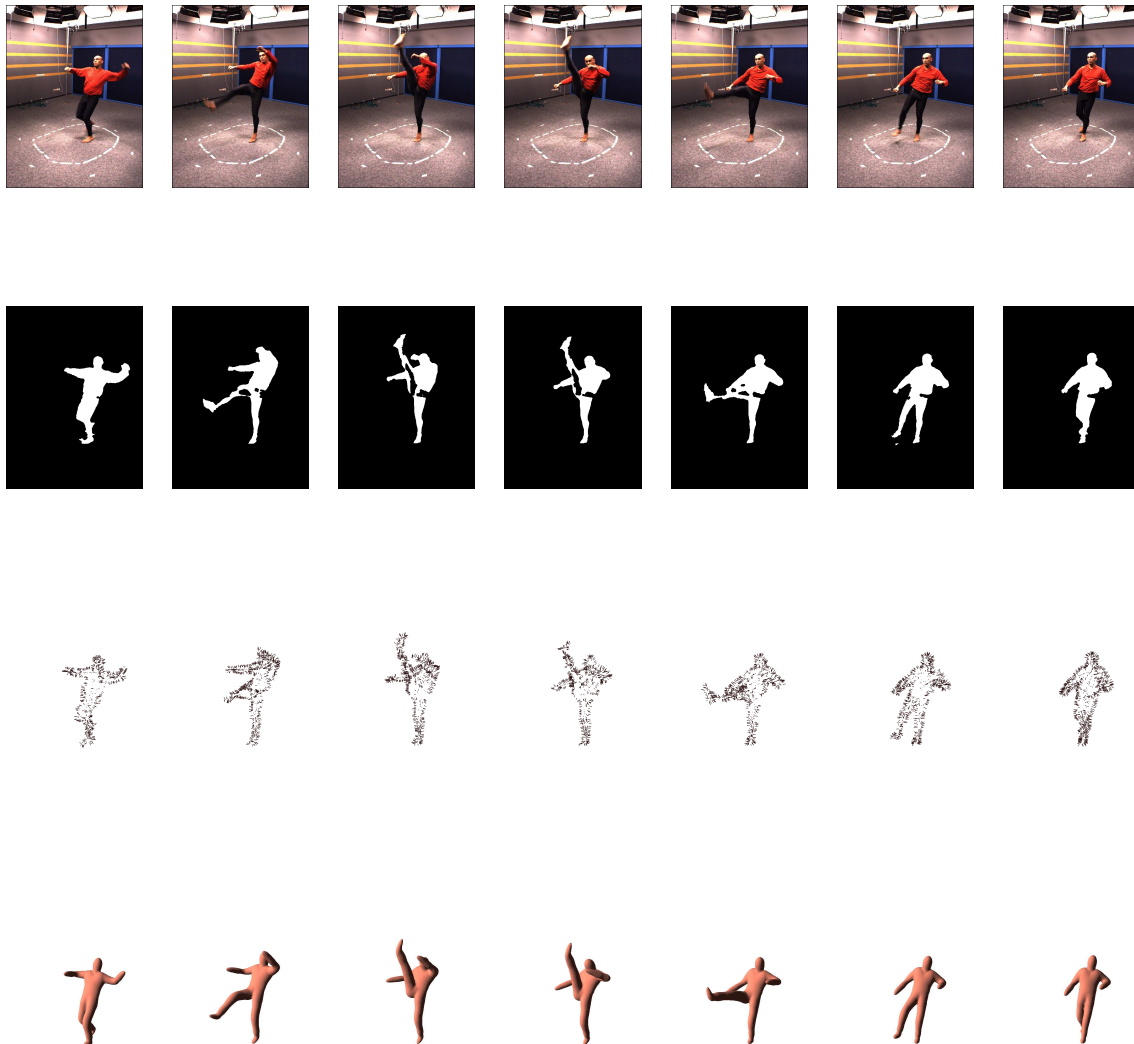


Figure 7. An image sequence of the first camera (this experiment used six cameras). The silhouettes extracted from these images. The 3-D visual hull (points and surface normals). The implicit surface fitted to the data using points and normals.

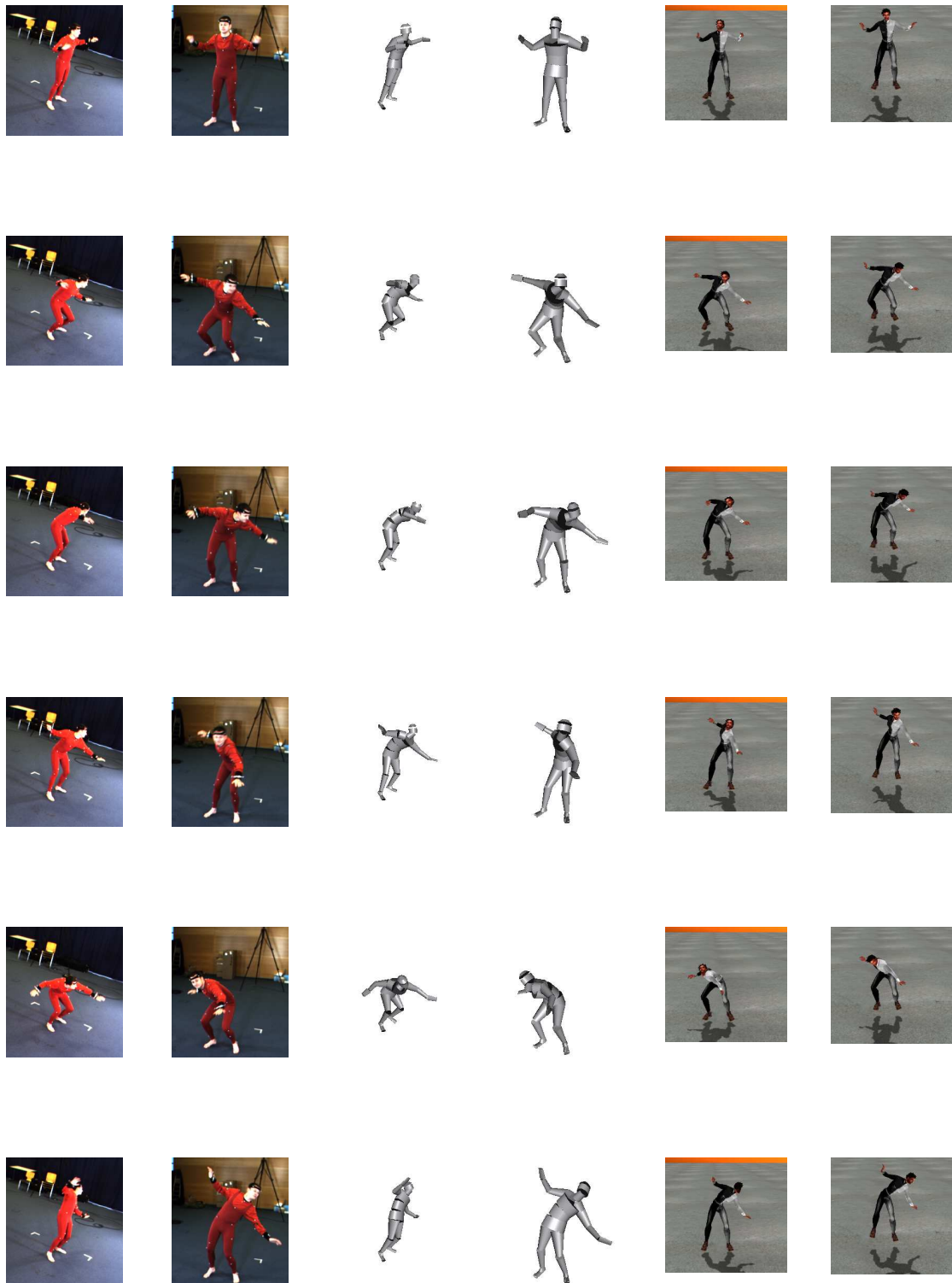


Figure 8. This figure shows a qualitative comparison between markerless and marker-based motion capture. The two image sequences shown on the left columns (together with four other image sequences which are not shown here) produced the results shown on the third, fourth and fifth columns. The last column shows the result obtained from a method that uses a 8-camera VICON system to locate body markers.

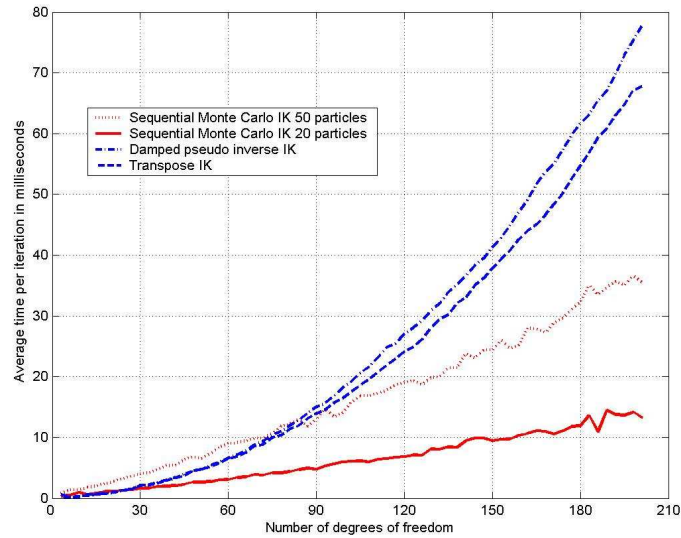


Figure 9. **Performances of our method** Comparison with state-of-the-art IK methods. One can effectively see the linear nature of the complexity of our method (instead of exponential with numeric IK).

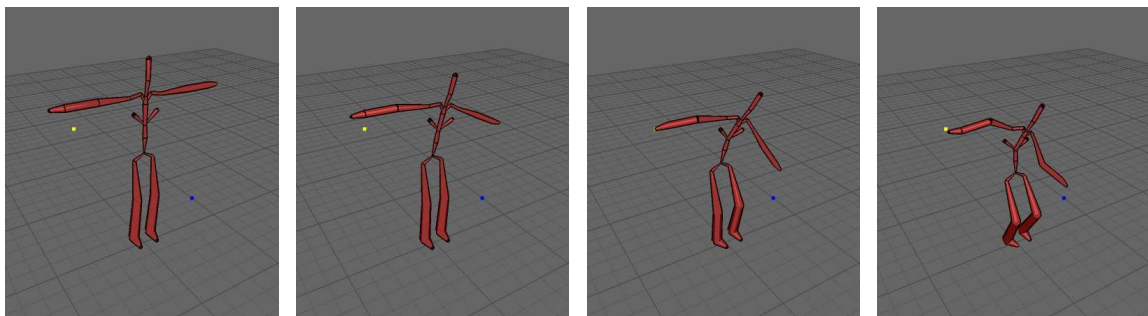


Figure 10. **human figure animation** In this animation, feet are constrained to lie on the floor, the right hand is linked with the yellow dot while the left arm has the blue dot as target. Notice how the knees bend for the task to be achieved

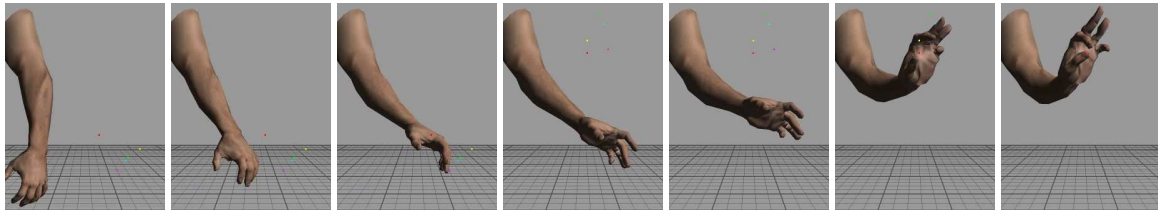


Figure 11. **Hand animation** In this animation the fingers were given a target position represented as colored dots in the images. The two strips correspond to two different tasks that were chained along the animation.

## 6.5. Multiple camera reconstruction

### 6.5.1. Point-based reconstruction using robust factorization

The problem of 3-D reconstruction from multiple images is central in computer vision. Bundle adjustment provides a general method and practical algorithms for solving this reconstruction problem using maximum likelihood. Nevertheless, bundle adjustment is non-linear in nature and sophisticated optimization techniques are necessary, which in turn require proper initialization. Moreover, the combination of bundle adjustment with robust statistical methods to reject outliers is not clear both from the points of view of convergence properties and of efficiency.

We addressed the problem of building a class of robust factorization algorithms that solve for the shape and motion parameters (i.e., 3-D reconstruction) with both affine (weak perspective) and perspective camera models. We introduce a Gaussian/uniform mixture model and its associated EM algorithm. This allows us to address robust parameter estimation with an unsupervised clustering approach. We devise both an affine factorization algorithm and an iterative perspective factorization algorithm which are robust in the presence of a large number of outliers. We carry out numerous experiments to validate our algorithms and to compare them with existing ones. We also compare our approach with factorization methods that use M-estimators.

This work is part of Andrei Zaharescu's PhD. A paper was recently submitted to the International Journal of Computer Vision.

### 6.5.2. Surface reconstruction based on mesh evolution

The point-based reconstruction algorithm just described provides sparse 3-D points that are impractical for rendering. Nevertheless, they can be used to build a rough mesh. We developed a method that starts with such a rough description and which consists in an evolution towards a very accurate description.

Most of the algorithms dealing with image based 3-D reconstruction involve the evolution of a surface based on a minimization criterion. The mesh parametrization, while allowing for an accurate surface representation, suffers from the inherent problems of not being able to reliably deal with self-intersections and topology changes. As a consequence, an important number of methods choose implicit representations of surfaces, e.g. level set methods, that naturally handle topology changes and intersections. Nevertheless, these methods rely on space discretizations, which introduce an unwanted precision-complexity trade-off. In this paper we explore a new mesh-based solution that robustly handles topology changes and removes self intersections, therefore overcoming the traditional limitations of this type of approaches. To demonstrate its efficiency, we present results on 3-D surface reconstruction from multiple images and compare them with state-of-the art results, [60] and Figure 13.

### 6.5.3. Multi-view stereo

We have addressed the problem of image-based surface reconstruction. The main contribution is the computation of the exact derivative of the reprojection error functional [42]. This allows its rigorous minimization via

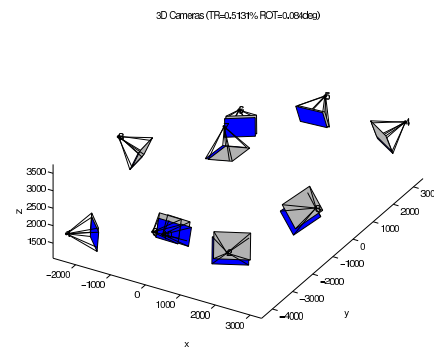
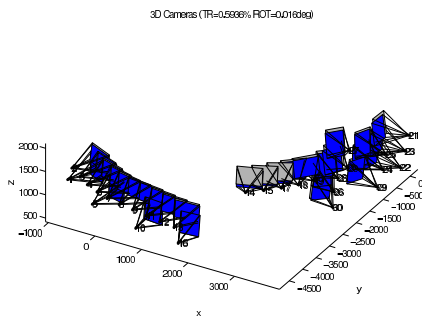
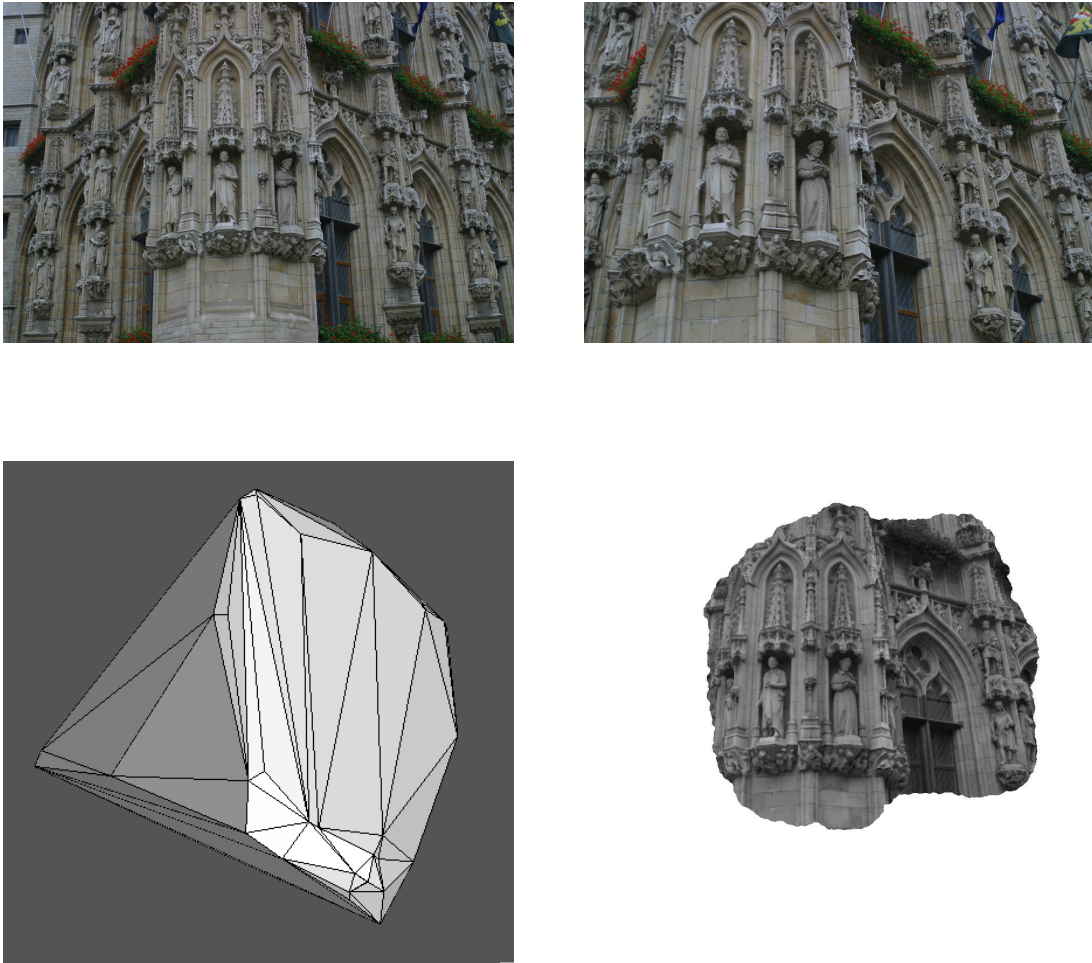


Figure 12. Robust perspective factorization applied to multiple-camera calibration. (Top-left): Partial view of a camera setup. (Top-right): The calibration data are gathered by moving a one-dimensional object in the common field of view of the cameras. (bottom): Final camera setups found by the algorithm.



*Figure 13. Transfomes is a method that starts with a rough mesh and produces a very detailed representation of a surface. In this example we show two raw images (out of a set of eight images), the starting mesh obtained from a sparse set of 3-D points, and the final reconstructed model.*

gradient descent surface evolution. The main difficulty has been to correctly take into account the visibility changes that occur when the surface moves. A geometric and analytical study of these changes is presented and used for the computation of the derivative.

Our analysis shows the strong influence that the movement of the contour generators (or “horizons”, see fig. 14) has on the reprojection error. As a consequence, during the proper minimization of the reprojection error, the contour generators of the surface are automatically moved to their correct location in the images. Therefore, current methods adding additional silhouettes or apparent contour constraints to ensure this alignment can now be understood and justified by a single criterion: the reprojection error.

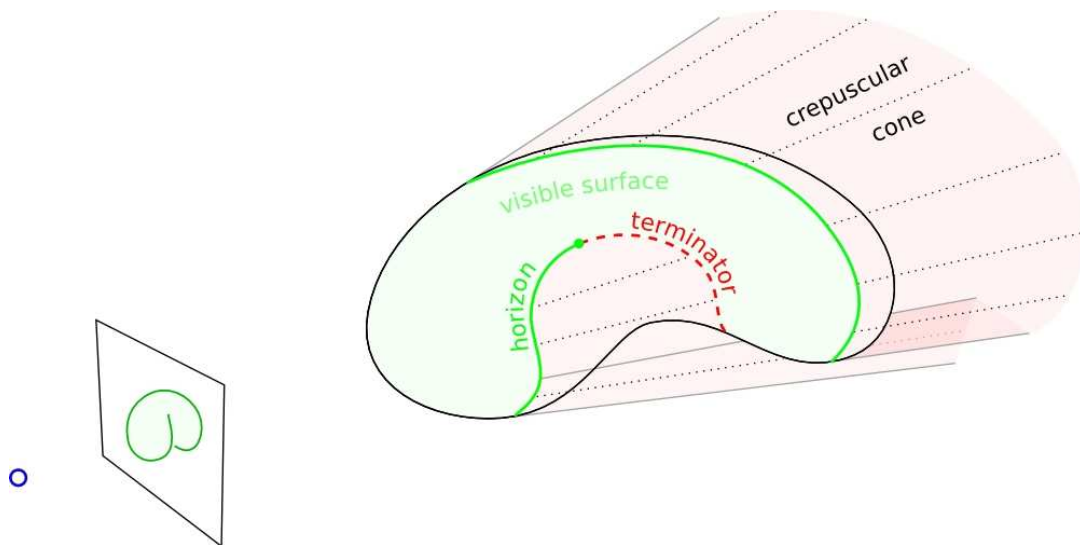


Figure 14. The banana shape seen from a vantage point. The horizon is drawn with green lines and the terminators with a dashed red line. The viewing ray segments are the crepuscular rays that form the crepuscular cone.

The impact of the proper handling of the visibility is proved in fig. 15.

The *balls* dataset (fig. 15) consists of 20 images of three balls floating above a plane. There is no texture or shading in any part of the scene. Therefore, the only information present in the images are the apparent contours. In addition, because of self-occlusions between the balls and the plane, the silhouettes of the foreground are not sufficient to distinguish that the balls are three separate objects. If you do not properly handle with the visibility (i.e. if you do as people do until now), then the minimization algorithm does not separate the balls during the evolution and, due to the lack of texture, does shrink and disappear. The shrinkage does happen even when initializing from the ground truth. The result displayed in the top-right of fig. 15 is the one computed when we properly handle with the visibility.

The *bowl* scene (fig. 15) contains a green ball inside a yellow bowl with Lambertian shading. The execution with the full flow, correctly recovered the concavity of the bowl and the shape of the ball. The execution using only the horizon term did not carve the concavity at all. The execution with the interior term, did carve the concavity, but not completely, keeping the ball and the bowl linked together. This shows how the interior and the horizon terms worked together, the first one carving the concavity and the second one enforcing the apparent contour of the ball on the images.

For more details see [42].



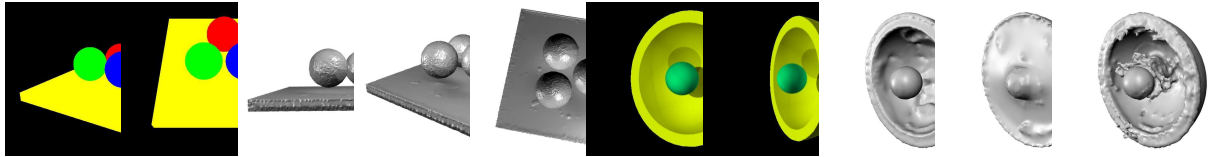


Figure 15. **Top:** two input images of the balls dataset and three renderings of the reconstruction obtained with the horizon term. **Bottom:** two input images of the bowl dataset and the reconstruction obtained by the full term, the horizon term and the interior term.

Another work in this area concerns the joint consideration of depth information and occupancy of space, i.e. which points are inside an object and which outside [43]. These two types of information are redundant, but considering them both explicitly, brings about two advantages. First, unlike other occupancy based models, it explicitly models the deterministic relationship between occupancy and depth and thus, it correctly handles occlusions. Second, unlike depth based approaches, determining depth from the occupancy automatically ensures the coherence of the resulting depth maps associated with different images.

## 6.6. Image-based modeling of reflectance properties

We develop a variational method to recover both the shape and the reflectance of a scene surface(s) using multiple images, assuming that illumination conditions are fixed and known in advance. Scene and image formation are modeled with known information about cameras and illuminants, and scene recovery is achieved by minimizing a global cost functional with respect to both shape and reflectance. Unlike most previous methods recovering only the shape of Lambertian surfaces, the proposed method considers general dichromatic surfaces. For more detail see [62], [58].

## 6.7. Motion Segmentation

### 6.7.1. Spectral clustering of motion trajectories

Recent progress in the acquisition of 3-D data from multi-camera setups opened a new way of looking at motion analysis. This work proposes a solution to the motion segmentation in the context of sparse scene flow. In particular, our interest focuses on the disassociation of motions belonging to different rigid objects, starting from the 3-D trajectories of features lying on their surfaces. We analyze these trajectories and propose a representation suitable for defining robust pairwise similarity measures between trajectories and handling missing data. The motion segmentation is treated as graph a multi-cut problem, and solved with spectral clustering techniques (two algorithms are presented). Experiments are done over simulated and real data in the form of sparse scene-flow; we also evaluate the results on trajectories from motion capture data. A discussion is provided on the results for each algorithm, the parameters and the possible use of these results in motion analysis [50].

### 6.7.2. Unsupervised articulated motion segmentation

We developed a novel tool for body-part segmentation and tracking in the context of multiple camera systems. Our goal is to produce robust motion cues over time sequences, as required by human motion analysis applications. Given time sequences of 3D body shapes, body-parts are consistently identified over time without any supervision or a priori knowledge. The approach first maps shape representations of a moving body to an embedding space using locally linear embedding. While this map is updated at each time step, the shape of the embedded body remains stable. Robust clustering of body parts can then be performed in the embedding space by k-wise clustering, and temporal consistency is achieved by propagation of cluster centroids. The

contribution with respect to methods proposed in the literature is a totally unsupervised spectral approach that takes advantage of temporal correlation to consistently segment body-parts over time. Comparisons on real data are run with direct segmentation in 3D by EM clustering and ISOMAP-based clustering: the way different approaches cope with topology transitions is discussed [41]

## 6.8. Articulated shape matching

Matching articulated shapes described as clouds of 3-D points reduces to maximal sub-graph isomorphism when representing each set of points as a weighted graph. Spectral graph theory can be used to map these graphs onto lower dimensional isometric spaces and match shapes by aligning their embeddings in virtue of their invariance to change of pose. Classical graph isomorphism schemes relying on the ordering of the eigenvalues to align Laplacian eigenvectors fail when handling large data-sets or noisy data. We derive a new formulation equivalent to finding the best alignment between two congruent K-D sets of points, where the dimension K of the embedded space results from the selection of the best subset of eigenfunctions of the Laplacian operator. This set is detected by matching the signatures of those eigenfunctions expressed as histograms, and provides a smart initialization for the alignment problem with a considerable impact on the overall performance. Dense matching then reduces to embedded point registration under orthogonal transformations, a task we cast into the framework of unsupervised clustering and solve using the EM algorithm. Maximal subset matching of non identical shapes is handled by defining an appropriate outlier class. Experimental results on challenging examples show how the algorithm naturally treats changes of topology, shape variations and different sampling densities, Figure 16 and [49], [48].

## 6.9. Temporal surface tracking

Tracking the surface of moving objects is of central importance when modeling dynamic scenes using multiple videos. This key step in the modeling pipeline yields temporal correspondences which in turn allows recovery of improved and consistent descriptions of object shapes and appearances. Furthermore, it is a necessary step for motion related applications such as motion capture.

We address the problem of capturing the evolution of a moving and deforming surface, in particular moving bodies, given multiple videos. Our approach is grounded on the observation that natural surfaces are usually arbitrary shaped and difficult to model *a priori*. In addition, shapes can significantly change or move over a time sequence. As an example, human body parts can present large motions as well as topological changes. To handle such deformations, we use meshes which are morphed from one frame to another. Like feature-based approaches, we use photometric cues provided by images and geometric cues provided by recovered meshes. However, instead of looking for a dense vertex match between 2 meshes, we use a sparse but robust match and its associated displacement vector field to drive a full mesh evolution. This is achieved by means of recent work which allows consistent mesh evolution with possible topological changes. Figure 17 illustrates

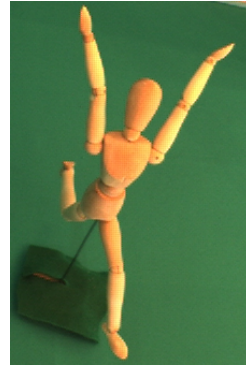
## 6.10. Action representation and recognition

Action recognition has received considerable attention over the past decades, as a result of the growing interest for automatic and advanced scene interpretations shown in several applications domains, e.g. video-surveillance or human machine interactions.

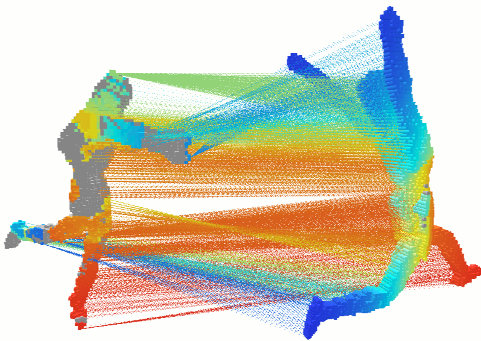
We considered the problem of recognizing actions from arbitrary sets of cameras. Our motivation comes from the observation that camera configurations for recognition are usually unknown and hence, can hardly be reproduced when learning. Thus the need for an approach that is robust to cameraconfigurations, possibly in several respects: the number of cameras and their viewpoints for instance. To this aim, we propose a new framework [57] where four dimensional action models are used to predict the observation from a single or few unknown viewpoints. To learn actions, we use three dimensional occupancy templates build from multiple viewpoints, in an exemplar-based HMM. The novelty is that three dimensional templates are not required during the recognition phase, instead learned 3D exemplars (see figure 18) are used to produce two dimensional image information that are confronted to the observations. Parameters that describe image projections are



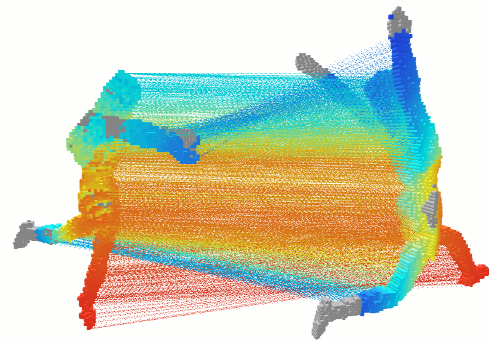
Position 1



Position 2



Initial correspondences



Final correspondences

Figure 16. This figure illustrates the articulated shape matching method under development. The method takes as input two positions of the same articulated object and established one-to-one correspondences between the two positions. More precisely, the shape is described by a graph whose nodes are voxels and whose edges are local connectivities between the voxels. Two graphs corresponding to two positions are mapped onto a lower dimensional eigenspace. The problem of graph matching is solved as a point matching problem in this reduced isometric space.

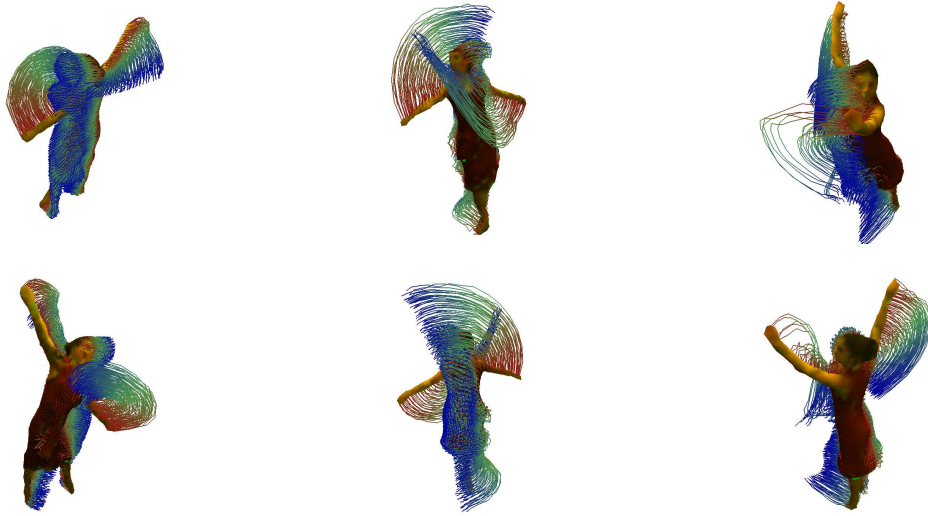


Figure 17. Temporal tracking: Results obtained on the dance-1 sequence (INRIA). Mesh point trajectories are presented in a color coded scheme where cooler colors represent earlier frames.

then added as latent variables in the recognition process. In this way, view changes are explicitly modeled, which avoids the loss of information that occurs with view invariant representations. In addition, the temporal Markov dependency applied to view parameters allows them to evolve during recognition as with a smoothly moving camera. The effectiveness of the framework was demonstrated on our real datasets and with innovative recognition scenarios.

## 6.11. Omnidirectional vision

Omnidirectional vision studies the modeling and use of cameras with very large field of view. Various technologies for achieving a large field of view exist (most common are different types of fisheye lenses and catadioptric cameras). Part of our research is concentrated on finding generic models for such cameras and algorithms for working with them. A good compromise between high generality and low complexity is our previously introduced model of generalized radial distortion. In [56], we propose an efficient algorithm for self-calibration of that model, from images of a planar scene. It allows to handle non-parametric and parametric versions of the generalized distortion model and gives very good results.

## 6.12. Other results

### 6.12.1. Conic fitting

Fitting conics to a set of 2D points is a classical problem occurring in computer vision and other areas. Practically all algorithms proposed in the literature, produce suboptimal and biased results, due to adopting cost functions that only approximately correspond to the Euclidean distance between points and conics. In [55], we describe how the Euclidean distance can be minimized, in a bundle adjustment manner, and show how this can be implemented very efficiently.

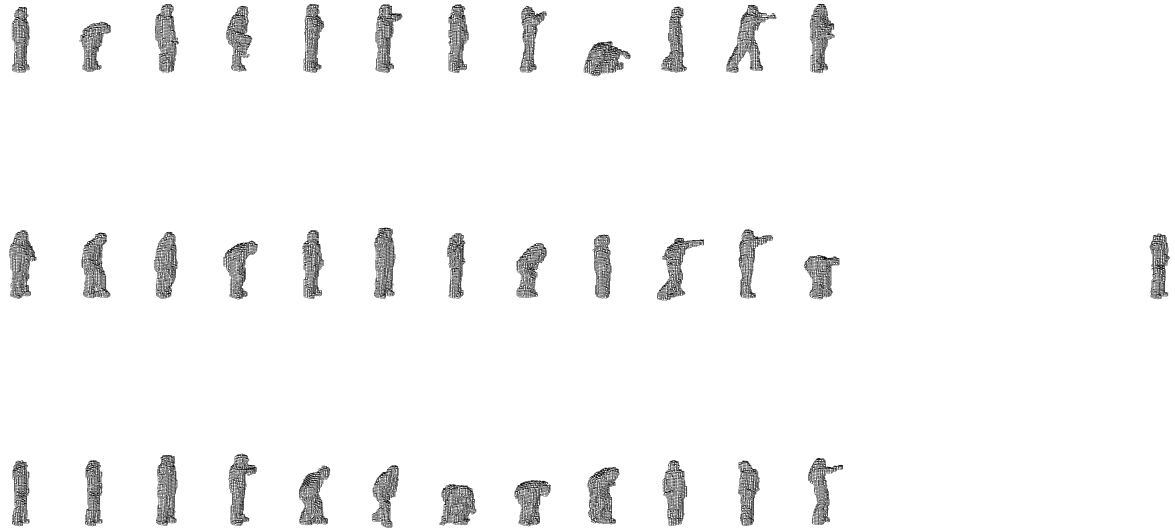


Figure 18. Action Representation: from left to right and top to bottom, 48 of the 52 exemplars used to model all the actions. The dataset is composed of 11 actions performed by 10 actors.

### 6.12.2. Automatically identifying foreground in multiple images

Identifying foreground regions in single or multiple images is a necessary preliminary step of several computer vision applications in object tracking, motion capture or 3D modeling for instance. In particular, several 3D modeling applications optimize an initial model obtained using silhouettes extracted as foreground image regions. Traditionally, foreground regions are segmented under the assumption that the background is static and known beforehand in each image. This operation is usually performed on an individual basis, even when multiple images of the same scene are considered. In the approach we developed [47], we took a different strategy and proposed a method that simultaneously extract foreground regions in multiple images without any *a priori* knowledge on the background. The interest arises in many applications where multiple images are considered and where background information are not available, for instance when a single image only is available per viewpoint.

## 7. Contracts and Grants with Industry

### 7.1. 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. ANR project CAVIAR

The global topic of the CAVIAR project (<http://www.anr-caviar.org/>) is to use omnidirectional cameras for aerial robotics. Our team implements calibration software for various kinds of omnidirectional cameras. We will develop approaches for matching images obtained with such cameras, as well as for performing camera motion estimation and 3D reconstruction of environments from them. This information is to be used for aiding an aerial robot's navigation and for 3D map generation.

This 3-year project started in December 2005. The partners are CREA (Amiens, coordinator), LAAS (Toulouse), ICARE (INRIA Sophia-Antipolis), and LE2I (Le Creusot). The current team members who are involved in this project are Peter Sturm and Simone Gasparini.

#### 8.1.2. ANR project DALIA

The project DALIA is aimed at visualizing, interacting and collaborating in heterogenous distributed environments. The main objective is to study collaborative interactive 3D applications dealing with large data sets of static nature, e.g. environments, as well as dynamic nature, e.g. a moving person. 4 partners are involved in this project: IPARLA (INRIA futur, Bordeaux), PERCEPTION and MOAIS (INRIA Rhone-Alpes). The team members involved in this project are Edmond Boyer and Benjamin Petit.

#### 8.1.3. ANR project FLAMENCO

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

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

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

The team members directly involved in this project are Peter Sturm, Emmanuel Prados (INRIA researchers) and Amael Delaunoy (PhD thesis). During 2007, they have focused on the illumination and the reflectance models.

### **8.1.4. 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.

The PERCEPTION team is concerned with the acquisition and reconstruction part of the project. The team members involved in this project are Edmond Boyer, Kiran Varanisi and Diana Mateus.

### **8.1.5. ARC–FANTASTIK**

In 3D control animation, one of the main difficulty is to take into account both the kinematic and dynamic constraints to obtain a physically plausible motion. Classical approaches are based on a global spacetime optimisation. The fact that they are both time consuming and non sequential make them difficult to use in practice. As an alternative, within this project, we propose to investigate the use of statistical tools, such as sequential Monte Carlo approaches combined with dimension reduction techniques, to the problem of motion control, where the evolution law will be defined using dynamic constraints, and the data collected from a motion capture system will constraint the solution sequentially.

The partners of this project are INRIA Rhône-Alpes (PERCEPTION and EVASION teams), the university of Bretagne Sud (équipe SAMSARA), and ENS Cachan (Centre de Mathématiques et de Leurs Applications). The team member involved in this project is Elise Arnaud.

## **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 has developed a real-time multi-camera and multi-PC system. The developments are 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 Stage Training or EST. Within these actions, VISITOR has been selected to host PhD students granted by the European commission. The PERCEPTION team actively participated in the project elaboration. Edmond Boyer is the coordinator of this project and we host 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) coordinated by Radu Horaud. This network gathers 11 partners from 11 European countries and has the ambition to address foundational issues in computational and cognitive vision systems through an European doctoral and post-doctoral program.

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

For 2007, VISIONTRAIN organized a thematic school (Computational and Neuro-physiological Models for Visual Perception, Les Houches Physics School, 25 - 30 March 2007, Les Houches France) which was attended by 75 participants, as well as two workshops at the University of Utrecht and at the Technion.

The PERCEPTION members involved in VISIONTRAIN are Radu Horaud, Andrei Zaharescu, and Fabio Cuzzolin.

#### **8.2.4. FP6 IST STREP project POP**

We are coordinators of the POP project (Perception on Purpose) involving the MISTIS and the PERCEPTION INRIA groups, as well as 4 other partners: University of Osnabruck (cognitive neuroscience), University Hospital Hamburg-Eppendorf (neurophysiology), University of Coimbra (robotics), and University of Sheffield (hearing and speech). POP's objectives are the followings:

The ease with which we make sense of our environment belies the complex processing required to convert sensory signals into meaningful cognitive descriptions. Computational approaches have so far made little impact on this fundamental problem. Visual and auditory processes have typically been studied independently, yet it is clear that the two senses provide complementary information which can help a system to respond robustly in challenging conditions. In addition, most algorithmic approaches adopt the perspective of a static observer or listener, ignoring all the benefits of interaction with the environment. This project proposes the development of a fundamentally new approach, perception on purpose, which is based on 5 principles. First, visual and auditory information should be integrated in both space and time. Second, active exploration of the environment is required to improve the audiovisual signal-to-noise ratio. Third, the enormous potential sensory requirements of the entire input array should be rendered manageable by multimodal models of attentional processes. Fourth, bottom-up perception should be stabilized by top-down cognitive function and lead to purposeful action. Finally, all parts of the system should be underpinned by rigorous mathematical theory, from physical models of low-level binocular and binaural sensory processing to trainable probabilistic models of audiovisual scenes. These ideas will be put into practice through behavioural and neuroimaging studies as well as in the construction of testable computational models. A demonstrator platform consisting of a mobile audiovisual head will be developed and its behaviour evaluated in a range of application scenarios. Project participants represent leading institutions with the expertise in computational, behavioural and cognitive neuroscientific aspects of vision and hearing needed both to carry out the POP manifesto and to contribute to the training of a new community of scientists.

#### **8.2.5. FP6-IST STREP project INTERACT**

The INTERACT project considers Human Machine Interfaces based on both speech and hand motion. The objective is the capability to manipulate virtual 3D objects using hands and speech. The resulting system will be based on computer vision techniques for the capturing hand motion and on speech recognition. 5 partners are involved in this project: PERCEPTION (INRIA Rhone-Alpes), EPTRON coordinator (Spain), Holographica (Hungary), Total-Immersion (France) and Vecsys (France).



## 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*, he is an *area editor* of *Computer Vision and Image Understanding*, and an *associated editor* of *Machine Vision Applications* and of *IET Computer Vision*.
- Edmond Boyer has been a member of the program committees of: CVPR07, ICCV07, BMVC07, CVMP07, 3DPVT07.
- Peter Sturm is a member of the editorial boards of the *Image and Vision Computing* journal and the *Journal of Computer Science and Technology*.
- Peter Sturm has been co-organizer of:
  - PACV – Workshop on Photometric Analysis For Computer Vision (in conjunction with ICCV)
  - BENCOS – ISPRS Workshop Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images (in conjunction with CVPR)
  - “Journée Thématique Vision omnidirectionnelle” of the GdR ISIS.
- Peter Sturm has been a member of the Program Committees of:
  - ICCV – IEEE International Conference on Computer Vision
  - DV – Workshop on Dynamical Vision (in conjunction with ICCV)
  - OMNIVIS – Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras (in conjunction with ICCV)
  - ACCV – Asian Conference on Computer Vision
  - BMVC – British Machine Vision Conference
  - WMVC – IEEE Workshop on Motion and Video Computing,
  - ISVC – International Symposium on Visual Computing
  - VISAPP – International Conference on Computer Vision Theory and Applications
  - ORASIS – Congrès francophone des jeunes chercheurs en vision par ordinateur
- Emmanuel Prados has been a member of the Program Committees of:
  - Organizer & General Co-Chair of PACV’07 (Photometric Analysis for Computer Vision); workshop in conjunction with ICCV’07, Rio de Janeiro, Brazil, October 14-21, 2007 ; with K. Ikeuchi, S. Soatto, P. Belhumeur and Peter Sturm.
  - a member of SSVM’07 program committee (Scale Space and Variational Methods Conference) : first joint Scale-Space and Variational Methods Conference - Ischia, Italy, May 30 - June 2, 2007.
  - the organizer of the Symposium “PDEs and image processing” in conjunction with the SciCADE 2007 (International Conference on SCIENTIFIC Computation And Differential Equations), Saint-Malo, France, 9-13 July 2007.
  - the organizer of the Symposium “Variational and PDE methods for computer vision and image processing” in conjunction with “Congrès SMAI 2007”, Praz sur Arly, France, 4-8 June 2007.

### 9.2. Services to the Scientific Community

- 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 Committee.
- Radu Horaud is the coordinator of the Visiontrain Marie Curie Research Training Network.
- Emmanuel Prados is the coordinator of the Flamenco Project (ANR-MDCA-2007-2010).
- Peter Sturm is chairing the "Commission Emplois Scientifiques" of INRIA Grenoble – Rhône-Alpes, that participates in the organization and selection of recruitment campaigns for post-docs and other positions.
- 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 Chairman of the Working Group "Géométrie et Image" of the GdR ISIS (Groupement de Recherche Information, Signal, Images et Vision).
- Peter Sturm is chairing the Committee on "Actions Incitatives", which is part of the INRIA COST – Conseil d'Orientation Scientifique et Technologique.

### 9.3. Teaching

- 3D Computer Vision, postgraduate course, University of Zaragoza, Spain, 20h, P. Sturm.
- Optimisation, M2R IVR, INPG, 6h, P. Sturm.
- Modélisation 3D à partir d'images et de vidéos, M2R INFORMATIQUE, 24h, E. Boyer and P. Sturm.
- Géométrie projective, M2R IVR, INPG, 6h, E. Boyer.
- Image retrieval, M2P, UJF, 15h, E. Arnaud.
- Computer Vision, M2P, UJF, 30h, E. Arnaud, E. Boyer
- Bayesian Networks and Graphical models, M2R, UJF, 10h, E. Arnaud
- probability, M1, UJF, 10h, E. Arnaud

### 9.4. Tutorials and invited talks

- Peter Sturm has given an invited talk on "Modélisation 3D et de l'apparence d'objets à partir d'images" at the "Journée d'étude 3D" of the association Club VISU, Montpellier, France.

### 9.5. Thesis

- Clément Menier
- David Knossow
- Edmond Boyer
- Peter Sturm acted as reviewer for the following PhD theses:
  - Etienne Mouragnon, Université Blaise Pascal, Clermont-Ferrand, 2007.
  - Christoph Strecha, Katholieke Universiteit Leuven, Belgium, 2007.
  - Carles Matabosch Geronès, Universitat de Girona, Spain, 2007.
  - Michela Farenzena, Università degli Studi di Verona, Italy, 2007.
- Edmond Boyer acted as a reviewer for the following PhD theses: Keith Forbes, University of Cape town.

## 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, n<sup>o</sup> 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, n<sup>o</sup> 2, 1998, p. 135-155.
- [3] R. HARTLEY, P. STURM. *Triangulation*, in "Computer Vision and Image Understanding", vol. 68, n<sup>o</sup> 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, n<sup>o</sup> 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, n<sup>o</sup> 4, August 1998, p. 525-532.
- [6] R. HORAUD, F. DORNAIKA. *Hand-Eye Calibration*, in "International Journal of Robotics Research", vol. 14, n<sup>o</sup> 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, n<sup>o</sup> 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, n<sup>o</sup> 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, n<sup>o</sup> 2-3, 2001, p. 101-126, <http://perception.inrialpes.fr/publication.php3?bibtex=PB01>.
- [14] E. PRADOS, F. CAMILLI, O. FAUGERAS. *A unifying and rigorous Shape From Shading method adapted to realistic data and applications*, in "Journal of Mathematical Imaging and Vision", 2006, <http://perception.inrialpes.fr/Publications/2006/PCF06>.

- [15] E. PRADOS, O. FAUGERAS. *A generic and provably convergent Shape-From-Shading Method for Orthographic and Pinhole Cameras*, in "International Journal of Computer Vision", vol. 65, n<sup>o</sup> 1/2, nov 2005, p. 97–125, <http://perception.inrialpes.fr/Publications/2005/PF05>.
- [16] E. PRADOS, O. FAUGERAS. *Shape from Shading: a well-posed problem ?*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, California", vol. II, IEEE, jun 2005, p. 870–877, <http://perception.inrialpes.fr/Publications/2005/PF05a>.
- [17] 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<sup>o</sup> 8, August 1997, p. 834-845.
- [18] 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<sup>o</sup> 1, January 1995, p. 34-46.
- [19] L. QUAN. *Conic Reconstruction and Correspondence from Two Views*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 18, n<sup>o</sup> 2, February 1996, p. 151-160.
- [20] L. QUAN. *Self-Calibration of An Affine Camera from Multiple Views*, in "International Journal of Computer Vision", vol. 19, n<sup>o</sup> 1, May 1996, p. 93-105.
- [21] A. RUF, R. HORAUD. *Visual Servoing of Robot Manipulators, Part I: Projective Kinematics*, in "International Journal of Robotics Research", vol. 18, n<sup>o</sup> 11, November 1999, p. 1101-1118, <http://hal.inria.fr/inria-00073002>.
- [22] C. SCHMID, R. MOHR, C. BAUCKHAGE. *Evaluation of Interest Point Detectors*, in "International Journal of Computer Vision", vol. 37, n<sup>o</sup> 2, 2000, p. 151-172, <http://perception.inrialpes.fr/publication.php3?bibtex=SMB00>.
- [23] 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<sup>o</sup> 5, May 1997, p. 530-534.
- [24] 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.
- [25] 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>.
- [26] 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", Jun 1997, p. 1100-1105.
- [27] 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.

- [28] B. TRIGGS. *Matching Constraints and the Joint Image*, in "IEEE Int. Conf. Computer Vision, Cambridge, MA", E. GRIMSON (editor), June 1995, p. 338-43.
- [29] B. TRIGGS. *Autocalibration and the Absolute Quadric*, in "IEEE Conf. Computer Vision & Pattern Recognition, Puerto Rico", 1997.

## Year Publications

### Doctoral dissertations and Habilitation theses

- [30] E. BOYER. *Modélisation de scènes dynamiques à partir de plusieurs caméras*, Ph. D. Thesis, Université Joseph Fourier, Janvier 2007, <http://perception.inrialpes.fr/Publications/2007/Boy07>.
- [31] D. KNOSSOW. *Paramétrage et Capture Multicaméras du Mouvement Humain*, PhD. manuscript, INPG, INRIA, 655 avenue de l'Europe, 38330 Montbonnot, April 2007, <http://perception.inrialpes.fr/Publications/2007/Kno07>.
- [32] C. MÉNIER. *Système de Vision Temps-Réel pour les Interactions*, Ph. D. Thesis, INPG, avril 2007, <http://perception.inrialpes.fr/Publications/2007/Men07>.

### Articles in refereed journals and book chapters

- [33] E. ARNAUD, E. MEMIN. *Partial Linear Gaussian Models for Tracking in Image Sequences Using Sequential Monte Carlo Methods*, in "International Journal of Computer Vision", vol. 74, n<sup>o</sup> 1, jan 2007, <http://perception.inrialpes.fr/Publications/2007/AM07>.
- [34] F. CUZZOLIN. *Two new Bayesian approximations of belief functions based on convex geometry*, in "IEEE Transactions on Systems, Man, and Cybernetics - Part B", 2007, <http://perception.inrialpes.fr/Publications/2007/Cuz07>.
- [35] D. KNOSSOW, R. RONFARD, R. P. HORAUD. *Human Motion Tracking with a Kinematic Parameterization of Extremal Contours*, in "International Journal of Computer Vision", To appear, 2007, <http://perception.inrialpes.fr/Publications/2007/KRH07>.
- [36] C. LENGLET, E. PRADOS, J.-P. PONS, R. DERICHE, O. FAUGERAS. *Brain connectivity mapping using riemannian geometry, control theory and PDES*, in "Submitted to SIAM Journal on Imaging Sciences (SIIMS)", 2007, <http://perception.inrialpes.fr/Publications/2007/LPPDF07>.
- [37] R. MOHR, M. DOUZE, P. STURM. *Géométrie projective, analyse numérique et vision par ordinateur*, in "Bulletin de l'Union des Professeurs de Spéciales – Mathématiques et Sciences Physiques", n<sup>o</sup> 219, jul 2007, p. 12–30, <http://perception.inrialpes.fr/Publications/2007/MDS07>.
- [38] T. PRIBANIC, P. STURM, M. CIFREK. *Calibration of 3D kinematic systems using orthogonality constraints*, in "Machine Vision and Applications", vol. 18, n<sup>o</sup> 6, nov 2007, p. 367–381, <http://perception.inrialpes.fr/Publications/2007/PSC07>.

### Publications in Conferences and Workshops

- [39] J. ALLARD, C. MÉNIER, B. RAFFIN, E. BOYER, F. FAURE. *Grimage: Markerless 3D Interactions*, in "Siggraph - Emerging Technologies", 2007, <http://perception.inrialpes.fr/Publications/2007/AMRBF07>.

- [40] M. COOKE, Y.-C. LU, Y. LU, R. P. HORAUD. *Active hearing, active speaking*, in "International Symposium on Auditory and Audiological Research (ISAAR 2007), Helsingor, Denmark", August 2007, <http://perception.inrialpes.fr/Publications/2007/CLLH07>.
- [41] F. CUZZOLIN, D. MATEUS, E. BOYER, R. P. HORAUD. *Robust Spectral 3D-bodypart Segmentation along Time*, in "Second Workshop on Human Motion, Understanding, Modeling, Capture and Animation, Rio de Janeiro, Brazil", Lecture Notes in Computer Science, vol. 4814, Springer, October 2007, p. 196–211, <http://perception.inrialpes.fr/Publications/2007/CMBH07a>.
- [42] P. GARGALLO, E. PRADOS, P. STURM. *Minimizing the Reprojection Error in Surface Reconstruction from Images*, in "Proceedings of the International Conference on Computer Vision, Rio de Janeiro, Brazil", IEEE Computer Society Press, 2007, <http://perception.inrialpes.fr/Publications/2007/GPS07>.
- [43] P. GARGALLO, P. STURM, S. PUJADES. *An Occupancy-Depth Generative Model of Multi-View Images*, in "Proceedings of the Asian Conference on Computer Vision, Tokyo, Japan", vol. 2, Springer, nov 2007, p. 373–383, <http://perception.inrialpes.fr/Publications/2007/GSP07>.
- [44] M. HANSARD, R. P. HORAUD. *A Model of Binocular Gaze Estimation*, in "Fourth Computational and Systems Neuroscience Meeting (COSYNE 2007), Salt Lake City, Utah, USA", February 2007, <http://perception.inrialpes.fr/Publications/2007/HH07>.
- [45] M. HANSARD, R. P. HORAUD. *Patterns of Binocular Disparity for a Fixating Observer*, in "Advances in Brain, Vision, and Artificial Intelligence, Second International Symposium, BVAI'07, Naples, Italy", Lecture Notes in Computer Science, Springer, 10-12 October 2007, p. 308–317, <http://perception.inrialpes.fr/Publications/2007/HH07a>.
- [46] D. KNOSSOW, J. VAN DE WEIJER, R. P. HORAUD, R. RONFARD. *Articulated-Body Tracking Through Anisotropic Edge Detection*, in "Dynamical Vision", Lecture Notes in Computer Science, vol. LNCS 4358, Springer, 2007, p. 86-99, <http://perception.inrialpes.fr/Publications/2007/KVHR07>.
- [47] W. LEE, W. WONTACK, E. BOYER. *Identifying Foreground from Multiple Images*, in "In Proceedings of the Eighth Asian Conference on Computer Vision", December 2007, <http://perception.inrialpes.fr/Publications/2007/LWB07>.
- [48] D. MATEUS, F. CUZZOLIN, R. P. HORAUD, E. BOYER. *Articulated Shape Matching by Robust Alignment of Embedded Representations*, in "IEEE Workshop on 3D Representation for Recognition (3DRR 2007), Rio de Janeiro, Brazil", IEEE Computer Society Press, October 2007, <http://perception.inrialpes.fr/Publications/2007/MCHB07a>.
- [49] D. MATEUS, F. CUZZOLIN, R. P. HORAUD, E. BOYER. *Articulated Shape Matching Using Locally Linear Embedding and Orthogonal Alignment*, in "IEEE Workshop on Non-rigid Registration and Tracking through Learning - NRTL 2007, Rio de Janeiro, Brazil", IEEE Computer Society Press, October 2007, <http://perception.inrialpes.fr/Publications/2007/MCHB07>.
- [50] D. MATEUS, R. P. HORAUD. *Spectral Methods for 3-D Motion Segmentation of Sparse Scene-Flow*, in "IEEE Workshop on Motion and Video Computing, WMVC'07", IEEE Computer Society Press, IEEE, February 2007, <http://perception.inrialpes.fr/Publications/2007/MH07>.

- [51] J. MORAT, F. DEVERNAY, S. CORNOU, J. IBANEZ GUZMAN. *Evaluation Method for Automotive Stereo-Vision Systems*, in "Proceedings of IEEE Intelligent Vehicles Symposium", jun 2007, <http://perception.inrialpes.fr/Publications/2007/MDCI07>.
- [52] J. MORAT, F. DEVERNAY, S. CORNOU. *Mesure de Trajectoire par Stéréo-Vision pour des Applications de Suivi de Véhicules à Basse Vitesse*, in "Actes des Journées ORASIS, Obernay", jun 2007, <http://perception.inrialpes.fr/Publications/2007/MDC07a>.
- [53] J. MORAT, F. DEVERNAY, S. CORNOU. *Tracking with Stereo Vision System for Low Speed Following Applications*, in "Proceedings of IEEE Intelligent Vehicles Symposium", Jun 2007, <http://perception.inrialpes.fr/Publications/2007/MDC07>.
- [54] L. SOARES, C. MÉNIER, B. RAFFIN, J.-L. ROCH. *Parallel Adaptive Octree Carving for Real-time 3D Modeling*, in "IEEE Virtual Reality", march 2007, <http://perception.inrialpes.fr/Publications/2007/SMRR07>.
- [55] P. STURM, P. GARGALLO. *Conic fitting using the geometric distance*, in "Proceedings of the Asian Conference on Computer Vision, Tokyo, Japan", vol. 2, Springer, 2007, p. 784–795, <http://perception.inrialpes.fr/Publications/2007/SG07>.
- [56] J.-P. TARDIF, P. STURM, S. ROY. *Plane-based self-calibration of radial distortion*, in "Proceedings of the International Conference on Computer Vision, Rio de Janeiro, Brazil", IEEE Computer Society Press, oct 2007, <http://perception.inrialpes.fr/Publications/2007/TSR07>.
- [57] D. WEINLAND, E. BOYER, R. RONFARD. *Action Recognition from Arbitrary Views using 3D Exemplars*, in "Proceedings of the International Conference on Computer Vision, Rio de Janeiro, Brazil", IEEE Computer Society Press, 2007, p. 1–7, <http://perception.inrialpes.fr/Publications/2007/WBR07>.
- [58] K.-J. YOON, A. DELAUNOY, P. GARGALLO, P. STURM. *Toward Global and Model based Multiview Stereo Methods for Shape and Reflectance Estimation*, in "Proceedings of the First International Workshop on Photometric Analysis For Computer Vision (in conjunction with ICCV'07)", 2007, <http://perception.inrialpes.fr/Publications/2007/YDGS07>.
- [59] K.-J. YOON, I. S. KWEON. *Stereo Matching with the Distinctive Similarity Measure*, in "Proceedings of the International Conference on Computer Vision, Rio de Janeiro, Brazil", IEEE Computer Society Press, 2007, <http://perception.inrialpes.fr/Publications/2007/YK07>.
- [60] A. ZAHARESCU, E. BOYER, R. P. HORAUD. *TransforMesh: a topology-adaptive mesh-based approach to surface evolution*, in "In Proceedings of the Eighth Asian Conference on Computer Vision, Tokyo, Japan", LNCS 4844, vol. II, Springer, November 2007, p. 166-175, <http://perception.inrialpes.fr/Publications/2007/ZBH07>.

### Internal Reports

- [61] N. WOTAWA, C. LENGLET, E. PRADOS, R. DERICHE, O. FAUGERAS. *Anatomical connections in the human visual cortex: validation and new insights using a DTI Geodesic Connectivity Mapping method*, Technical report, n<sup>o</sup> RR-6176, INRIA Rhône-alpes, 2007, <http://hal.inria.fr/inria-00144016/>.

- [62] K.-J. YOON, E. PRADOS, P. STURM, A. DELAUNOY, P. GARGALLO. *Shape and Reflectance Recovery using Multiple Images with Known Illumination Conditions*, Technical report, n<sup>o</sup> RR-6309, INRIA Rhône-alpes, September 2007, <http://hal.inria.fr/inria-00175274>.