INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team perception

Rhône-Alpes

THEME COG

Table of contents

1. Team

Head of the team

Radu Horaud [Research Director (DR), HdR]

Administrative assistant

Anne Pasteur [Secretary (SAR) Inria]

Personnel Inria

Frédéric Devernay [Research Associate (CR), team member until August 2006]
Emmanuel Prados [Research Associate (CR)]
Rémi Ronfard [Research Associate (CR), team member until August 2006]
Peter Sturm [Research Director (DR), HdR]

Research scientists (external)

Elise Arnaud [Université Joseph Fourier Grenoble, since September 2006]
Edmond Boyer [Université Joseph Fourier Grenoble]

Ph.D students

Ouideh Bentrah [in collaboration with IGN]
Thomas Bonfort [INRIA grant]
Jean-Sébastien Franco [MESR grant,[team member until March 2006]
Pau Gargallo [MESR grant]
Frédéric Huguet [MESR grant]
Aude Jacquot [CIFRE funding with Thales]
David Knossow [INRIA grant]
Diana Mateus [Marie-Curie grant]
Clément Ménier [ENS PhD student]
Julien Morat [CIFRE funding with Renault]
Ramya Narasimha [INRIA grant]
Srikumar Ramalingam [Marie-Curie grant, co-supervision agreement with U California]
Daniel Weinland [Marie-Curie Grant]
Andrei Zaharescu [Marie-Curie grant]

Post-doctoral fellows

Fabio Cuzzolin
Miles Hansard
Kuk-Jin Yoon

Technical staff

Hervé Mathieu [Research Engineer (IR)]
Florian Geffray
Bertrand Holveck
Marc Lapierre [Team member until March 2006]
Loïc Lefort
Philippe Martin
Serji Pujades [Team member until August 2006]
Elise Taillant

Visitors

Jean-Philippe Tardif [PhD student at Université de Montréal, Canada]
Kiyoung Kim [PhD student at Gwangju Institute of Science and Technology, South Korea]
Nitin Jindal [M.S. Trainee]

2. Overall Objectives

2.1. Overall Objectives

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

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

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

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

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

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

3. Scientific Foundations

3.1. The geometry of multiple images

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

3.2. Multiple-camera acquisition of visual data

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

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

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

3.3. Shape Acquisition

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

3.4. Motion Analysis

The perception of motion is one of the major goals in computer vision with a wide range of promising applications. A prerequisite for motion analysis is motion modelling. Motion models span from rigid motion to complex articulated and/or deformable motion. Deformable objects form an interesting case because the models are closely related to the underlying physical phenomena. In the recent past, robust methods were developed for analysing rigid motion. This can be done either in image space or in 3D space. Image-space analysis is appealing but it requires sophisticated non-linear minimization methods and a probabilistic framework. An intrinsic difficulty with methods based on 2D data is the ambiguity of associating a multiple degree of freedom 3D model with image contours, texture and optical flow. Methods using 3D data are more relevant with respect to our recent research investigations. 3D data are produced using stereo or a multiple-camera setup. These data are matched against an articulated object model (based on cylindrical parts, implicit surfaces, conical parts, and so forth). The matching is carried out iteratively using various methods, such as ICP (iterative close point) or EM (expectation/maximization).

Challenging problems are the detection of motion and motion tracking. When a vision systems observes complex articulated motion, such as the motion of the hands, it is crucial to be able to detect motion cues and to interpret them in terms of moving parts, independently of a prior model. Another difficult problem is to track articulated motion over time and to estimated the motions associated with each individual degree of freedom.

4. Application Domains

4.1. 3D modeling and rendering

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

4.2. Mixed and Augmented Reality

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

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

4.3. Human Motion Capture and Analysis

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

4.4. Multi-media and interactive applications

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

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

4.5. Car driving technologies

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

4.6. Defense technologies

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

5. Software

5.1. Platforms

5.1.1. The Grimage platform

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

5.1.2. The MV platform

The MV-platform is a software platform developed in conjunction with Grimage. It allows real-time image acquisition, 3D modeling and interaction using multiple cameras. This platform has been developed by the PERCEPTION team in collaboration with the MOAIS team through different National and European projects.

5.2. Camera calibration

5.2.1. MVCamera: Multi-camera calibration

We released a second version of our multi-camera calibration software (MVCAMERA) in October 2006. The new software includes an enhanced graphical user interface (GUI) to allow even non-expert users to perform and validate a calibration. The new version automates the Initialization of the camera positions and orientations in a global reference coordinate system. The software has been extensively tested and validated internally in the GRIMAGE room, with camera configurations of up to 40 cameras, and delivered to some of our industry partners.

5.2.2. *Mono and stereo calibration*

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

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

The software package is downloadable at http://perception.inrialpes.fr/Soft/calibration/index.html.

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

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

- A *frontend* is directly connected to the camera driver and is in charge of doing the image acquisition. Blinky makes the images available either to the local host by using shared memory or to other hosts by networking, allowing images to be captured transparently accross a local network. Actually, the frontend can be seen as an image server, which can also be used to change the camera parameters (shutter speed, aperture, zoom...). The frontend is also able to record the raw video stream to a file, which can later be used as if it were a live camera.
- A *backend* is the user application that captures the images and processes them. Each camera is designated by a device name which is valid accross the network, and acquiring an image is as simple as reading a file. Multiple backends can connect to a single frontend, allowing for many applications to run at the same time using the same cameras. A single backend can also connect to multiple cameras, and a single function call is necessary to acquire a set of synchronized images from multiple cameras spread over the network.

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

- *blinkyf1394*, a frontend for IIDC-compliant IEEE1394 cameras on Linux.
- *blinkyflnrd*, a frontend for CameraLink cameras, using the Arvoo Leonardo frame grabber on Linux.
- *blinkyfdummy*, a dummy frontend serving predictable images.
- *blinkybenchmark*, a backend to test the acquisition rate (FPS) of a camera.
- *blinkysaveimages*, a backend that images aquired for one or several cameras to any standard image format.
- *blinkysdl*, a backend that visualizes the video stream using the SDL graphical library.

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

5.4. MVVideo: Multiple-camera video player and recorder

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

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

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

5.5. Point tracking in video sequences

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

5.6. Real-time shape acquisition and visualization

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

5.7. Stereo-based 3D reconstruction

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

5.8. Visual-Hull library (EPVH)

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

5.9. MVACTOR - Human-body modelling for motion capture and tracking

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

That same software is also used for performing markerless motion capture using such body models. MVCAMERA was developped by INRIA as part of the SEMOCAP project.

5.10. Dynamic texture reconstruction and visualization

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

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

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

6. New Results

6.1. Self-calibration of general radial distortions

In line with our recent works on a generic framework for camera calibration, see e.g. [25], we address here a special case of particular practical interest. We propose self-calibration approaches for a generic radial distortion model that handles classical radial distortion cases but also fisheye cameras and most common catadioptric cameras (systems composed of cameras and mirrors). The model only imposes that the 3D-to-2D mapping be radially symmetric about the optical axis and the image center; besides this, the mapping can be arbitrary, which comes down to having an arbitrary, possibly non-parametric, radial distortion function. In [71], [66], [32], we propose two different self-calibration approaches for this camera model. The approaches are elegant and produce excellent results.

6.2. Object Tracking

Within a contract with Thalèes Optronique we have extended our previous approach [76] for object tracking in videos. In [31], [56], we propose an approach that allows to switch between different geometric models for the tracked object, in order to adapt the model to the current situation. For example, in our target application of tracking in aerial image sequences, objects appear in widely different scales throughout a sequence. Often, tracking starts when the object only covers a diameter of a dozen pixels, in which case a planar geometric model is usually sufficient for tracking. Later in the sequence, when the aircraft carrying the camera approaches, the object's 3D structure may become discernible, and the planar model should be abandoned in the favor of a 3D model. Our approach automatically switches between different such geometric models. To be precise, this switch is done in a "soft" manner: tracking is done within the particle filtering framework and each particle is allowed to change its geometric model.

We also combine these geometric models with image appearance, encoded by greylevel histograms. This combined approach gave the most promising results. Even for difficult real sequences, good results were obtained, although in these cases, the approach remains sensitive to the choice of tuning parameters.

6.3. Tracking articulated motion

One of the major research efforts has been to address the problem of articulated motion tracking.

6.3.1. Human-body tracking using an implicit surface, 3D points, and surface normals.

We addressed the problem of human motion tracking by registering a surface to a set of 3-D data. We developed an *implicit and articulated surface model* that combines a volumetric representation with a kinematic description. The 3-D data consist in both 3-D points and normals. We developed a distance function between such a 3-D datum and an ellipsoid and we showed how it can be used to fit an implicit surface to a set of 3-D observations. The tracking per se is carried out by a model-to-data fitting technique. The latter is cast into a maximum likelihood estimator implemented as an expectation-maximization algorithm. The method has been successfully applied to the problem of tracking human motions with six cameras, see figure 1.

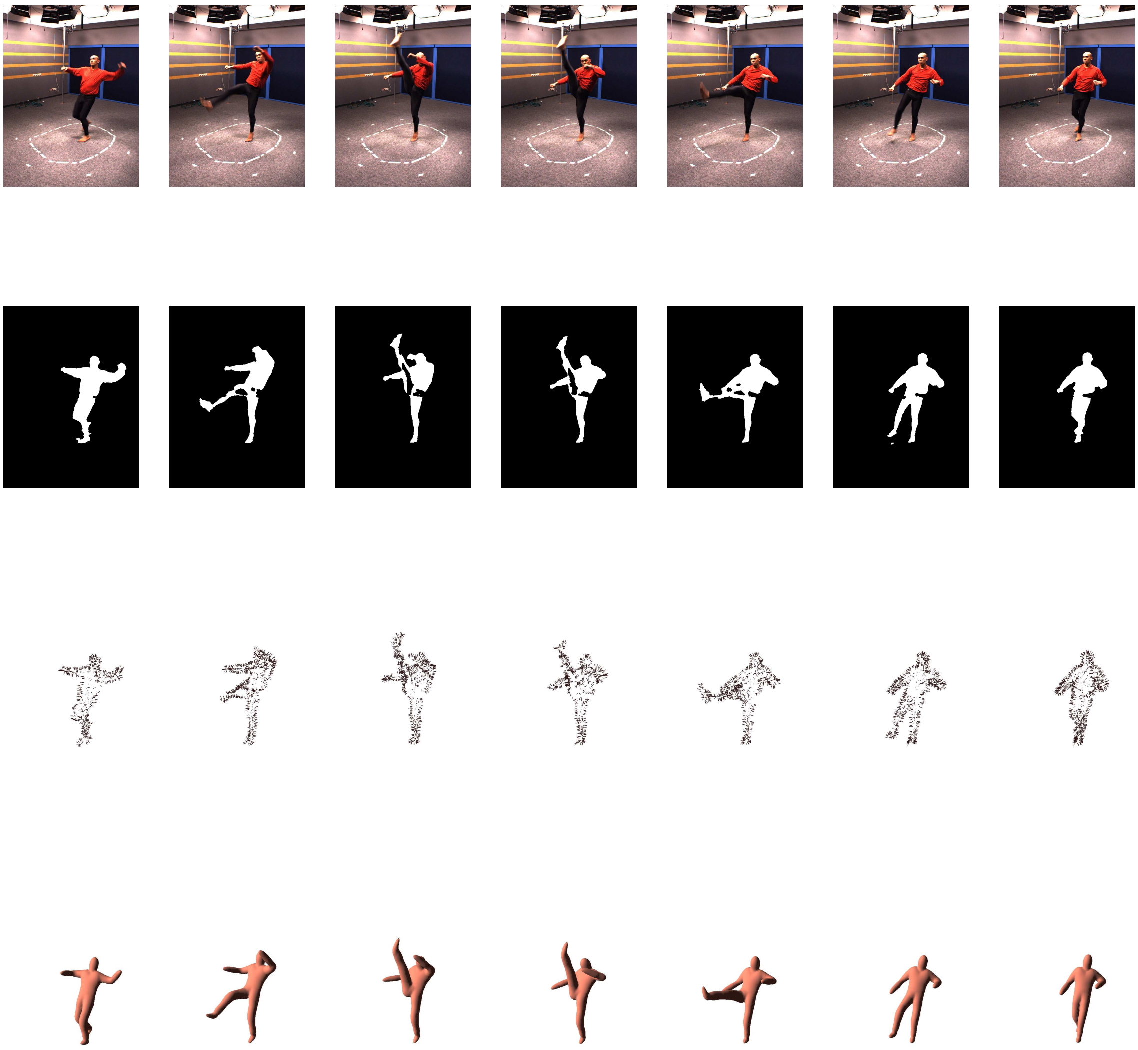

Figure 1. Human-motion tracking using 3-D observations (points and normals) and an articulated and implicit surface. The first row shows one image sequence (out of 6 sequences) gathered at 28 frames per second. The second row shows the 2-D silhouettes and the third row shows the 3-D data extracted from these silhouettes. The last row shows the results of fitting the model to the data.

6.3.2. Human-body tracking using the kinematics of extremal contours.

This paper addresses the problem of human motion tracking from image sequences. 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. The tracker has been successfully applied to several complex motions gathered at 30 frames/second, see figure 2.

6.3.3. Hand tracking using 3-D point set registration

We revisited the point set registration problem in order to extend it from rigid objects to articulated objects. We showed that the problem can be solved by combinging a probabilistic framework (supervised clustering) with a closed-form solution for the articulated pose problem. We proposed to replace one-to-one point assignments by many-to-one. A data point can be assigned either to a model point or to an outlier category. Several data points can assigned to the same model point. We devised an expectation-maximization algorithm that alternates between updating the assignments and updating the kinematic parameters. The method has been successfully applied to the problem of hand tracking.

6.4. Action recognition

PERCEPTION started research activities on Action recognition last year. We have been investigating the use of volumetric reconstructions from multiple cameras for the purpose of recognizing primitive actions such as crossing arms, waiving, standing up, sitting down, etc. To that end, we derived motion descriptors based on the cylindric Fourier decomposition of a new representation called the Motion History Volume. In [73], we applied those descriptors to automatically segment and classify actions from multiple videos into consistent groups.

6.5. Multiple camera reconstruction and motion segmentation

6.5.1. Robust perspective factorization.

One of our major research topics is the problem of recovering structure and motion from a large number of intrinsically calibrated perspective cameras. We describe a method that combines (1) weak-perspective reconstruction in the presence of noisy and missing data and (2) an algorithm that updates weak-perspective reconstruction to perspective reconstruction by incrementally estimating the projective depths. The method also solves for the reversal ambiguity associated with affine factorization techniques. The method has been successfully applied to the problem of calibrating the external parameters (position and orientation) of several multiple-camera setups. Results obtained with synthetic and experimental data compare favourably with results obtained with non-linear minimization such as bundle adjustment.

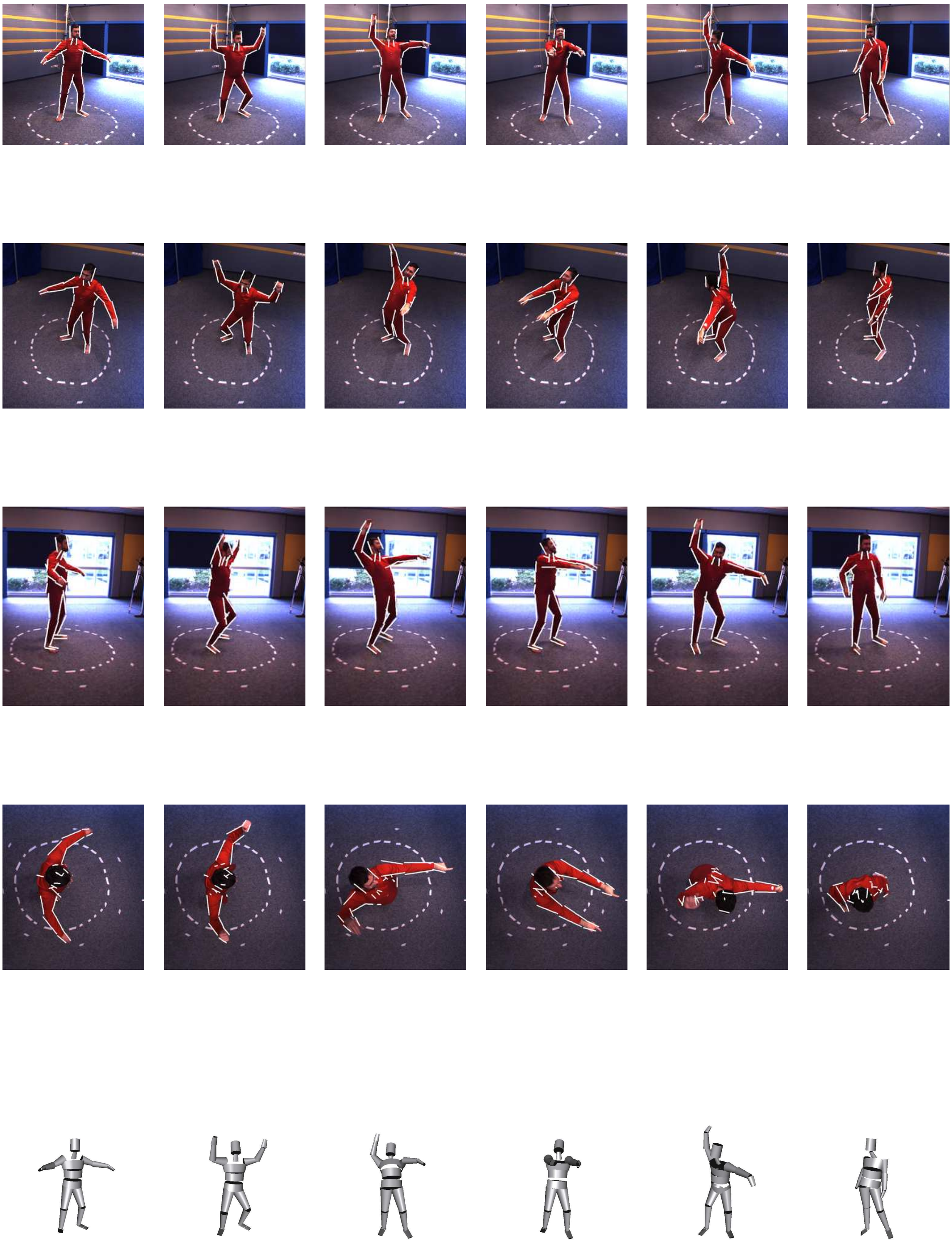

Figure 2. A human character is observed with four cameras (from top to bottom) and his motions are tracked over time (from left to right). The tracker is based on matching image edges with edges predicted from an articulated model. The predicted edges are shown overlapped onto the input images. The last row shows the pose (position and orientation) of the articulated human-body model.

6.5.2. 3-D motion segmentation.

We investigated the problem of motion segmentation within the context of spectral clustering. The input data consisted in sparse scene flow. The latter consists in a set of trajectories associated with the motion of 3-D points. We addressed the problem of segmenting these trajectories based on the rigidity constraints. Since it appeared difficult to represent these trajectories in some parameter space, we decided to use spectral methods. We proposed a similarity measure between two trajectories and we showed how it can be plugged into a spectral method. In practice we implemented and tested two algorithms: the first one seeks degrees of correlation within the data, the second one uses the Multiway Normalized Cut method.

6.6. Computational modelling of binocular human vision

This year we started to investigate computational stereopsis from the point of view of biological plausibility. So far we concentrated onto two topics: the control of eye movements for achievieng binocular gaze and the implementation of intensity-based stereo matching.

6.6.1. Binocular gaze and epipolar geometry.

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.

6.6.2. Stereo correspondence 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. Statitistical 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 3), 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 would like 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.

We would also like to relate the problem of dense stereo matching with areas in the visual cortex which are believed to play a role in depth perception. A large number of investigators identified single neurons as well as populations of neurons whose responses are simultaneously associated with left-eye and right-eye receptive fields. However, how the brain achieves stereo correspondence (which is a prerequisite for depth perception) is not clear and remains to be investigated. Moreover, the link between correspondence and binocular vergence control must be investigated as well.

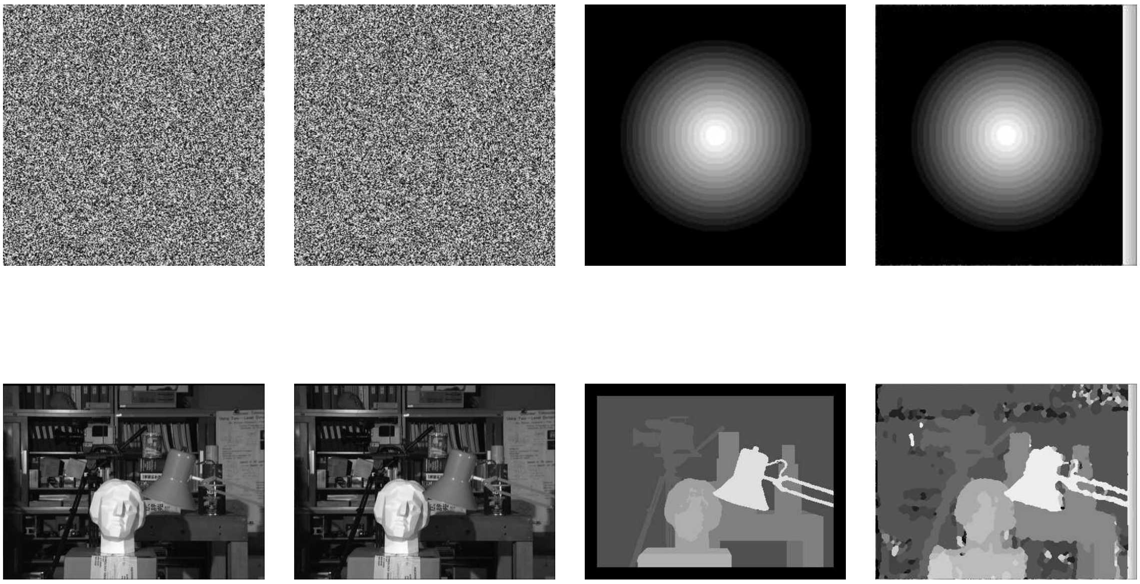

Figure 3. Recently we started to address the problem of dense stereo matching using a mean-field approximation (MFA) method. The dense disparity (or depth) field is the hidden variable in a Markov random field (MRF) model. This work aims at establishing a link between computational stereo and the biological plausibility of MRF/MFA methods. From left to right: two stereo pairs, the ground-truth, and the depth map found by our method. The second example is from the wellknown Tsukuba database. This work is a collaboration between the PERCEPTION and the MISTIS teams and is carried out with the European project POP.

6.7. Other Results

6.7.1. Self-Calibration and Euclidean Reconstruction using Images of Circles

It has been well-known that images of circles, even in the absence of knowledge about their relative position or their diameteres, provide information on the camera's intrinsic parameters [77]. In the last few years a fair number of algorithms that exploit this have been proposed, for camera self-calibration and 3D modelling, especially in the case of so-called "turntable" image sequences. Nonetheless, no general and complete solution has so far been developed: in the basic case of disposing of images of two circles, several special cases exist and no previous approach has given the correct solution in all cases. In [54], we propose such a complete general approach. It is based on a unified problem formulation and gives the minimum number of possible solutions for all special cases; further, algorithms for using images of any number of circles are proposed.

7. Contracts and Grants with Industry

7.1. Thalès-Optronics SA

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

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

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

7.2. Renault SA

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

In June 2004 we started a 3 year collaboration with the French car manufacturer Renault SA (Direction de la Recherche). Within this collaboration Renault co-funds a PhD thesis with ANRT. The topic of the collaboration and of the thesis is the detection and classification of obstacles which are ahead of a vehicle. We currently develop a prototype system based on *stereoscopic vision* with the following functionalities: low speed following, pre-crash, and pedestrian detection. In particular we study the robustness of the image processing algorithms with respect to camera/stereo calibration problems (the system should be able to self-detect such problems).

8. Other Grants and Activities

8.1. National initiatives

8.1.1. 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 will implement 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 team members who have been or will be involved in this project are Peter Sturm, Srikumar Ramalingam, and a post-doc to be hired in 2007.

8.1.2. ANR project STANDS-MSD (young researchers program)

STANDS-MDS (Spatio-Temporal Analysis of Deformable Structures in Meteosat Second Generation Images) is a 3-year project that will start on December 1, 2006 . This project aims at developing some computer vision methods devoted to the MSG (Meteosat Second Generation) meteorological images. More precisely, the project will focus on the detection and the tracking of two climatic events of interest : the sea breeze phenomenon and the convective cells.

The consortium is composed of two INRIA teams specialized in computer vision (The VISTA and the PERCEPTION teams), the climatology team COSTEL (Climate and Land Cover with Remote Sensing) situated in Rennes, as well as other laboratories (LASMEA, GREYC, LMD).

The team member involved in this project is Elise Arnaud. She will work on tracking of fluid deformable structures using stochastic filtering methods.

8.1.3. RNTL-OCETRE

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

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

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

8.1.4. RIAM–SEMOCAP

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

8.1.5. ACI–GEOLSTEREO

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

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

8.1.6. ARC–Georep

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

8.2. Projects funded by the European Commission

8.2.1. FP6-IST STREP project Holonics

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

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

8.2.2. FP6/Marie-Curie EST Visitor

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

8.2.3. FP6/Marie-Curie RTN VISIONTRAIN

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

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

8.2.4. 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.

8.3. Associate Team

Our associate team with Brown University (Providence, Rhode Island, USA) on the theme of 3D cinematography was startec in 2005 and renewed in 2006.

Two students from INRIA visited Brown University in 2006. Three Brown students visited INRIA for internships in the summer. A one-day scientific seminar was held at INRIA and attended by all PERCEPTION team members as well as Gabriel Taubin and his three student from Brown in July 2006. Rémi Ronfard spent 4 weeks in Brown University in November 2006 with the summer interns, to finalize their projects and send them to international conferences (CVPR and ICCV).

Gabriel Taubin and Remi Ronfard co-organized and co-chaired an international workshop on 3D cinematography in New York City on June 22, 2006. This workshop)was associated with CVPR, and included keynote speeches by Takeo Kanade and Iannis Aloimonos, among others. The workshop was well attended by an international audience, including many associate team members. A second edition is already scheduled to take place in Minneapolis in June 2007. The IEEE journal Computer Graphics and Applications will include a report on the workshop, due in April 2007.

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*.
- Peter Sturm is a member of the editorial board of the *Image and Vision Computing Journal*.
- Peter Sturm was Area Chair of ECCV (European Conference on Computer Vision, Graz, Austria) and member of the Program Committees of:
    - RFIA, Congrès de Reconnaissance des Formes et Intelligence Artificielle, Tours, France.
    - CVPR, IEEE Conf. on Computer Vision and Pattern Recognition, New York, USA.

  - BMVC, British Machine Vision Conference, Edinburgh, UK.
  - ICIP, IEEE International Conference on Image Processing, Atlanta, USA.
  - 3DPVT, Int. Symp. 3D Data Processing, Visualization, ..., Chapel Hill, USA.
  - PCV, Symposium “Photogrammetric Computer Vision” of ISPRS Commission III, Bonn, Germany.
  - ACCV, Asian Conference on Computer Vision, Hyderabad, India.
  - ICPR, 18th International Conference on Pattern Recognition, Hongkong.
  - VISAPP, Int. Conf. Computer Vision Theory and Applications, Setubal, Portugal.
  - ICVGIP, Indian Conf. Computer Vision, Graphics and Image Proc., Madurai, India.
  - ISVC, International Symposium on Visual Computing, Lake Tahoe, USA.
  - DEFORM, Workshop on Image Registration in Deformable Environments, Edinburgh, UK, held in conjunction with BMVC.
  - DV, 2nd Workshop on Dynamical Models for Computer Vision, held in conjunction with ECCV).

- Emmanuel Prados was member of the Program Committees of:
  - SSVM, Scale Space and Variational Methods Conference, Ischia, Italy.
  - The thematic school “Optimization Methods in Computer Vision”, Les Houches-Servoz, France.
- Rémi Ronfard was a member of the Program Committees of:
  - CVMP, Second European Conference on Visual Media Production, London, UK.

  and a co-organizer of:
  - 3DCINE, First Workshop on Threez-Dimensional Cinematography, held in conjunction with CVPR, New York, USA.
- Rémi Ronfard was a reviewer for the SIGGRAPH and ICIP conferences
- Rémi Ronfard was a co-editor of a special issue of Computer Vision and Image Understanding published in November-December 2006.
- Edmond Boyer was Area Chair of the British Machine Vision Conference (BMVC) and member of the program committees of: CVPR, ECCV, 3DPVT, ICPR, BMVC, WRUPKV (ECCV Workshop on Prior Knowledge in Vision).

9.2. Services to the Scientific Community

- Radu Horaud is in charge of European coordination at INRIA Rhône-Alpes.
- Peter Sturm co-organized the BIRS Invitational Workshop on Mathematical Methods in Computer Vision, Banff, Canada.
- Peter Sturm organized the “Journée Thématique Modélisation 3D à partir d’images” of the GdR ISIS, Paris, France.
- Peter Sturm co-organized the “Journées Nationales des ARC (Actions de Recherche Coopérative)” at INRIA Rhône-Alpes.
- 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 a member of the INRIA Committee on “Actions Incitatives” (part of COST – Conseil d’Orientation Scientifique et Technologique).
- Remi Ronfard is member of the "Commission de Specialistes" for recruitments at the University Joseph Fourier of Grenoble.
- Edmond Boyer is member of the IMAG (Intitut d’Informatique et de Mathématiques Appliquées de Grenoble) Scientific Committe.
- Edmond Boyer is member of the "Commission de specialistes" for recruitments at the University Joseph Fourier of Grenoble and at the Institut National Polytechnique de Grenoble.
- Edmond Boyer is coordinator of the Marie-Curie Visitor Project and member of the Visitor Scientific Committe.
- Radu Horaud is the coordinator of the Visiontrain Marie Curie Research Training Network.

9.3. Teaching

- 3D Computer Vision, postgraduate course, University of Zaragoza, Spain, 20h, P. Sturm.
- Analyse d’images, DESS INFORMATIQUE, UNIV. JOSEPH FOURIER, 30H, R. Ronfard.
- Optimisation, M2R IVR, INPG, 6h, P. Sturm.
- Vision 3D, M2R IVR, INPG, 12h, P. Sturm.
- Géométrie projective, M2R IVR, INPG, 6h, E. Boyer.
- Stereoscopic Perception, Mastère Photogrammétrie Numérique, ENSG, Marne-la-Vallée, 17h, F. Devernay.
- Computer Vision, Mastère 2 Pro IICAO, Université Joseph Fourier, Grenoble, 24h, F. Devernay.
- Image retrieval, Mastère 2 Pro, Universite Joseph Fourier, Grenoble, 15h, E. Arnaud.

9.4. Tutorials and invited talks

- Peter Sturm, Srikumar Ramalingam, and Rahul Swaminathan (Telekom Labs Berlin) gave a tutorial on *General Imaging – Design, Calibration and Applications* at ECCV (European Conference on Computer Vision, Graz, Austria).
- Peter Sturm gave a tutorial on *Modeling and Analysing Images of Generic Cameras* at the ISPRS Symposium “Photogrammetric Computer Vision”, Bonn, Germany.
- Peter Sturm gave an invited talk at the Sanken International Symposium 2006 (Osaka, Japan) and at the Computer Vision Colloquium (associated with ECCV’06 Area Chair meeting, Graz, Austria).

9.5. Thesis

- Thomas Bonfort defended his PhD thesis in February 2006, and Srikumar Ramalingam and Aude Jacquot in November 2006.
- Peter Sturm defended his *Habilitation à diriger des Recherches* in May 2006.
- Peter Sturm acted as reviewer for the following PhD theses:
  - Bertrand Vandeportaele, Institut National Polytechnique de Toulouse, France, 2006.
  - Benjamin Albouy, Université d’Orl’eans, France, 2006.
  - Eric Royer, Universit’e Blaise Pascal, Clermont-Ferrand, France, 2006.
- Peter Sturm acted as reviewer for the following DRT thesis (Diplôme de Recherche Technologique):
  - Rada Stegerean, Institut National Polytechnique de Grenoble, France, 2006.

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^o 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^o 2, 1998, p. 135-155.

[3] R. HARTLEY, P. STURM. *Triangulation*, in "Computer Vision and Image Understanding", vol. 68, n^o 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^o 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^o 4, August 1998, p. 525-532.

[6] R. HORAUD, F. DORNAIKA. *Hand-Eye Calibration*, in "International Journal of Robotics Research", vol. 14, n^o 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^o 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^o 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^o 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^o 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^o 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^o 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^o 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^o 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^o 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^o 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^o 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", Juin 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] T. BONFORT. *Reconstruction de surfaces réfléchissantes à partir d'images*, Ph. D. Thesis, Institut National Polytechnique de Grenoble, feb 2006, http://perception.inrialpes.fr/Publications/2006/Bon06.

[31] A. JACQUOT. *Suivi d'objets en imagerie aérienne*, Ph. D. Thesis, Institut National Polytechnique de Grenoble, nov 2006, http://perception.inrialpes.fr/Publications/2006/Jac06.

[32] S. RAMALINGAM. *Generic Imaging Models: Calibration and 3D Reconstruction Algorithms*, Ph. D. Thesis, Institut National Polytechnique de Grenoble, nov 2006, http://perception.inrialpes.fr/Publications/2006/Ram06.

[33] P. STURM. *Recherches en vision par ordinateur*, Habilitation à diriger des Recherches, Institut National Polytechnique de Grenoble, may 2006, http://perception.inrialpes.fr/Publications/2006/Stu06.

Articles in refereed journals and book chapters

[34] E. ARNAUD, E. MEMIN. *Partial Linear Gaussian Models for Tracking in Image Sequences Using Sequential Monte Carlo Methods*, in "International Journal of Computer Vision", to appear 2006, http://perception.inrialpes.fr/Publications/2006/AM06.

[35] F. CAMILLI, E. PRADOS. *Shape-from-Shading with discontinuous image brightness*, in "Applied Numerical Mathematics", vol. 56, n^{o} 9, sept 2006, p. 1225–1237, http://perception.inrialpes.fr/Publications/2006/CP06.

[36] R. P. HORAUD, D. KNOSSOW, M. MICHAELIS. *Camera Cooperation for Achieving Visual Attention*, in "Machine Vision and Applications", vol. 16, n^{o} 6, February 2006, p. 331-342, http://perception.inrialpes.fr/Publications/2006/HKM06.

[37] H. MATHIEU, P. FUCHS. *Les Capteurs de Localisation*, in "Le Traité de la Réalité Virtuelle : Interfaçage, immersion et interaction en environnement virtuel", P. FUCHS (editor). , vol. 2, chap. LES CAPTEURS DE LOCALISATION, Les Presses de l'École des Mines de Paris, mar 2006, http://perception.inrialpes.fr/Publications/2006/MF06.

[38] 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.

[39] E. PRADOS, F. CAMILLI, O. FAUGERAS. *A viscosity solution method for Shape-From-Shading without image boundary data*, in "Mathematical Modelling and Numerical Analysis (M2AN)", vol. 40, n^{o} 2, 2006, p. 393–412, http://perception.inrialpes.fr/Publications/2006/PCF06a.

[40] E. PRADOS, O. FAUGERAS. *Shape From Shading*, in "Handbook of Mathematical Models in Computer Vision", N. PARAGIOS, Y. CHEN, O. FAUGERAS (editors). , chap. 23, Springer, 2006, p. 375–388, http://perception.inrialpes.fr/Publications/2006/PF06a.

[41] S. RAMALINGAM, S. LODHA, P. STURM. *A Generic Structure-from-Motion Framework*, in "Computer Vision and Image Understanding", vol. 103, n^o 3, sep 2006, p. 218–228, http://perception.inrialpes.fr/Publications/2006/RLS06.

[42] P. STURM, S. RAMALINGAM, S. LODHA. *On Calibration, Structure from Motion and Multi-View Geometry for Generic Camera Models*, in "Imaging Beyond the Pinhole Camera", K. DANIILIDIS, R. KLETTE (editors). , Computational Imaging and Vision, vol. 33, Springer, aug 2006, http://perception.inrialpes.fr/Publications/2006/SRL06.

[43] D. WEINLAND, R. RONFARD, E. BOYER. *Free Viewpoint Action Recognition using Motion History Volumes*, in "Computer Vision and Image Understanding", vol. 104, n^o 2-3, November/December 2006, http://perception.inrialpes.fr/Publications/2006/WRB06a.

Publications in Conferences and Workshops

[44] J. ALLARD, J.-S. FRANCO, C. MÉNIER, E. BOYER, B. RAFFIN. *The GrImage Platform: A Mixed Reality Environment for Interactions*, in "Proceedings of the 4th International Conference on Computer Vision Systems", jan 2006, http://perception.inrialpes.fr/Publications/2006/AFMBR06.

[45] E. ARNAUD, E. DELPONTE, F. ODONE, A. VERRI. *Trains of keypoints for 3D object recognition*, in "International Conference on Pattern Recognition, ICPR'06", aug 2006.

[46] E. ARNAUD, E. MEMIN, G. A. SOSA. *A fluid motion estimator for Schlieren image Velocimetry*, in "European Conference on Image Processing", may 2006.

[47] N. BIRKBECK, D. COBZAS, P. STURM, M. JÄGERSAND. *Variational Shape and Reflectance Estimation under Changing Light and Viewpoints*, in "Proceedings of the 9th European Conference on Computer Vision, Graz, Austria", vol. 1, Springer, may 2006, p. 536-549, http://perception.inrialpes.fr/Publications/2006/BCSJ06.

[48] T. BONFORT, P. STURM, P. GARGALLO. *General Specular Surface Triangulation*, in "Proceedings of the Asian Conference on Computer Vision, Hyderabad, India", vol. II, jan 2006, p. 872–881, http://perception.inrialpes.fr/Publications/2006/BSG06.

[49] E. BOYER. *On Using Silhouettes for Camera Calibration*, in "In Proceedings of the seventh Asian Conference on Computer Vision", LNCS, Springer, January 2006, http://perception.inrialpes.fr/Publications/2006/Boy06.

[50] E. DELPONTE, E. ARNAUD, F. ODONE, A. VERRI. *Analysis on a local approach to 3D object recognition*, in "Symposium of the German Association for Pattern Recognition, DAGM'06", sept 2006.

[51] F. DEVERNAY, D. MATEUS, M. GUILBERT. *Multi-Camera Scene Flow by Tracking 3-D Points and Surfels*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York", IEEE Computer Society, IEEE, June 2006, http://perception.inrialpes.fr/Publications/2006/DMG06.

[52] G. DEWAELE, F. DEVERNAY, R. P. HORAUD, F. FORBES. *The Alignment Between 3-D Data and Articulated Shapes with Bending Surfaces*, in "Proceedings of the 9th European Conference on Computer Vision, Graz, Austria", A. LEONARDIS, H. BISCHOF, A. PINZ (editors). , LNCS, vol. III, Springer, May 2006, p. 578–591, http://perception.inrialpes.fr/Publications/2006/DDHF06.

[53] J.-S. Franco, M. Lapierre, E. Boyer. *Visual Shapes of Silhouette Sets*, in "Proceedings of the 3rd International Symposium on 3D Data Processing, Visualization and Transmission, Chapel Hill (USA)", 2006, http://perception.inrialpes.fr/Publications/2006/FLB06.

[54] P. Gurdjos, P. Sturm, Y. Wu. *Euclidean Structure from N>=2 Parallel Circles: Theory and Algorithms*, in "Proceedings of the 9th European Conference on Computer Vision, Graz, Austria", vol. 1, Springer, may 2006, p. 238-252, http://perception.inrialpes.fr/Publications/2006/GSW06.

[55] B.-W. Hong, E. Prados, L. A. Vese, S. Soatto. *Shape Representation based on Integral Kernels: Application to Image Matching and Segmentation*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, NY, June, 2006", vol. 1, IEEE, June 2006, p. 833–840, http://perception.inrialpes.fr/Publications/2006/HPVS06.

[56] A. Jacquot, P. Sturm, O. Ruch. *3D Object Tracking Based on Multi-Model Particle Filtering*, in "Proceedings of the Sixth IEEE International Workshop on Visual Surveillance, Graz, Austria", may 2006, http://perception.inrialpes.fr/Publications/2006/JSR06.

[57] D. Knossow, R. Ronfard, R. P. Horaud, F. Devernay. *Tracking with the Kinematics of Extremal Contours*, in "Computer Vision – ACCV 2006, Hyderabad, India", P. Narayanan, S. Nayar, H.-Y. Shum (editors). , LNCS, Springer, January 2006, p. 664–673, http://perception.inrialpes.fr/Publications/2006/KRHD06.

[58] D. Knossow, J. van de Weijer, R. P. Horaud, R. Ronfard. *Articulated-body Tracking through Anisotropic Edge Detection*, in "Workshop on Dynamical Vision, ECCV'06", May 2006, http://perception.inrialpes.fr/Publications/2006/KVHR06.

[59] C. Ménier, E. Boyer, B. Raffin. *3D Skeleton-Based Body Pose Recovery*, in "Proceedings of the 3rd International Symposium on 3D Data Processing, Visualization and Transmission, Chapel Hill (USA)", june 2006, http://perception.inrialpes.fr/Publications/2006/MBR06.

[60] P. Paul, A. Quere, E. Arnaud, X. Morandi, P. Jannin. *A surface registration approach for video-based analysis of intraoperative brain surface deformations*, in "Workshop on Augmented Environments for Medical Imaging and Computer-Aided Surgery (Miccai workshop) AMI-ARC'06", oct 2006.

[61] E. Prados, O. Faugeras. *Rôle clé de la Modélisation en Shape From Shading*, in "Actes du 15ème Congrès de Reconnaissance des Formes et Intelligence Artificielle, Tours, France - RFIA 2006", AFRIF-AFIA, 2006, http://perception.inrialpes.fr/Publications/2006/PF06.

[62] E. Prados, C. Lenglet, J.-P. Pons, N. Wotawa, R. Deriche, O. Faugeras, S. Soatto. *Control Theory and Fast Marching Techniques for Brain Connectivity Mapping*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, NY, June, 2006", vol. 1, IEEE, June 2006, p. 1076–1083, http://perception.inrialpes.fr/Publications/2006/PLPWDFS06a.

[63] E. Prados, S. Soatto. *Méthode de "Fast Marching" générique pour le "Shape From Shading"*, in "Actes du 15ème Congrès de Reconnaissance des Formes et Intelligence Artificielle, Tours, France - RFIA 2006", AFRIF-AFIA, 2006, http://perception.inrialpes.fr/Publications/2006/PS06.

[64] T. Pribanić, P. Sturm, P. Bačić. *A Human Body Model for Movement Analysis Using Optoelectronic System*, in "Proceedings of the 5th World Congress of Biomechanics, Munich, Germany", 2006, http://perception.inrialpes.fr/Publications/2006/PSB06.

[65] T. PRIBANIĆ, P. STURM, I. BRIGIĆ. *Body segments angle computation from single view*, in "Proceedings of the 9th International Symposium on the 3-D Analysis of Human Movement, Valenciennes, France", 2006, http://perception.inrialpes.fr/Publications/2006/PSB06a.

[66] S. RAMALINGAM, P. STURM, E. BOYER. *A Factorization Based Self-Calibration for Radially Symmetric Cameras*, in "Proceedings of the Third International Symposium on 3D Data Processing, Visualization and Transmission, Chapel Hill, USA", jun 2006, http://perception.inrialpes.fr/Publications/2006/RSB06.

[67] S. RAMALINGAM, P. STURM, S. LODHA. *Theory and Calibration Algorithms for Axial Cameras*, in "Proceedings of the Asian Conference on Computer Vision, Hyderabad, India", vol. I, jan 2006, p. 704–713, http://perception.inrialpes.fr/Publications/2006/RSL06.

[68] R. SOSA, E. ARNAUD, G. ARTANA. *Study of induced flow by a sliding descharge*, in "Proceedings of the international Symposium on Electrohydrodynamics, Buenos Aires, Argentina", dec 2006.

[69] R. SOSA, E. ARNAUD, E. MEMIN, G. ARTANA. *Schlieren Image Velocimetry applied to EHD flow*, in "Proceedings of the international Symposium on Electrohydrodynamics, Buenos Aires, Argentina", dec 2006.

[70] P. STURM, T. BONFORT. *How to Compute the Pose of an Object without a Direct View?*, in "Proceedings of the Asian Conference on Computer Vision, Hyderabad, India", vol. II, jan 2006, p. 21–31, http://perception.inrialpes.fr/Publications/2006/SB06.

[71] J.-P. TARDIF, P. STURM, S. ROY. *Self-Calibration of a General Radially Symmetric Distortion Model*, in "Proceedings of the 9th European Conference on Computer Vision, Graz, Austria", Lecture Notes in Computer Science, vol. 4, Springer, may 2006, p. 186-199, http://perception.inrialpes.fr/Publications/2006/TSR06.

[72] D. WANG, E. PRADOS, S. SOATTO. *Towards Robust and Physically Plausible Shaded Stereoscopic Segmentation*, in "Computer Vision and Pattern Recognition Workshop (CVPRW), 2006 Conference on", IEEE, 2006, p. 171–171, http://perception.inrialpes.fr/Publications/2006/WPS06.

[73] D. WEINLAND, R. RONFARD, E. BOYER. *Automatic Discovery of Action Taxonomies from Multiple Views*, in "In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", 2006, http://perception.inrialpes.fr/Publications/2006/WRB06.

[74] A. ZAHARESCU, R. P. HORAUD, R. RONFARD, L. LEFORT. *Multiple Camera Calibration using Robust Perspective Factorization*, in "Proceedings of the 3rd International Symposium on 3D Data Processing, Visualization and Transmission, Chapel Hill (USA)", IEEE Computer Society Press, 2006, http://perception.inrialpes.fr/Publications/2006/ZHRL06.

Internal Reports

[75] D. KNOSSOW, R. RONFARD, R. P. HORAUD. *Human Motion Tracking with a Kinematic Parameterization of Extremal Contours*, Research Report, n^{o} RR-6007, INRIA, 655 avenue de l'Europe, 38330 Montbonnot St Martin, October 2006, http://perception.inrialpes.fr/Publications/2006/KRH06.

References in notes

[76] A. JACQUOT, P. STURM, O. RUCH. *Adaptive Tracking of Non-Rigid Objects Based on Color Histograms and Automatic Parameter Selection*, in "Proceedings of the ieee Workshop on Motion and Video Computing, Breckenridge, Colorado, USA", January 2005, p. 103–109.

[77] P. STURM, L. QUAN. *Affine Stereo Calibration*, Technical report, n^{o} LIFIA-29, lifia–imag, Grenoble, France, 1995.