INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Magrit

Visual Augmentation of Complex Environments

Nancy - Grand Est

Theme : Vision, Perception and Multimedia Understanding

Table of contents

1. Team

Research Scientist

Marie-Odile Berger [Research associate (CR) INRIA,Team Leader, HdR]
Erwan Kerrien [Research associate (CR)]

Faculty Member

Gilles Simon [Assistant professor, Université Henri Poincaré]
Frédéric Sur [Assistant professor, Institut National Polytechnique de Lorraine]
Pierre-Fréderic Villard [Assistant professor, Université Henri Poincaré, since September 2009]
Brigitte Wrobel-Dautcourt [Assistant professor, Université Henri Poincaré]

External Collaborator

René Anxionnat [Medical Doctor, PhD, Professor CHRU Nancy]

Technical Staff

Blaise Potard [Engineer, from January to August 2009]

PhD Student

Michael Aron [INRIA, since October 2005]
Srikrishna Bhat [INRIA, since December 2008]
Nicolas Noury [INRIA since September 2006]
Nicolas Padoy [ENS fellow, AMN, since October 2005]

Post-Doctoral Fellow

Evren Imre [INRIA, until December 2008]
Ting Peng [INRIA, since December 2008]

Administrative Assistant

Isabelle Herlich [INRIA]

2. Overall Objectives

2.1. Introduction

Augmented reality (AR) is a field of computer research which deals with the combination of real world and computer generated data in order to provide the user with a better understanding of his surrounding environment. Usually this refers to a system in which computer graphics are overlaid onto a live video picture or projected onto a transparent screen as in a head-up display.

Though there exist a few commercial examples demonstrating the effectiveness of the AR concept for certain applications, the state of the art in AR today is comparable to the early years of Virtual Reality. Many research ideas have been demonstrated but few have matured beyond lab-based prototypes.

Computer vision plays an important role in AR applications. Indeed, the seamless integration of computer generated objects at the right place according to the motion of the user needs automatic real-time detection and tracking. In addition, 3D reconstruction of the scene is needed to solve occlusions and light inter-reflexion between objects and to make easier the interactions of the user with the augmented scene. Since fifteen years, much work has been successfully devoted to the problem of structure and motion, but these works are often formulated as off-line algorithms and require batch processing of several images acquired in a sequence. The challenge is now to design robust solutions to these problems with the aim to let the user free of his motion during AR applications and to widen the range of AR application to large and/or unstructured environments. More specifically, the Magrit team aims at addressing the following problems:

- On-line pose computation for structured and non structured environments: this problem is the cornerstone of AR systems and must be achieved in real time with a good accuracy.

- Long term management of AR applications: a key problem of numerous algorithms is the gradual drifting of the localization over time. One of our aims is to develop methods that improve the accuracy and the repeatability of the pose during arbitrarily long periods of motion.
- 3D modeling for AR applications: this problem is fundamental to manage light interactions between real and virtual objects, to solve occlusions and to obtain realistic fused images.

2.2. Highlights

Nicolas Padoy, Diana Mateus, Daniel Weinland, Marie-Odile Berger and Nassir Navab received the best paper award for the paper "Workflow Monitoring Based on 3D Motion Features" at VOEC'09 (Workshop on Video-Oriented Object and Event Classification, ICCV Worshop).

3. Scientific Foundations

3.1. Scientific Foundations

The aim of the Magrit project is to develop vision based methods which allow significant progress of AR technologies in terms of ease of implementation, usability, reliability and robustness in order to widen the current application field of AR and to improve the freedom of the user during applications. Our main research directions concern two crucial issues, camera tracking and scene modeling. Methods are developed with a view to meet the expected robustness and to provide the user with a good perception of the augmented scene.

3.1.1. Camera calibration and registration

Keywords: *Registration*, *augmented reality*, *tracking*, *viewpoint computation.*

One of the most basic problems currently limiting Augmented Reality applications is the registration problem. The objects in the real and virtual worlds must be properly aligned with respect to each other, or the illusion that the two worlds coexist will be compromised.

As a large number of potential AR applications are interactive, real time pose computation is required. Although the registration problem has received a lot of attention in the computer vision community, the problem of real-time registration is still far from being a solved problem, especially for unstructured environments. Ideally, an AR system should work in all environments, without the need to prepare the scene ahead of time, and the user should walk anywhere he pleases.

For several years, the Magrit project has been aiming at developing on-line and markerless methods for camera pose computation. We have especially proposed a real-time system for camera tracking designed for indoor scenes [1]. The main difficulty with online tracking is to ensure robustness of the process. For off-line processes, robustness is achieved by using spatial and temporal coherence of the considered sequence through move-matching techniques. To get robustness for open-loop systems, we have developed a method which combines the advantage of move-matching methods and model-based methods [6] by using a piecewise-planar model of the environment. This methodology can be used in a wide variety of environments: indoor scenes, urban scenes ... We are also concerned with the development of methods for camera stabilization. Indeed, statistical fluctuations in the viewpoint computations lead to unpleasant jittering or sliding effects, especially when the camera motion is small. We have proved that the use of model selection allows us to noticeably improve the visual impression and to reduce drift over time.

An important way to improve the reliability and the robustness of pose algorithms is to combine the camera with another form of sensor in order to compensate for the shortcomings of each technology. Each technology approach has limitations: on the one hand, rapid head motions cause image features to undergo large motion between frames that can cause visual tracking to fail. On the other hand, inertial sensors response is largely independent from the user's motion but their accuracy is bad and their response is sensitive to metallic objects in the scene. We have proposed a system that makes an inertial sensor cooperate with the camera-based system in order to improve the robustness of the AR system to abrupt motions of the users, especially head motions. This work contributes to reduce the constraints on the users and the need to carefully control the environment during an AR application [1]. This research area has been continued within the ASPI project in order to build a dynamic articulatory model from various image modalities and sensor data.

Obtaining a model of the scene where the AR applications is to take place is often required by pose algorithms. However, obtaining a model either by automatic or interactive means is a tedious task, especially for large environments. In addition, models may be described in terms of 3D features which cannot be identified in the images. Pose by recognition is thus an appealing approach which allows to link photometric knowledge learned on the scene to the camera pose. We are currently considering learning-based techniques, the aim of which is to allow pose computation from video sequences previously acquired on the site where the application is to be used.

Finally, it must be noted that the registration problem must be addressed from the specific point of view of augmented reality: the success and the acceptance of an AR application does not only depend on the accuracy of the pose computation but also on the visual impression of the augmented scene. The search for the best compromise between accuracy and perception is therefore an important issue in this project. This research topic has been addressed in our project both in classical AR [7] and in medical imaging in order to choose the camera model, including intrinsic parameters, which describes at best the considered camera.

3.1.2. Scene modeling

Keywords: *Fusion*, *medical imaging*, *reconstruction*.

Modeling the scene is a fundamental issue in AR for many reasons. First, pose computation algorithms often use a model of the scene or at least some 3D knowledge on the scene. Second, effective AR systems require a model of the scene to support occlusion and to compute light reflexions between the real and the virtual objects. Unlike pose computation which has to be computed in a sequential way, scene modeling can be considered as an off-line or an on-line problem according to the application.

In our past activities, scene modeling was mainly addressed as an off-line and possibly interactive process, especially to build models for medical imaging from several images modalities. Since three years, one of our research directions is about online scene reconstruction, with the aim to be able to handle AR applications in vast environments without the need to instrument the scene.

Interactive scene modeling from various image modalities is mainly considered in our medical activities. For the last 15 years, we have been working in close collaboration with the neuroradiology laboratory (CHU-University Hospital of Nancy) and GE Healthcare. As several imaging modalities are now available in a per-operative context (2D and 3D angiography, MRI, ...), our aim is to develop a multi-modality framework to help therapeutic decision and treatment.

We have mainly been interested in the effective use of a multimodality framework in the treatment of arteriovenous malformations (AVM) and aneurysms in the context of interventional neuroradiology. The goal of interventional gestures is to guide endoscopic tools towards the pathology with the aim to perform embolization of the AVM or to fill the aneurysmal cavity by placing coils. An accurate definition of the target is a parameter of great importance for the success of the treatment. We have proposed and developed multimodality and augmented reality tools which make cooperate various image modalities (2D and 3D angiography, fluoroscopic images, MRI, ...) in order to help physicians in clinical routine. One of the success of this collaboration is the implementation of the concept of *augmented fluoroscopy* [4], which helps the surgeon to guide endoscopic tools towards the pathology. Lately, in cooperation with the Alcove EPI, we have

proposed new methods for implicit modeling of the aneurysms with the aim to obtain near real time simulation of the coil deployment in the aneurysm [3]. Multi-modality techniques for reconstruction are also considered within the european ASPI project, the aim of which is to build a dynamic model of the vocal tract from various images modalities (MRI, ultrasound, video) and magnetic sensors.

On-line reconstruction of the scene structure needed by pose or occlusion algorithms is highly desirable for numerous AR applications for which instrumentation is not conceivable. Hence, structure and pose must be sequentially estimated over time. This process largely depends on the quality of the matching stage which allows to detect and to match features over the sequence. Ongoing research are thus conducted on the use of probabilistic methods to establish robust correspondences of features over time. The use of *a contrario* decision is especially under study to achieve this aim [5].

Most automatic techniques aim at reconstructing a sparse and thus unstructured set of points of the scene. Such models are obviously not appropriate to perform interaction with the scene. In addition, they are incomplete in the sense that they may omit features which are important for the accuracy of the pose recovered from 2D/3D correspondences. We have thus investigated interactive techniques with the aim to obtain reliable and structured models of the scene. The goal of our approach is to develop immersive and intuitive interaction techniques which allow scene modeling during the application [19].

4. Application Domains

4.1. Augmented reality

We have a significant experience in the AR field especially through the European project ARIS (2001–2004) which aimed at developing effective and realistic AR systems for e-commerce and especially for interior design. Beyond this restrictive application field, this project allowed us to develop nearly real time camera tracking methods for multi-planar environments. Since then, we have amplified our research on multi-planar environments in order to obtain effective and robust AR systems in such environments. We currently investigate both automatic and interactive techniques for scene reconstruction/structure from motion methods in order to be able to consider large and unknown environments.

4.2. Medical imaging

For 15 years, we have been working in close collaboration with the University hospital of Nancy and GE Healthcare in interventional neuroradiology with the aim to develop tools allowing the physicians to take advantage of the various existing imaging modalities on the brain in their clinical practice. As several imaging modalities that bring complementary information on the various brain pathologies are now available in a pre-operative context (2D and 3D subtracted angiography, fluoroscopy, MRI,...) our aim is to develop a multi-modality framework to help therapeutic decisions. Recently, we have investigated the use of AR tools for neuronavigation. The aim of the PhD thesis of Sebastien Gorges which ended in May 2007 was to design tools for neuronavigation that take advantage of a real-time imagery (fluoroscopy) and a pre-operative 3D imagery (3D angiography). We are currently involved in the SOFA project with the ALCOVE team and the University hospital of Nancy. Our aim is to develop simulation tools of the interventional act adapted to the patient's anatomy and physiology, in order to help the surgeon with planning the coil placement, rehearsing the therapeutic gesture, and to provide new tools to improve the medical training to the technique.

4.3. Augmented head

There is a strong evidence that visual information of the speaker, especially jaws and lips but also tongue position, noticeably improves the speech intelligibility. Hence, having a realistic augmented head displaying both external and internal articulators could help language learning technology in giving the student a feedback on how to change articulation in order to achieve a correct pronunciation. This task is complex and necessitates a multidisciplinary effort involving speech production modeling and image analysis. The long term aim of the

project is the design of a 3D +t articulatory model to be used for the realistic animation of an augmented/talking head. Within this project, we have especially worked on the tracking of the visible articulators using stereo-vision techniques and we intend to supplement the model with internal articulators (tongue, larynx) obtained from medical imaging (ultrasound images for tongue tracking and MRI for global model). These activities have been conducted within the European ASPI project (2005-2009) and are continued within the ANR ARTIS (2009-2012).

5. Software

5.1. RAlib

Our software efforts are integrated in a library called RAlib which contains our research development on image processing, registration (2D and 3D) and visualization. This library is licensed by the APP (French agency for software protection).

The visualization module is called QGLSG: it enables the visualization of images, 2D and 3D objects under a consistent perspective projection. It is based on Qt (http://www.trolltech.com) and OpenScenegraph (http://www.openscenegraph.org/projects/osg) libraries. The latter was integrated in the project by Frédéric Speisser, who was part of Magrit project-team between September 2006 and September 2008 as an INRIA assistant engineer. The QGLSG library integrates innovative features such as online camera distortion correction (which has since been integrated in the latest releases of OpenScengraph, though independently from our code), and invisible objects that can be incorporated in a scene so that virtual objects can cast shadows on real objects, and occlusion between virtual and real objects are easier to handle. The library was also ported to Mac OS and Windows and a full doxygen documentation was written.

The library was consolidated this year through the design of applications used internally within Talking-Head related projects, but also Augmented Reality projects. In particular, a software called iSketchup was shown at ISMAR conference this year to demonstrate our new results on interactive scene modeling.

6. New Results

6.1. New Results

6.1.1. Scene and camera reconstruction

Participants: Marie-Odile Berger, Srikrishna Bhat, Evren Imre, Nicolas Noury, Gilles Simon, Frédéric Sur.

On the theme of scene and camera reconstruction, we investigate both fully automatic methods and learning-based techniques for pose and structure recovery. Interactive techniques are also considered in order to obtain well-structured description of the scene and to meet the required robustness with the help of the user.

6.1.1.1. Structure from motion via a contrario models

Structure from motion problems call for probabilistic frameworks to meet robustness requirements. Features (e.g. points of interest) are extracted from images, then they are matched under projective constraints. This determines both the structure of the scene and the position of the camera. The problem is difficult since the position of the features may not be accurately known, and matching may introduce false correspondences which can endanger the reconstruction process. We aim at developing new probabilistic methods to tackle these problems. This year we focused on two directions. On the one hand we studied a probabilistic *a contrario model* to incorporate point location uncertainty in a Ransac-like robust matching algorithm. On the other hand we brought to completion our previous work about point of interest matching based on epipolar constraint and photometric consistency. The proposed algorithm gives interesting results with respect to repeated patterns and strong viewpoint changes [20].

6.1.1.2. Improved inverse-depth parameterization for SLAM

The monocular simultaneous localization and mapping (SLAM) problem involves the estimation of the location of a set of landmarks in an unknown environment (mapping) as well as the estimation of the camera pose via the photometric measurements of these landmarks by a camera. Since the computational complexity of the structure from motion techniques is deemed prohibitively high, the literature is dominated by extended Kalman filter (EKF) and particle filter (PF) based approaches. However, the non-Gaussianity of the depth estimate uncertainty degrades the performance of EKF-SLAM systems that use a 3-d cartesian landmark parameterization, especially in low parallax configurations. The inverse depth parameterization (IDP) proposed in [24] alleviates this problem through a redundant representation. In addition, this approach successfully deals with the feature initialization problem in monocular SLAM. However, it is computationally expensive, and when a set of landmarks is initialized from the same image, it fails to enforce the common origin constraint. We thus proposed in [16] two improvements of the classical inverse-depth parameterization. The key-idea is to factor out the common pose parameters when several landmarks are initialized from the same image. In the first extension (IDP1), only the pose is factored out whereas in the second one (IDP2), both the pose and the orientation are factored out. Experiments proved that IDP2 is superior to the classical IDP both in computational cost and in performance, whereas IDP1 delivers a similar performance at a much lower computational cost. This approach is also useful in particle filter based SLAM systems as the landmarks are estimated with a Kalman filter [15].

6.1.1.3. Learning-based techniques for pose computation

Recent advances in object recognition based on local descriptors have shown the possibility of efficient image matching and retrieval from a database [25] and pave the way towards more robust methods for pose computation. Most approaches attempt to quantize the SIFT descriptors extracted from a set of images of the environment into clusters, called visual words, which are likely to represent a unique feature of the world. This year, we started to investigate the joint use of tracking based methods and recognition methods with a view to handle large environments in AR applications.

Because Euclidean distance between SIFT descriptors fails to provide a good dissimilarity measure, we have devised a different way of forming visual words using transitive closure relationships: two SIFT features belong to the same visual word if their Euclidean distance is less than a given threshold. An object is then represented as a set of feature vectors instead of a single feature vector. Our experiments proved that this representation allows noticeable improvements of the robustness of detection. Unfortunately, representing a word with a list of vectors is not scalable. We are thus investigating methods to find a suitable distance measure from the visual words obtained on a short video of the environment where the AR applications has to take place. Our objective is to obtain specific representations which can be efficiently matched.

6.1.1.4. Online reconstruction for AR tasks

Acquiring the 3D geometry of arbitrary scenes has been a primary objective of both the computer vision and graphics communities for many decades. Applications are numerous in various domains such as construction, GIS and 3D maps, virtual tours, visual effects and AR. Existing modeling methods usually rely on two separate stages. First, some data about the scene (photographs, videos, laser measurements, ...) are acquired on-site. Then these data are processed off-line using specific manipulations and algorithms. Unfortunately, this process can be time-consuming and tedious. Moreover, there is no guarantee after the first stage that the required model is fully extractable from the acquired data and additional acquisitions are sometimes needed to supplement the missing parts. We thus propose to bridge the gap between data acquisition and their exploitation. A purely image-based system has been developed, which allows a user to interactively capture the 3D geometry of a polyhedral scene with the aid of its physical presence [19].

This system can be seen as an immersive version of the widely used 3D drawing software Google SketchUp™ (http://sketchup.google.com). This software combines some of the features of pencil-and-paper sketching and some of the features of CAD systems to provide a lightweight, gesture-based interface for 3D polyhedral modeling. By indicating two orthogonal vanishing points, the user is able to align the world axes to match a photo perspective. With this done, he can create models using the photo as a direct reference; mouse strokes

are converted into 3D-space using inverse ray intersections with the previously defined geometry or the ground plane by default. These principles have been taken up in our implementation, but with the crucial difference that we consider dynamic video images instead of static ones. The system alternates between two operating modes: (i) a modeling mode where the scene geometry is defined by applying pure rotations to the camera and using an eye cursor to perform point-and-click operations and (ii) a tracking mode, where 6 degrees-of-freedom camera tracking is performed based on the available geometry, enabling the user to get closer to some parts of the scene or make some new faces visible before keeping on modeling. Switches between these two modes are done automatically using Akaike's model selection. As a result, we get a user-friendly interface which is particularly suitable for mobile devices such as PDAs and mobile phones.

6.1.2. Medical imaging

Participants: René Anxionnat, Marie-Odile Berger, Erwan Kerrien, Nicolas Padoy, Pierre-Frédéric Villard.

6.1.2.1. Simulation for planning the embolization of intracranial aneurisms

The endovascular treatment for an intracranial aneurism consists in filling the aneurismal cavity by placing coils. These are sorts of long platinum springs that, once deployed, wind into a compact ball. Considering the location of the lesion, close to the brain, and its small size, a few millimeters, the interventional gesture requires a good planning and cannot but be performed by a very experienced surgeon. A simulation tool of the interventional act, available in the operating room, reliable, adapted to the patient's anatomy and physiology, would help to plan the coil placement, rehearse the procedure, and improve the medical training to the technique.

Our research activity is led in collaboration with Alcove project-team at INRIA Lille-Nord Europe and the Department of Interventional Neuroradiology at University Hospital of Nancy. It started in 2007 with the SIMPLE project (INRIA colaborative research initiative (ARC)) and was pursued this year in the context of the SOFA-InterMedS initiative (INRIA large-scale initiative action (AE)).

Our task consists in providing precise in-vivo data about the patient and in particular a precise geometric model of the patient's arterial wall. Despite the very high quality of the available 3D images (3D rotational angiography), tomographic reconstruction artefacts perturb the isosurface that should correspond to the arterial wall. Taking this isosurface as an initialization, we proposed to improve it within an active surface framework where the arterial wall is deformed until its X-ray projection fits a set a registered 2D angiographic images taken on the patient.

This year, we first addressed the validation of our model both on silicon phantoms and actual patient data. Our models were used in conjunction with a first prototype developed by Alcove project-team on SOFA software platform (http://www.sofa-framework.org) to simulate coil deployment under various conditions on real patient data. The methodology we followed, the clinical metrics we designed and the results of our investigations were presented in major conferences, both in medicine [10] and in medical imaging [14].

However, our algorithm produces a triangulated mesh model for the arterial wall, which requires a difficult compromise to be made on the simulation side between real time processing and the physical realism of the coil behavior. Therefore, we started to investigate the implicit modeling of blood vessels in 3D, in particular radial basis functions (RBF) with Pierre Glanc's engineering internship. This work will be pursued by a PhD student, under the shared direction of Magrit and Alcove teams, whom we shall welcome at the end of this year. The major axes of research concern the design of the profile function of the RBF, model fitting to the data, as well as the compacity of the model, in order to ensure both real-time simulation and geometric accuracy of the model.

6.1.2.2. Surgical workflow analysis

The focus of this work is the development of statistical methods that permit the modeling and monitoring of surgical processes, based on signals available in the surgery room. Previous works in the domain of activity recognition have addressed different problems such as the identification of either isolated actions or well-defined interactions among objects in a scene. In this work, we address the activity recognition problem in the context of a workflow. In this case, activities follow a well-defined structure over a long period of time and can

be semantically grouped in relevant phases. The major characteristics of the phase recognition problem are the temporal dependencies between phases and their highly varying durations. We have addressed the problem of recognizing phases, based on exemplary recordings. We have proposed to use Workflow-HMMs, a form of HMMs augmented with phase probability variables that model the complete workflow process [17]. This model takes into account the full temporal context which improves on-line recognition of the phases, especially in case of partial labeling. Targeted applications are workflow monitoring in hospitals and factories, where common action recognition approaches are difficult to apply. To avoid interfering with the normal workflow, we capture the activity of a room with a multiple-camera system. Additionally, we propose to rely on real-time low-level features (3D motion flow) to maintain a generic approach. Our method has been successfully demonstrated on sequences of medical procedures performed in a mock-up operating room. The sequences followed a complex workflow, containing various alternatives.

6.1.3. Modeling face and vocal tract dynamics

Participants: Michael Aron, Marie-Odile Berger, Erwan Kerrien, Ting Peng, Blaise Potard, Brigitte Wrobel-Dautcourt.

Being able to produce realistic facial animation is crucial for many speech applications in language learning technologies. In order to reach realism, it is necessary to acquire 3D models of the face and of the internal articulators (tongue, palate,...) from various image modalities.

6.1.3.1. A shape-based variational framework for curve segmentation

MRI provides us with a convenient and powerful tool for observing the internal articulators which are involved in speech production. In this study, we acquired 3D MRI data with a group of articulations from different speakers. With the help of the tongue model of a reference speaker, we aim to extract tongue contours from mid-sagittal images of a new speaker, and then to build his/her tongue model. This will enable us, in the future, to compare tongue models between speakers, and explore how to adapt the reference speaker's tongue model to the new speaker. To reach this aim, we have proposed a shape-based variational framework to curve evolution for the segmentation of tongue contours from MRI mid-sagittal images. The method starts with the construction of a PCA model on tongue contours of different articulations of a reference speaker. Tongue contours for a new speaker are constrained to belong to this shape space. An objective function is defined which integrates both global and local image information. The global term extracts roughly the object in the whole image domain; while the local term improves precision inside a small neighborhood around the contour. Promising results on several speaker's MRI data and comparisons with other approaches demonstrated the efficiency of our new model [18].

6.1.3.2. Modeling the vocal tract

Our long term objective is to provide intuitive and near-automatic tools for building a dynamic 3D model of the vocal tract from various image and sensor modalities (MRI, ultrasound (US), video, magnetic sensors ...).

Combining several modalities requires that all geometrical and temporal data be consistent together. It also requires to define appropriate image processing techniques to extract the articulators (tongue, palate, lips...) from the data. A fast, low cost and easily reproducible acquisition system had been designed in order to temporally align the data in a previous work [22]. This year, we focussed on the problem of fusing image modalities [11]. As 3D measures of the face can be extracted from both MRI and stereoscopic images, MRI and video sequences were registered through an iterative closest point algorithm. All the modalities were then registered using EM sensors glued on the US probe and under the speaker's ears. As a result, dynamic articulatory data including points on the lips, the tongue and the palate are now available. These data were used very recently to perform articulatory inversion. To the best of our knowledge, this is the first work that demonstrates the potential of static and dynamic data fusion in the construction of articulatory databases.

We also addressed this year the problem of assessing the quality of the obtained fused data. This amounts to evaluate the uncertainty on each transformation used used to align the data in a common coordinate system. Monte Carlo statistical methods were used to estimate the uncertainty on this complex registration process: starting from the uncertainty on the sensors and on the image features, we are able to estimate the accuracy on

each articulator through exhaustive sampling and propagation techniques [21]. This study enabled us to isolate the major sources of error in the registration process. Not surprisingly EM sensors are an important factor, but US resolution was also found to be critical. uncertainty on articulatory data.

6.1.3.3. Realistic face animation

Reaching realism in facial animation needs to acquire and to animate dense 3D models of the face which are often acquired with 3D scanners. However, acquiring the dynamics of the speech from 3D scans is difficult as the acquisition time generally allows only sustained sounds to be recorded. On the contrary, acquiring the speech dynamics on a sparse set of points is easy using a stereovision system recording a speaker with markers painted on his/her face. We have proposed in [12] an approach to animate a very realistic dense talking head which makes use of a reduced set of 3D dense meshes acquired for sustained sounds as well as the speech dynamics learned on a speaker equiped with painted markers. Our contributions are twofold: We first proposed an appropriate principal component analysis (PCA) with missing data techniques in order to compute the basic modes of the speech dynamics despite possible unobservable points in the sparse meshes obtained by the stereovision system. A method for densifying the modes was then proposed to compute dense modes for spatial animation from the sparse modes learned with the stereovision system.

7. Other Grants and Activities

7.1. National Initiatives

7.1.1. SOFA-InterMedS

Participants: René Anxionnat, Marie-Odile Berger, Erwan Kerrien, Pierre Glanc.

The SOFA-InterMedS large-scale INRIA initiative (http://www.inria.fr/recherche/equipes/sofa-intermeds.en.html) is a research-oriented collaboration across several INRIA teams, international research groups and clinical partners. Its main objective is to leverage specific competences available in each team to further develop the multidisciplinary field of Medical Simulation research.

Our action within the initiative takes place in close collaboration with both Alcove INRIA project-team in Lille and the Department of diagnostic and therapeutic interventional neuroradiology of Nancy University Hospital. We aim at providing in-vivo models of the patient's organs, and in particular a precise geometric model of the arterial wall. Such a model is used by Alcove team to simulate the coil deployment within an intracranial aneurysm. The associated medical team in Nancy, and in particular our external collaborator René Anxionnat, is in charge of validating our results.

In order to consolidate last year's results on blood vessel modeling as triangulated meshes, we focused on the first half of this year on validating our simulation results on real patient data. This involved designing a methodology that implied defining tests and describing clinical metrics. Our findings were presented in major conferences both in medicine [10] and medical imaging [14].

One of the conclusions was that the current mesh model induces a very tight compromise to be made between real-time simulation and the physical realism of the coil behavior. On-going investigations concern using radial basis functions to build an implicit model of the blood vessels, with an emphasis on preserving the accuracy of the geometric model while improving simulation time.

7.1.2. ANR ARTIS (2009-2012)

Participants: Marie-Odile Berger, Erwan Kerrien, Ting Peng.

The main objective of this fundamental research project is to develop inversion tools and to design and implement methods that allows producing augmented speech from the speech sound signal alone or with video images of the speaker's face. The Magrit team is especially concerned with the development of procedures allowing the automatic construction of a speaker's model from various imaging modalities. This year, we have developed variational guided methods which allow the internal articulators of a speaker to be segmented from MRI images given the knowledge of these articulators for a reference speaker.

7.1.3. ANR Visac (2009-2012)

Participants: Marie-Odile Berger, Brigitte Wrobel-Dautcourt, Blaise Potard.

The ANR Visac is about acoustic-visual speech synthesis by bimodal concatenation. The major challenge of this project is to perform speech synthesis with its acoustic and visible components simultaneously. Within this project, the role of the magrit team is twofold. One of them is to build a stereovision system able to record synchronized audio-visual sequences at a high frame rate. Second, a highly realistic dense animation of the head must be produced. During this first year, the stereovision system was noticeably improved in order to make possible the recording of long audio-visual sequences. The acquired corpus has been processed in order to obtain a corpus of 3D faces.

7.2. International initiatives

7.2.1. Animation of the scientific community

- M.-O. Berger was a member of the program committee of the conferences ISMAR 09 (International Symposium on Mixed and Augmented Reality), MICCAI 09 (International Conference on Medical Image Computing and Computer Assisted Intervention), RFIA 2010 (Reconnaissance des formes et intelligence artificielle) and of the workshops AMI-ARCS and M2CAI (workshop on Modeling and Monitoring of Computer Assisted Interventions).
- E. Kerrien was a member of the program committe of ORASIS 2009
- G. Simon was a member of the program committee of ISMAR 09 and ORASIS 2009.
- Pierre-Frédéric Villard was a member of the program committee of MICCAI and ISBMS (International Symposium on Biomedical Simulation).
- The members of the team frequently review articles and papers for IJCARS (International Journal of Computer Assisted Radiology and Surgery), TVCG (IEEE Transactions on Visualization and Computer Graphics), CAVW (Computer Animation and Virtual Worlds), EURASIP Journal on Image and Video Processing, IEEE Transactions on Image Processing, Signal, Image and Video Processing, SIAM Journal on Imaging Sciences.

8. Dissemination

8.1. Teaching

- Several members of the group, in particular assistant professors and Ph.D. students, actively teach at Henri Poincaré Nancy 1, Nancy 2 universities and INPL.
- Other members of the group also teach in the computer science Master of Nancy, in the “Master en sciences de la vie et de la santé” (SVS) and in the ”DIU Chirurgie Robotique ”.

8.2. Participation to conferences and workshops

Members of the group participated in the following events: International Conference on Acoustics, Speech, and Signal Processing (Taipei), International Conference on Robotics and Automation (Kobe), DAGM (Jena Allemagne), International Symposium on Mixed and Augmented Reality (Orlando, USA), International Conference on Medical Image Computing and Computer Assisted Intervention (Londres).

9. Bibliography

Major publications by the team in recent years

[1] M. ARON, G. SIMON, M.-O. BERGER. *Use of Inertial Sensors to Support Video Tracking*, in "Computer Animation and Virtual Worlds", vol. 18, 2007, p. 57-68, http://hal.inria.fr/inria-00110628/en/.

[2] M.-O. BERGER, R. ANXIONNAT, E. KERRIEN, L. PICARD, M. SODERMAN. *A methodology for validating a 3D imaging modality for brain AVM delineation: Application to 3DRA.*, in "Computerized Medical Imaging and Graphics", vol. 32, 2008, p. 544-553, http://hal.inria.fr/inria-00321688/en/.

[3] J. DEQUIDT, M. MARCHAL, C. DURIEZ, E. KERRIEN, S. COTIN. *Interactive Simulation of Embolization Coils: Modeling and Experimental Validation*, in "Medical Imaging Computing and Computer Assisted Intervention, MICCAI Lecture Notes in Computer Science, USA", vol. 5241, 2008, p. 695-702, http://hal.inria.fr/inria-00336907/en/.

[4] S. GORGES, E. KERRIEN, M.-O. BERGER, Y. TROUSSET, J. PESCATORE, R. ANXIONNAT, L. PICARD, S. BRACARD. *3D Augmented Fluoroscopy in Interventional Neuroradiology: Precision Assessment and First Evaluation on Clinical Cases*, in "Workshop on Augmented environments for Medical Imaging and Computer-aided Surgery - AMI-ARCS 2006 (held in conjunction with MICCAI'06), Copenhagen Denmark", Wolfgang Birkfellner, Nassir Navab and Stephane Nicolau, 11 2006, http://hal.inria.fr/inria-00110850/en/.

[5] P. MUSÉ, F. SUR, F. CAO, Y. GOUSSEAU, J.-M. MOREL. *An a contrario decision method for shape element recognition*, in "International Journal of Computer Vision", vol. 69, n^{o} 3, 2006, p. 295-315, http://hal.inria.fr/inria-00104260/en/.

[6] G. SIMON, M.-O. BERGER. *Pose estimation for planar structure*, in "IEEE Computer Graphics and Applications", vol. 22, n^{o} 6, 2002, p. 46-53.

[7] J.-F. VIGUERAS, G. SIMON, M.-O. BERGER. *Calibration Errors in Augmented Reality: A Practical Study*, in "International Symposium on Mixed and Augmented Reality, Vienna Astria", 10 2005, p. 154-163, http://hal.inria.fr/inria-00000383/en/.

Year Publications

Doctoral Dissertations and Habilitation Theses

[8] M. ARON. *Acquisition et modélisation de données articulatoires dans un contexte multimodal*, Université Henri Poincaré - Nancy I, 11 2009, http://tel.archives-ouvertes.fr/tel-00432124/en/, Ph. D. Thesis.

Articles in International Peer-Reviewed Journal

[9] P.-F. VILLARD, F. P. VIDAL, C. HUNT, F. BELLO, N. W. JOHN, S. JOHNSON, D. A. GOULD. *A prototype percutaneous transhepatic cholangiography training simulator with real-time breathing motion*, in "International Journal of Computer Assisted Radiology and Surgery", vol. 4, n^{o} 6, 2009-11, p. 571–583, http://hal.archives-ouvertes.fr/hal-00430223/en/GB.

Articles in Non Peer-Reviewed Journal

[10] R. ANXIONNAT, F. ROCCA, S. BRACARD, J. DEQUIDT, E. KERRIEN, C. DURIEZ, M.-O. BERGER, S. COTIN. *Evaluation of a computer-based simulation for the endovascular treatment of intracranial aneurysms*, 2009, http://hal.inria.fr/inria-00432289/en/.

International Peer-Reviewed Conference/Proceedings

[11] M. ARON, A. TOUTIOS, M.-O. BERGER, E. KERRIEN, B. WROBEL-DAUTCOURT, Y. LAPRIE. *Registration of Multimodal Data for Estimating the Parameters of an Articulatory Model*, in "IEEE International

Conference on Acoustics, Speech, and Signal Processing (ICASSP), Taiwan", 2009, http://hal.inria.fr/inria-00350298/en/.

[12] M.-O. BERGER, J. PONROY, B. WROBEL-DAUTCOURT. *Realistic Face Animation for Audiovisual Speech Applications: A Densification Approach Driven by Sparse Stereo Meshes*, in "Computer Vision/Computer Graphics CollaborationTechniques 4th International Conference, MIRAGE 2009, France Rocquencourt", vol. 5495, 2009-05-04, http://hal.inria.fr/inria-00429338/en/.

[13] A. BONNEAU, J. BUSSET, B. WROBEL-DAUTCOURT. *Contextual effects on protrusion and lip opening for /i,y/*, in "10th Annual Conference of the International Speech Communication Association - Interspeech 2009, Royaume-Uni Brighton", ISCA, 2009-09, http://hal.inria.fr/inria-00433381/en/.

[14] J. DEQUIDT, C. DURIEZ, S. COTIN, E. KERRIEN. *Towards interactive planning of coil embolization in brain aneurysms*, in "Medical Image Computing and Computer Assisted Intervention - MICCAI 2009 12th International conference on Medical Image Computing and Computer-Assisted Intervention 8211; MICCAI 2009, part I, UK, London", G.-Z. YANG, D. HAWKES, D. RUECKERT, A. NOBLE, C. TAYLOR (editors), vol. 5761, Springer Berlin / Heidelberg, 2009-10-01, p. 377-385, http://hal.inria.fr/inria-00430867/en/.

[15] E. IMRE, M.-O. BERGER. *A 3-Component Inverse Depth Parameterization for Particle Filter SLAM*, in "DAGM Symposium, Pattern recognition, Germany, Jena", vol. 5748, 2009-09-09, p. 1–10, http://hal.inria.fr/inria-00429327/en/.

[16] E. IMRE, M.-O. BERGER, N. NOURY. *Improved Inverse Depth Parameterization for Monocular Simultaneous Localization and Mapping*, in "IEEE International Conference on Robotics and Automation, Japan, Kobe", 2009-05-12, http://hal.inria.fr/inria-00429318/en/.

[17] N. PADOY, D. MATEUS, D. WEINLAND, M.-O. BERGER, N. NAVAB. *Workflow Monitoring based on 3D Motion Features*, in "Workshop on Video-Oriented Object and Event Classification, ICCV Worshop, Japan, Kyoto", 2009-09-28, http://hal.inria.fr/inria-00429355/en/DE.

[18] TING. PENG, E. KERRIEN, M.-O. BERGER. *A shape base framework to segmentation of tongue contours from MRI data*, in "International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Dallas USA", 03 2010, http://hal.inria.fr/inria-00442138/en/.

[19] G. SIMON. *Immersive Image-Based Modeling of Polyhedral Scenes*, in "8th IEEE/ACM International Symposium on Mixed and Augmented Reality, USA, Orlando", 2009-10-20, http://hal.inria.fr/inria-00429847/en/.

National Peer-Reviewed Conference/Proceedings

[20] N. NOURY, F. SUR, M.-O. BERGER. *Modèle a contrario pour la mise en correspondance robuste sous contraintes épipolaires et photométriques*, in "RFIA, 17ième congrès francophone AFRIF-AFIA, Reconnaissance des Formes et Intelligence Artificielle, Caen France", 01 2010, http://hal.inria.fr/inria-00432992/en/.

Workshops without Proceedings

[21] M. ARON, M.-O. BERGER, E. KERRIEN. *Evaluation of the uncertainty of multimodal articulatory data*, in "Ultrafest V, USA, New Heaven", 2010-03-19, http://hal.inria.fr/inria-00429330/en/.

Scientific Books (or Scientific Book chapters)

[22] M. ARON, M.-O. BERGER, E. KERRIEN, Y. LAPRIE. *Acquisition multimodale de données articulatoires*, in "L'imagerie médicale pour l'étude de la parole", A. MARCHAL (editor), Hermes, 2009-11-02, http://hal.inria.fr/inria-00429585/en/.

[23] P.-F. VILLARD, W. BOURNE, F. BELLO. *Interactive Simulation of Diaphragm Motion Through Muscle and Rib Kinematics*, in "Recent Advances in the 3D Physiological Human", N. MAGNENAT-THALMANN, J. J. ZHANG, D. D. FENG (editors), Springer, 2009-10-30, p. 120–133, http://hal.archives-ouvertes.fr/hal-00430234/en/GB.

References in notes

[24] J. MONTIEL, J. CIVERA, A. DAVISON. *Unified Inverse Depth Parametrization for Monocular SLAM*, in "Robotics: Science and Systems", July 2006, http://pubs.doc.ic.ac.uk/inverse-depth-slam/.

[25] J. SIVIC, A. ZISSERMAN. *Video Google: A Text Retrieval Approach to Object Matching in Videos*, in "Proceedings of the International Conference on Computer Vision", vol. 2, October 2003, p. 1470–1477, http://www.robots.ox.ac.uk/~vgg.