IN PARTNERSHIP WITH:
Institut polytechnique de Grenoble

Activity Report 2018

Project-Team MORPHEO

Capture and Analysis of Shapes in Motion

IN COLLABORATION WITH: Laboratoire Jean Kuntzmann (LJK)

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Vision, perception and multimedia interpretation

Table of contents

Project-Team MORPHEO

Creation of the Team: 2011 March 01, updated into Project-Team: 2014 January 01

Keywords:

Computer Science and Digital Science:

A5.1.8. - 3D User Interfaces
A5.4. - Computer vision
A5.4.4. - 3D and spatio-temporal reconstruction
A5.4.5. - Object tracking and motion analysis
A5.5.1. - Geometrical modeling
A5.5.4. - Animation
A5.6. - Virtual reality, augmented reality
A6.2.8. - Computational geometry and meshes

Other Research Topics and Application Domains:

B2.6.3. - Biological Imaging
B2.8. - Sports, performance, motor skills
B9.2.2. - Cinema, Television
B9.2.3. - Video games
B9.4. - Sports

1. Team, Visitors, External Collaborators

Research Scientists

Edmond Boyer [Team leader, Inria, Senior Researcher, HDR]
Julien Pansiot [Inria, Starting Research Position]
Stefanie Wuhrer [Inria, Researcher, HDR]

Faculty Members

Jean Sébastien Franco [Institut polytechnique de Grenoble, Associate Professor]
Sergi Pujades Rocamora [Univ Grenoble Alpes, Associate Professor, from Dec 2018]

Post-Doctoral Fellow

Adnane Boukhayma [Inria, until Feb 2018]

PhD Students

Matthieu Armando [Inria]
Jean Basset [Inria, from Sep 2018]
Victoria Fernandez Abrevaya [Univ Grenoble Alpes]
Claude Goubet [Inria]
Roman Klokov [Inria]
Vincent Leroy [Inria]
Di Meng [Inria, from Oct 2018]
Abdullah Haroon Rasheed [Inria]
Romain Rombourg [Institut polytechnique de Grenoble]
Nitika Verma [Univ Grenoble Alpes]
Jinlong Yang [Inria, until Sep 2018]

Technical staff

Eymeric Amselem [Inria, from Jul 2018]

Laurence Boissieux [Inria]
Tomas Svaton [Inria, from Apr 2018]

Intern

Ivan Iudintsev [Inria, from Feb 2018 until Jun 2018]

Administrative Assistant

Nathalie Gillot [Inria]

External Collaborators

Franck Hétroy-Wheeler [Univ de Strasbourg, HDR]
Sergi Pujades Rocamora [Max Planck Institute for Perceiveing Systems, from Sep 2018 until Nov 2018]
Jinlong Yang [Facebook, from Oct 2018]

2. Overall Objectives

2.1. Overall Objectives

Figure 1. Dynamic Geometry Modeling

MORPHEO's ambition is to perceive and to interpret shapes that move using multiple camera systems. Departing from standard motion capture systems, based on markers, that provide only sparse information on moving shapes, multiple camera systems allow dense information on both shapes and their motion to be recovered from visual cues. Such ability to perceive shapes in motion brings a rich domain for research investigations on how to model, understand and animate real dynamic shapes, and finds applications, for instance, in gait analysis, bio-metric and bio-mechanical analysis, animation, games and, more insistently in recent years, in the virtual and augmented reality domain. The MORPHEO team particularly focuses on three different axes within the overall theme of 3D dynamic scene vision or 4D vision:

1. Shape and appearance models: how to build precise geometric and photometric models of shapes, including human bodies but not limited to, given temporal sequences.
2. Dynamic shape vision: how to register and track moving shapes, build pose spaces and animate captured shapes.
3. Inside shape vision: how to capture and model inside parts of moving shapes using combined color and X-ray imaging.

The strategy developed by MORPHEO to address the mentioned challenges is based on methodological tools that include in particular geometry, Bayesian inference and numerical optimization. Following the evolution in computer vision, our strategy has also evolved towards data driven approaches, as they have proved to be beneficial on different components of 3D vision solutions. Thus, our methodology include now machine learning tools whose potential in 4D vision are still to be fully investigated.

3. Research Program

3.1. Shape and Appearance Modeling

Standard acquisition platforms, including commercial solutions proposed by companies such as Microsoft, 3dMD or 4DViews, now give access to precise 3D models with geometry, e.g. meshes, and appearance information, e.g. textures. Still, state-of-the-art solutions are limited in many respects: They generally consider limited contexts and close setups with typically at most a few meter side lengths. As a result, many dynamic scenes, even a body running sequence, are still challenging situations; They also seldom exploit time redundancy; Additionally, data driven strategies are yet to be fully investigated in the field. The MORPHEO team builds on the Kinovis platform for data acquisition and has addressed these issues with, in particular, contributions on time integration, in order to increase the resolution for both for shapes and appearances, on representations, as well as on exploiting recent machine learning tools when modeling dynamic scenes. Our originality lies, for a large part, in the larger scale of the dynamic scenes we consider as well as in the time super resolution strategy we investigate. Another particularity of our research is a strong experimental foundation with the multiple camera Kinovis platforms.

3.2. Dynamic Shape Vision

Dynamic Shape Vision refers to research themes that consider the motion of dynamic shapes, with e.g. shapes in different poses, or the deformation between different shapes, with e.g. different human bodies. This includes for instance shape tracking, shape registration, all these themes being covered by MORPHEO. While progress has been made over the last decade in this domain, challenges remain, in particular due to the required essential task of shape correspondence that is still difficult to perform robustly. Strategies in this domain can be roughly classified into two categories: (i) data driven approaches that learn shape spaces and estimate shapes and their variations through space parameterizations; (ii) model based approaches that use more or less constrained prior models on shape evolutions, e.g. locally rigid structures, to recover correspondences. The MORPHEO team is substantially involved in the second category that leaves more flexibility for shapes that can be modeled, an important feature with the Kinovis platform. The team is anyway also considering the first category with faces and body under clothes modeling, classes of shapes that are more likely to evolve in spaces with reasonable dimensions. The originality of MORPHEO in this axis is to go beyond static shape poses and to consider also the dynamics of shape over several frames when modeling moving shapes, this in particular with shape tracking, animation and, more recently, with face registration.

3.3. Inside Shape Vision

Another research axis is concerned with the ability to perceive inside moving shapes. This is a more recent research theme in the MORPHEO team that has gained importance. It was originally the research associated to the Kinovis platform installed in the Grenoble Hospitals. This platform is equipped with two X-ray cameras and ten color cameras, enabling therefore simultaneous vision of inside and outside shapes. We believe this opens a new domain of investigation at the interface between computer vision and medical imaging. Interesting issues in this domain include the links between the outside surface of a shape and its inner parts, especially with the human body. These links are likely to help understanding and modeling human motions. Until now, numerous dynamic shape models, especially in the computer graphic domain, consist of a surface, typically a mesh, bound to a skeletal structure that is never observed in practice but that help anyway parameterizing human motion. Learning more accurate relationships using observations can therefore significantly impact the domain.

3.4. Shape Animation

3D animation is a crucial part of digital media production with numerous applications, in particular in the game and motion picture industry. Recent evolutions in computer animation consider real videos for both the creation and the animation of characters. The advantage of this strategy is twofold: it reduces the creation cost and increases realism by considering only real data. Furthermore, it allows to create new motions, for real characters, by recombining recorded elementary movements. In addition to enable new media contents to be produced, it also allows to automatically extend moving shape datasets with fully controllable new motions. This ability appears to be of great importance with the recent advent of deep learning techniques and the associated need for large learning datasets. In this research direction, we investigate how to create new dynamic scenes using recorded events.

4. Application Domains

4.1. 4D modeling

Modeling shapes that evolve over time, analyzing and interpreting their motion has been a subject of increasing interest of many research communities including the computer vision, the computer graphics and the medical imaging communities. Recent evolutions in acquisition technologies including 3D depth cameras (Time-of-Flight and Kinect), multi-camera systems, marker based motion capture systems, ultrasound and CT scanners have made those communities consider capturing the real scene and their dynamics, create 4D spatio-temporal models, analyze and interpret them. A number of applications including dense motion capture, dynamic shape modeling and animation, temporally consistent 3D reconstruction, motion analysis and interpretation have therefore emerged.

4.2. Shape Analysis

Most existing shape analysis tools are local, in the sense that they give local insight about an object's geometry or purpose. The use of both geometry and motion cues makes it possible to recover more global information, in order to get extensive knowledge about a shape. For instance, motion can help to decompose a 3D model of a character into semantically significant parts, such as legs, arms, torso and head. Possible applications of such high-level shape understanding include accurate feature computation, comparison between models to detect defects or medical pathologies, and the design of new biometric models.

4.3. Human Motion Analysis

The recovery of dense motion information enables the combined analysis of shapes and their motions. Typical examples include the estimation of mean shapes given a set of 3D models or the identification of abnormal deformations of a shape given its typical evolutions. The interest arises in several application domains where temporal surface deformations need to be captured and analyzed. It includes human body analyses for which potential applications are anyway numerous and important, from the identification of pathologies to the design of new prostheses.

4.4. Virtual and Augmented Reality

This domain has actually seen new devices emerged that enable now full 3D visualization, for instance the HTC Vive, the Microsoft Hololens and the Magic leap one. These devices create a need for adapted animated 3D contents that can either be generated or captured. We believe that captured 4D models will gain interest in this context since they provide realistic visual information on moving shapes that tend to avoid negative perception effects such as the uncanny valley effect. Besides 3D visualization devices, many recent applications also rely on everyday devices, such as mobile phones, to display augmented reality contents with free viewpoint ability. In this case, 3D and 4D contents are also expected.

5. Highlights of the Year

5.1. Highlights of the Year

MORPHEO created holograms for an augmented reality application developed for the clothing retailer Zara. This application enables the brand's customers to enjoy a virtual and interactive shopping experience via their smartphones in one of the 120 stores across the world taking part in this experiment. Last January, all of the holograms presented in the Zara AR application were captured using the Kinovis 4D platform. The challenge with regard to the acquisition of the 12 sequences created was to accurately reproduce the models in sweeping movements and with complex clothing effects due to the materials and styles chosen.

6. New Software and Platforms

6.1. 4D repository

KEYWORDS: 4D - Dynamic scene
FUNCTIONAL DESCRIPTION: This website hosts dynamic mesh sequences reconstructed from images captured using a multi-camera set up. Such mesh-sequences offer a new promising vision of virtual reality, by capturing real actors and their interactions. The texture information is trivially mapped to the reconstructed geometry, by back-projecting from the images. These sequences can be seen from arbitrary viewing angles as the user navigates in 4D (3D geometry + time) . Different sequences of human / non-human interaction can be browsed and downloaded from the data section.

- Contact: Edmond Boyer
- URL: http://4drepository.inrialpes.fr/

6.2. Lucy Viewer

KEYWORDS: Data visualization - 4D - Multi-Cameras
SCIENTIFIC DESCRIPTION: Lucy Viewer is an interactive viewing software for 4D models, i.e, dynamic three-dimensional scenes that evolve over time. Each 4D model is a sequence of meshes with associated texture information, in terms of images captured from multiple cameras at each frame. Such data is available from the 4D repository website hosted by Inria Grenoble.

With Lucy Viewer, the user can use the mouse to zoom in onto the 4D models, zoom out, rotate, translate and view from an arbitrary angle as the 4D sequence is being played. The texture information is read from the images at each frame in the sequence and applied onto the meshes. This helps the user visualize the 3D scene in a realistic manner. The user can also freeze the motion at a particular frame and inspect a mesh in detail. Lucy Viewer lets the user to also select a subset of cameras from which to apply texture information onto the meshes. The supported formats are meshes in .OFF format and associated images in .PNG or .JPG format.
FUNCTIONAL DESCRIPTION: Lucy Viewer is an interactive viewing software for 4D models, i.e, dynamic three-dimensional scenes that evolve over time. Each 4D model is a sequence of meshes with associated texture information, in terms of images captured from multiple cameras at each frame.

- Participants: Edmond Boyer, Jean-Sébastien Franco, Matthieu Armando and EYMERIC AMSELEM
- Contact: Edmond Boyer
- URL: https://kinovis.inria.fr/lucyviewer/

6.3. Shape Tracking

FUNCTIONAL DESCRIPTION: We are developing a software suite to track shapes over temporal sequences. The motivation is to provide temporally coherent 4D Models, i.e. 3D models and their evolutions over time , as required by motion related applications such as motion analysis. This software takes as input a temporal sequence of 3D models in addition to a template and estimate the template deformations over the sequence that fit the observed 3D models.

- Contact: Edmond Boyer

6.4. QuickCSG V2

KEYWORDS: 3D modeling - CAD - 3D reconstruction - Geometric algorithms
SCIENTIFIC DESCRIPTION: See the technical report "QuickCSG: Arbitrary and Faster Boolean Combinations of N Solids", Douze, Franco, Raffin.

The extension of the algorithm to self-intersecting meshes is described in "QuickCSG with self-intersections", a document inside the package.
FUNCTIONAL DESCRIPTION: QuickCSG is a library and command-line application that computes Boolean operations between polyhedra. The basic algorithm is described in the research report "QuickCSG: Arbitrary and Faster Boolean Combinations of N Solids", Douze, Franco, Raffin. The input and output polyhedra are defined as indexed meshes. In version 2, that was developed in the context of a software transfer contract, the meshes can be self-intersecting, in which case the inside and outside are defined by the non-zero winding rule. The operation can be any arbitrary Boolean function, including one that is defined as a CSG tree. The focus of QuickCSG is speed. Robustness to degeneracies is obtained by carefully applied random perturbations.

- Authors: Matthys Douze, Jean-Sébastien Franco and Bruno Raffin
- Contact: Jean-Sébastien Franco
- URL: https://kinovis.inria.fr/quickcsg/

6.5. CVTGenerator

KEYWORDS: Mesh - Centroidal Voronoi tessellation - Implicit surface
FUNCTIONAL DESCRIPTION: CVTGenerator is a program to build Centroidal Voronoi Tessellations of any 3D meshes and implicit surfaces.

- Partner: INP Grenoble
- Contact: Li Wang
- URL: http://cvt.gforge.inria.fr/

6.6. Kinovis Platform

Kinovis (http://kinovis.inrialpes.fr/) is a multi-camera acquisition project that was selected within the call for proposals ”Equipements d’Excellence” of the program "Investissement d’Avenir" funded by the French government. The project involves 2 institutes: the Inria Grenoble Rhône-Alpes, the Université Joseph Fourier and 4 laboratories: the LJK (laboratoire Jean Kuntzmann - applied mathematics), the LIG (Laboratoire d’informatique de Grenoble - Computer Science), the Gipsa lab (Signal, Speech and Image processing) and the LADAF (Grenoble Hospitals - Anatomy). The Kinovis environment is composed of 2 complementary platforms (see Figure 2). A first platform located at Inria Grenoble with a 10mx10m acquisition surface is equipped with 68 color cameras and 20 IR motion capture (mocap) cameras. It is the evolution of the Grimage platform towards the production of better models of more complex dynamic scenes. A second platform located at Grenoble Hospitals (CHU), within the LADAF anatomy laboratory, is equipped with 10 color and 2 X-ray cameras to enable combined analysis of internal and external shape structures, such as skeletons and their surrounding bodies. Both platforms have already demonstrated their potential through a range of projects lead by the team and externally. Members of Morpheo are highly involved in this project. Edmond Boyer is coordinating this project, and Julien Pansiot is managing the technical resources of both platforms.

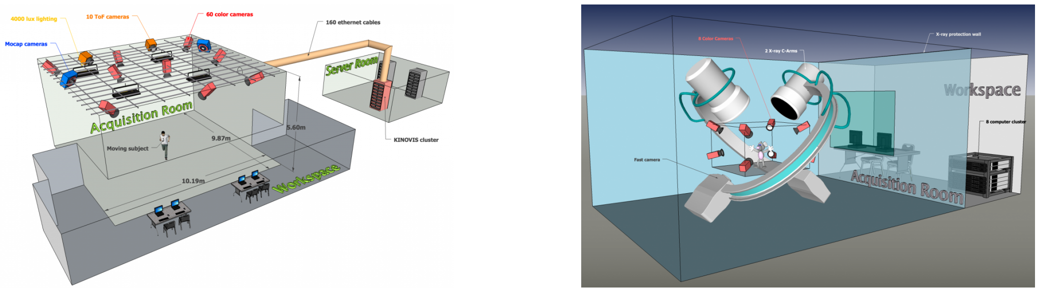

Figure 2. Kinovis platforms: on the left the Inria platform; on the right Grenoble Hospital platform.

7. New Results

7.1. Surface Motion Capture Animation Synthesis

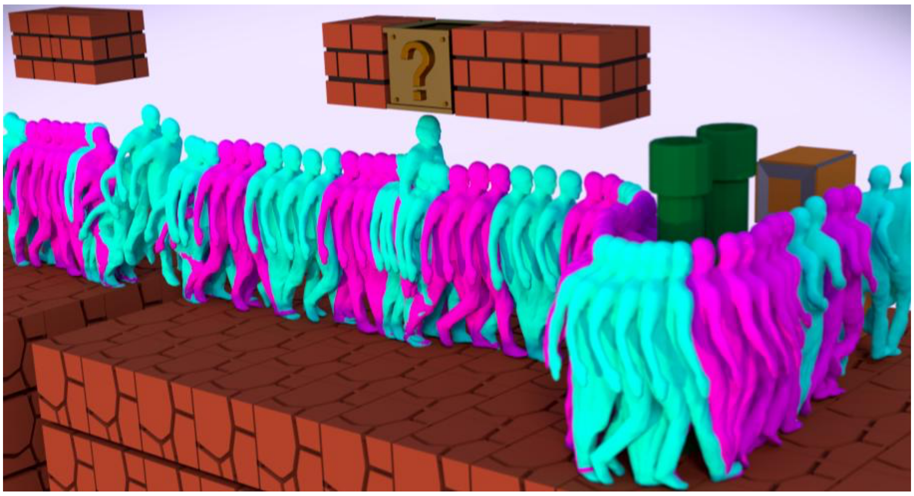

Figure 3.

We propose to generate novel animations from a set of elementary examples of video-based surface motion capture, under user-specified constraints. 4D surface capture animation is motivated by the increasing demand from media production for highly realistic 3D content. To this aim, data driven strategies that consider video-based information can produce animation with real shapes, kinematics and appearances. Our animations rely on the combination and the interpolation of textured 3D mesh data, which requires examining two aspects: (1) Shape geometry and (2) appearance. First, we propose an animation synthesis structure for the shape geometry, the Essential graph, that outperforms standard Motion graphs in optimality with respect to quantitative criteria, and we extend optimized interpolated transition algorithms to mesh data. Second, we propose a compact view-independent representation for the shape appearance. This representation encodes subject appearance changes due to viewpoint and illumination, and due to inaccuracies in geometric modelling independently. Besides providing compact representations, such decompositions allow for additional applications such as interpolation for animation (see figure 3).

This result was published in a prominent computer graphics journal, IEEE Transactions on Visualization and Computer Graphics [2].

7.2. A Multilinear Tongue Model Derived from Speech Related MRI Data of the Human Vocal Tract

We present a multilinear statistical model of the human tongue that captures anatomical and tongue pose related shape variations separately. The model is derived from 3D magnetic resonance imaging data of 11 speakers sustaining speech related vocal tract configurations. To extract model parameters, we use a minimally supervised method based on an image segmentation approach and a template fitting technique. Furthermore, we use image denoising to deal with possibly corrupt data, palate surface information reconstruction to handle palatal tongue contacts, and a bootstrap strategy to refine the obtained shapes. Our evaluation shows that, by limiting the degrees of freedom for the anatomical and speech related variations, to 5 and 4, respectively, we obtain a model that can reliably register unknown data while avoiding overfitting effects. Furthermore, we show that it can be used to generate plausible tongue animation by tracking sparse motion capture data.

This result was published in Computer Speech and Language 51 [3].

7.3. CBCT of a Moving Sample from X-rays and Multiple Videos

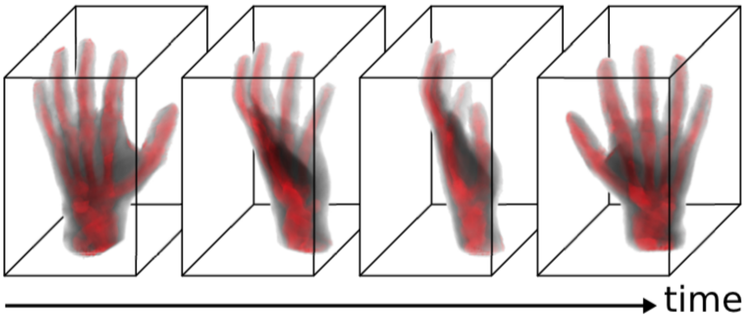

Figure 4. Dense volumetric attenuation reconstruction from a rigidly moving sample captured by a single planar X-ray imaging device and a surface motion capture system. Higher attenuation (here bone structure) is highlighted in red.

We consider dense volumetric modeling of moving samples such as body parts. Most dense modeling methods consider samples observed with a moving X-ray device and cannot easily handle moving samples. We propose instead a novel method to observe shape motion from a fixed X-ray device and to build dense in-depth attenuation information. This yields a low-cost, low-dose 3D imaging solution, taking benefit of equipment widely available in clinical environments. Our first innovation is to combine a video-based surface motion capture system with a single low-cost/low-dose fixed planar X-ray device, in order to retrieve the sample motion and attenuation information with minimal radiation exposure. Our second innovation is to rely on Bayesian inference to solve for a dense attenuation volume given planar radioscopic images of a moving sample. This approach enables multiple sources of noise to be considered and takes advantage of very limited prior information to solve an otherwise ill-posed problem. Results show that the proposed strategy is able to reconstruct dense volumetric attenuation models from a very limited number of radiographic views over time on synthetic and in-situ data, as illustrated in Figure 4.

This result was published in a prominent medical journal, IEEE Transactions on Medical Imaging [4].

7.4. Automatic camera calibration using multiple sets of pairwise correspondences

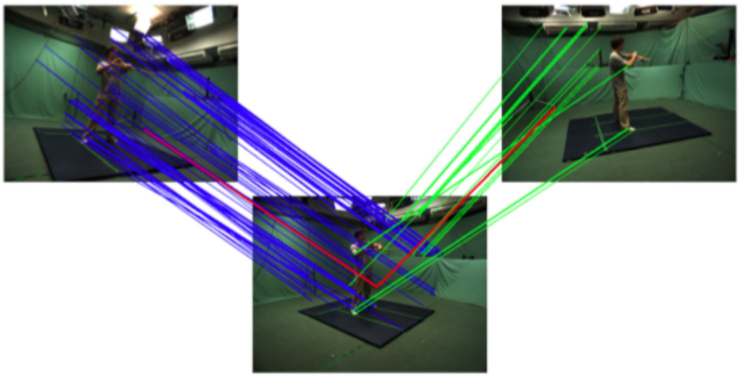

Figure 5. Correspondences extracted from SIFT features. Given the wide baseline between the views there is a single reliable triple correspondence (red) while there are many reliable pairwise correspondences (blue and green).

We propose a new method to add an uncalibrated node into a network of calibrated cameras using only pairwise point correspondences (see figure 5). While previous methods perform this task using triple correspondences, these are often difficult to establish when there is limited overlap between different views. In such challenging cases we must rely on pairwise correspondences and our solution becomes more advantageous. Our method includes an 11-point minimal solution for the intrinsic and extrinsic calibration of a camera from pairwise correspondences with other two calibrated cameras, and a new inlier selection framework that extends the traditional RANSAC family of algorithms to sampling across multiple datasets. Our method is validated on different application scenarios where a lack of triple correspondences might occur: addition of a new node to a camera network; calibration and motion estimation of a moving camera inside a camera network; and addition of views with limited overlap to a Structure-from-Motion model.

This result was published in a prominent medical journal, IEEE Transactions on Pattern Analysis and Machine Intelligence [5].

7.5. Multilinear Autoencoder for 3D Face Model Learning

Generative models have proved to be useful tools to represent 3D human faces and their statistical variations (see figure 6). With the increase of 3D scan databases available for training, a growing challenge lies in the ability to learn generative face models that effectively encode shape variations with respect to desired attributes, such as identity and expression, given datasets that can be diverse. This paper addresses this challenge by proposing a framework that learns a generative 3D face model using an autoencoder architecture, allowing hence for weakly supervised training. The main contribution is to combine a convolutional neural network-based en-coder with a multilinear model-based decoder, taking therefore advantage of both the convolutional network robustness to corrupted and incomplete data, and of the multilinear model capacity to effectively model and decouple shape variations. Given a set of 3D face scans with annotation labels for the desired attributes, e.g. identities and expressions, our method learns an expressive multilinear model that decouples shape changes due to the different factors. Experimental results demonstrate that the proposed

Figure 6. Shape variations caused by different expressions of the same subject.

method outperforms recent approaches when learning multilinear face models from incomplete training data, particularly in terms of space decoupling, and that it is capable of learning from an order of magnitude more data than previous methods.

This result was published in IEEE Winter Conference on Applications of Computer Vision [6].

7.6. Spatiotemporal Modeling for Efficient Registration of Dynamic 3D Faces

We consider the registration of temporal sequences of 3D face scans. Face registration plays a central role in face analysis applications, for instance recognition or transfer tasks, among others. We propose an automatic approach that can register large sets of dynamic face scans without the need for landmarks or highly specialized acquisition setups. This allows for extended versatility among registered face shapes and deformations by enabling to leverage multiple datasets, a fundamental property when e.g. building statistical face models. Our approach is built upon a regression-based static registration method, which is improved by spatiotemporal modeling to exploit redundancies over both space and time. We experimentally demonstrate that accurate registrations can be obtained for varying data robustly and efficiently by applying our method to three standard dynamic face datasets.

This work has been published in 3D Vision 2018 [7].

7.7. Shape Reconstruction Using Volume Sweeping and Learned Photoconsistency

The rise of virtual and augmented reality fuels an increased need for contents suitable to these new technologies including 3D contents obtained from real scenes (see figure 7). We consider in this paper the problem of 3D shape reconstruction from multi-view RGB images. We investigate the ability of learning-based strategies to effectively benefit the reconstruction of arbitrary shapes with improved precision and robustness. We especially target real life performance capture, containing complex surface details that are difficult to recover with existing approaches. A key step in the multi-view reconstruction pipeline lies in the search for matching features between viewpoints in order to infer depth information. We propose to cast the matching on a 3D receptive field along viewing lines and to learn a multi-view photoconsistency measure for that purpose. The intuition is that deep networks have the ability to learn local photometric configurations in a broad way, even with respect to different orientations along various viewing lines of the same surface point. Our results demonstrate this ability, showing that a CNN, trained on a standard static dataset, can help recover surface details on dynamic scenes that are not perceived by traditional 2D feature based methods. Our evaluation also shows that our solution compares on par to state of the art reconstruction pipelines on standard evaluation datasets, while yielding significantly better results and generalization with realistic performance capture data.

This work has been published in the European Conference on Computer Vision 2018 [9] and Reconnaissance des Formes, Image, Apprentissage et Perception 2018 [8].

7.8. FeaStNet: Feature-Steered Graph Convolutions for 3D Shape Analysis

Figure 7. Challenging scene captured with Kinovis. (left) one input image, (center) reconstructions obtained with our previous work based on classical 2D features, (right) proposed solution. Our results validate the key improvement of a CNN-learned disparity to MVS for performance capture scenarios. Results particularly improve in noisy, very low contrast and low textured regions such as the arm, the leg or even the black skirt folds.

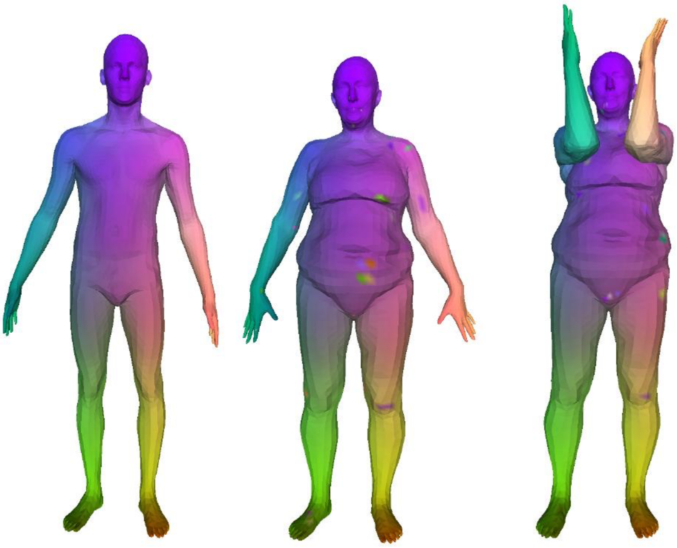

Figure 8. Two examples of texture transfer from a reference shape in neutral pose (left) using shape correspondences predicted by FeaStNet (multi-scale architecture, without refinement).

Convolutional neural networks (CNNs) have massively impacted visual recognition in 2D images, and are now ubiquitous in state-of-the-art approaches. CNNs do not easily extend, however, to data that are not represented by regular grids, such as 3D shape meshes or other graph-structured data, to which traditional local convolution operators do not directly apply. To address this problem, we propose a novel graph-convolution operator to establish correspondences between filter weights and graph neighborhoods with arbitrary connectivity. The key novelty of our approach is that these correspondences are dynamically computed from features learned by the network, rather than relying on predefined static coordinates over the graph as in previous work. We obtain excellent experimental results that significantly improve over previous state-of-the-art shape correspondence results (see figure 8). This shows that our approach can learn effective shape representations from raw input coordinates, without relying on shape descriptors.

This work has been published in the IEEE Conference on Computer Vision and Pattern Recognition 2018 [11].

7.9. Analyzing Clothing Layer Deformation Statistics of 3D Human Motions

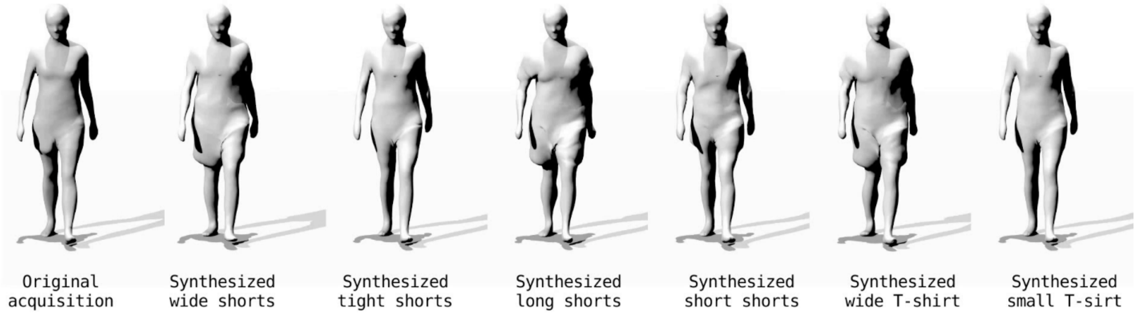

Figure 9. Examples of clothing re-synthesis based on our clothing layer regression model.

Recent capture technologies and methods allow not only to retrieve 3D model sequence of moving people in clothing, but also to separate and extract the underlying body geometry and motion component and separate the clothing as a geometric layer. So far this clothing layer has only been used as raw offsets for individual applications such as retargeting a different body capture sequence with the clothing layer of another sequence, with limited scope, e.g. using identical or similar motions. The structured, semantics and motion-correlated nature of the information contained in this layer has yet to be fully understood and exploited. To this purpose we propose a comprehensive analysis of the statistics of this layer with a simple two-component model, based on PCA subspace reduction of the layer information on one hand, and a generic parameter regression model using neural networks on the other hand, designed to regress from any semantic parameter whose variation is observed in a training set, to the layer parameterization space. We show that this model not only allows to reproduce previous motion retargeting works, but generalizes the data generation capabilities of the method to other semantic parameters such as clothing variation and size (see figure 9), or physical material parameters with synthetically generated training sequence, paving the way for many kinds of capture data-driven creation and augmentation applications.

This work has been published in the European Conference on Computer Vision 2018 [12].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

A collaboration with the French Start up Holooh started in 2017 and was pursued in 2018. Holooh aims at producing high quality holograms for VR and AR applications, especially for the fashion and music domains. Holooh's objective is to set up a multi-camera studio in Paris for that purpose. Edmond Boyer is involved in the collaboration.

8.2. Bilateral Grants with Industry

The Morpheo Inria team and Microsoft research set up a collaboration on the capture and modelling of moving shapes using multiple videos. Two PhD proposals will be part of this collaboration with the objective to make contributions on 4D Modeling. The PhDs will take place at Inria Grenoble Rhône-Alpes and will involve regular visits and stays at Microsoft in Redmond (USA) and Cambridge (UK). At Microsoft, Steve Sullivan, Andrew Fitzgibbon and Marta Wilczkowiak will be participating to the project.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

9.1.1.1. ANR PRCE CaMoPi – Capture and Modelling of the Shod Foot in Motion

The main objective of the CaMoPi project is to capture and model dynamic aspects of the human foot with and without shoes. To this purpose, video and X-ray imagery will be combined to generate novel types of data from which major breakthroughs in foot motion modelling are expected. Given the complexity of the internal foot structure, little is known about the exact motion of its inner structure and the relationship with the shoe. Hence the current state-of-the art shoe conception process still relies largely on ad-hoc know-how. This project aims at better understanding the inner mechanisms of the shod foot in motion in order to rationalise and therefore speed up and improve shoe design in terms of comfort, performance, and cost. This requires the development of capture technologies that do not yet exist in order to provide full dense models of the foot in motion. To reach its goals, the CaMoPi consortium comprises complementary expertise from academic partners : Inria (combined video and X-ray capture and modeling) and Mines St Etienne (finite element modeling), as well as industrial : CTC Lyon (shoe conception and manufacturing, dissemination). The project has effectively started in October 2017 with Claude Goubet's recruitment as a PhD candidate followed by Tomas Svaton as an engineer in April 2018.

9.1.1.2. ANR project Achmov – Accurate Human Modeling in Videos

The technological advancements made over the past decade now allow the acquisition of vast amounts of visual information through the use of image capturing devices like digital cameras or camcorders. A central subject of interest in video are the humans, their motions, actions or expressions, the way they collaborate and communicate. The goal of ACHMOV is to extract detailed representations of multiple interacting humans in real-world environments in an integrated fashion through a synergy between detection, figure-ground segmentation and body part labeling, accurate 3D geometric methods for kinematic and shape modeling, and large-scale statistical learning techniques. By integrating the complementary expertise of two teams (one French, MORPHEO and one Romanian, CLVP), with solid prior track records in the field, there are considerable opportunities to move towards processing complex real world scenes of multiple interacting people, and be able to extract rich semantic representations with high fidelity. This would enable interpretation, recognition and synthesis at unprecedented levels of accuracy and in considerably more realistic setups than currently considered. This project has funded the work of two soon to defend PhD students Vincent Leroy and Jinlong Yang, and ended during the year 2018.

9.1.2. Competitivity Clusters

9.1.2.1. FUI project Creamove

Creamove is a collaboration between the Morpheo team of the Inria Grenoble Rhône-Alpes, the 4D View Solution company specialized in multi-camera acquisition systems, the SIP company specialized in multimedia and interactive applications and a choreographer. The objective is to develop new interactive and artistic applications where humans can interact in 3D with virtual characters built from real videos. Dancer performances will be pre-recorded in 3D and used on-line to design new movement sequences based on inputs coming from human bodies captured in real time. Website: http://www.creamove.fr.

9.1.2.2. FUI24 SPINE-PDCA

The goal of the SPINE-PDCA project is to develop a unique medical platform that will streamline the medical procedure and achieve all the steps of a minimally invasive surgery intervention with great precision through a complete integration of two complementary systems for pre-operative planning (EOS platform from EOS IMAGING) and imaging/intra-operative navigation (SGV3D system from SURGIVISIO). Innovative low-dose tracking and reconstruction algorithms will be developed by Inria, and collaboration with two hospitals (APHP Trousseau and CHU Grenoble) will ensure clinical feasibility. The medical need is particularly strong in the field of spinal deformity surgery which can, in case of incorrect positioning of the implants, result in serious musculoskeletal injury, a high repeat rate (10 to 40% of implants are poorly positioned in spine surgery) and important care costs. In paediatric surgery (e. g. idiopathic scoliosis), the rate of exposure to X-rays is an additional major consideration in choosing the surgical approach to engage. For these interventions, advanced linkage between planning, navigation and postoperative verification is essential to ensure accurate patient assessment, appropriate surgical procedure and outcome consistent with clinical objectives. The project has effectively started in October 2018 with Di Meng's recruitment as a PhD candidate.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Reviewer

- Jean-Sébastien Franco reviewed for 3DV, CVPR and ECCV 2018, and the CVPR 2018 Humans Workshop.
- Sergi Pujades reviewed for 3DV.
- Julien Pansiot reviewed for ECCV 2018 and CVPR 2019.
- Stefanie Wuhrer reviewed for CVPR, ECCV, and SIGGRAPH.
- Edmond Boyer reviewed for 3DV, CVPR and RFIAP 2018.
- Edmond Boyer was area chair for BMVC and ECCV 2018.

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- Edmond Boyer is associate editor of the International Journal of Computer Vision (Springer).

10.1.2.2. Reviewer - Reviewing Activities

- Stefanie Wuhrer reviewed for IJCV and PAMI.
- Edmond Boyer reviewed for IJCV and PAMI.
- Sergi Pujades reviewed for the Journal of Imaging (MDPI), Sensors and IEEE TVCG.
- Jean-Sébastien Franco reviewed for IEEE Computer Graphics and Applications.

10.1.3. Invited Talks

- Edmond Boyer gave invited talks at: Microsoft Cambridge (visit), Institut Descarte Paris (conférence maths et mouvement), New York University (visit) and Montpellier University (Module Image).

10.1.4. Scientific Expertise

- Jean-Sebastien Franco reviewed for the Euregio Science Fund (EGTC) - Interregional Project Networks (IPN) in 2018.
- Jean-Sebastien Franco was a member of the recruiting commitee of Universite Grenoble Alpes for an Assistant Professor position in 2018.
- Jean-Sebastien Franco was a member of the Ensimag Engineering school - Grenoble INP steering commitee (Conseil d'École) in 2018.
- Jean-Sebastien Franco was a member of the recruiting commitee of Ensimag - Grenoble INP Engineering school for temporary research and teaching associates (ATER) in 2018.
- Sergi Pujades reviewed for the DFG (German ANR).

10.1.5. Research Administration

- Edmond Boyer is auditor for the Computer Vision European Association.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence: Jean-Sébastien Franco, Introduction to Imperative Programming, 57h, Ensimag 1st year, Grenoble INP.

Licence: Jean-Sébastien Franco, C Programming project, 27h, Ensimag 3rd year, Grenoble INP.

Master: Jean-Sébastien Franco, Supervision of the 2nd year program (300 students), 36h, Ensimag 2nd year, Grenoble INP.

Master: Jean-Sébastien Franco, Introduction to Computer Graphics, 44h, Ensimag 2nd year, Grenoble INP.

Master: Jean-Sébastien Franco, Introduction to Computer Vision, 27h, Ensimag 3rd year, Grenoble INP.

Master: Jean-Sébastien Franco, End of study project tutoring (PFE), 12h, Ensimag 3rd year, Grenoble INP.

Master: Edmond Boyer, 3D Modeling, 18h, M2R Mosig GVR, Grenoble INP.

Master: Edmond Boyer, Introduction to Visual Computing, 30h, M1 MoSig, Université Grenoble Alpes.

Master: Stefanie Wuhrer, 3D Graphics, 13.5h, M1 MoSig and MSIAM, Université Grenoble Alpes.

10.2.2. Supervision

PhD in progress: Victoria Fernandez Abrevaya, 3D Dynamic Human Motion Representations, Université Grenoble Alpes (France), started 01/10/2016, supervised by Edmond Boyer and Stefanie Wuhrer.

PhD in progress: Jean Basset, Learning Morphologically Plausible Pose Transfer, Université Grenoble Alpes (France), started 01/10/2018, supervised by Edmond Boyer, Franck Multon and Stefanie Wuhrer.

PhD in progress: Abdullah Haroon Rasheed, Cloth Modeling and Simulation, Université Grenoble Alpes (France), started 01/11/2017, supervised by Florence Bertails-Descoubes, Jean-Sébastien Franco and Stefanie Wuhrer.

PhD in progress: Jinlong Yang, Learning shape spaces of dressed 3D human models in motion, Université Grenoble Alpes (France), started 01/10/2015, supervised by Franck Hétroy-Wheeler and Stefanie Wuhrer.

HdR : Stefanie Wuhrer, Deformation Models for Human Shape Analysis, Université Grenoble Alpes (France), September 2018

PhD in progress: Claude Goubet, Dense 4D Modelling of Moving Shapes, Université Grenoble Alpes (France), started 02/10/2017, supervised by Edmond Boyer and Julien Pansiot

PhD in progress: Di Meng, Deep learning for low-dose CBCT reconstruction and registration, Université Grenoble Alpes (France), started 01/10/2018, supervised by Edmond Boyer and Julien Pansiot.

PhD in progress: Matthieu Armando, Temporal Integration for Shape and Appearance Modeling, Université Grenoble Alpes (France), started 01/01/2018, supervised by Edmond Boyer and Jean-Sébastien Franco.

PhD in progress: Vincent Leroy, 4D shape reconstruction from photoconsistency cues, Université Grenoble Alpes, started 01/10/2015, supervised by Edmond Boyer and Jean-Sébastien Franco.

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] S. WUHRER. *Deformation Models for Human Shape Analysis*, Université Grenoble Alpes (France), September 2018, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01894716

Articles in International Peer-Reviewed Journals

[2] A. BOUKHAYMA, E. BOYER. *Surface Motion Capture Animation Synthesis*, in "IEEE Transactions on Visualization and Computer Graphics", 2018, pp. 1-14 [*DOI :* 10.1109/TVCG.2018.2831233], https://hal.inria.fr/hal-01781164

[3] A. HEWER, S. WUHRER, I. STEINER, K. RICHMOND. *A Multilinear Tongue Model Derived from Speech Related MRI Data of the Human Vocal Tract*, in "Computer Speech and Language", September 2018, vol. 51, pp. 68-92, https://arxiv.org/abs/1612.05005 [*DOI :* 10.1016/J.CSL.2018.02.001], https://hal.archives-ouvertes.fr/hal-01418460

[4] J. PANSIOT, E. BOYER. *CBCT of a Moving Sample from X-rays and Multiple Videos*, in "IEEE Transactions on Medical Imaging", August 2018, pp. 1-11 [*DOI :* 10.1109/TMI.2018.2865228], https://hal.inria.fr/hal-01857487

[5] F. VASCONCELOS, J. P. BARRETO, E. BOYER. *Automatic camera calibration using multiple sets of pairwise correspondences*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", April 2018, vol. 40, n^{o} 4, pp. 791-803 [*DOI :* 10.1109/TPAMI.2017.2699648], https://hal.inria.fr/hal-01675686

International Conferences with Proceedings

[6] V. FERNÁNDEZ ABREVAYA, S. WUHRER, E. BOYER. *Multilinear Autoencoder for 3D Face Model Learning*, in "WACV 2018 - IEEE Winter Conference on Applications of Computer Vision", Lake Tahoe, NV/CA, United States, IEEE, March 2018, pp. 1-9 [*DOI :* 10.1109/WACV.2018.00007], https://hal.archives-ouvertes.fr/hal-01700934

[7] V. FERNÁNDEZ ABREVAYA, S. WUHRER, E. BOYER. *Spatiotemporal Modeling for Efficient Registration of Dynamic 3D Faces*, in "3DV 2018 - 6th International Conference on 3D Vision", Verona, Italy, September 2018, pp. 1-10, https://hal.inria.fr/hal-01855955

[8] V. LEROY, J.-S. FRANCO, E. BOYER. *Apprentissage de la Cohérence Photométrique pour la Reconstruction de Formes Multi-Vues*, in "RFIAP 2018 - Reconnaissance des Formes, Image, Apprentissage et Perception", Marne la Vallée, France, June 2018, https://hal.archives-ouvertes.fr/hal-01857627

[9] V. LEROY, J.-S. FRANCO, E. BOYER. *Shape Reconstruction Using Volume Sweeping and Learned Photoconsistency*, in "European Conference on Computer Vision", Munich, Germany, September 2018, https://hal.archives-ouvertes.fr/hal-01849286

[10] A.-H. RASHEED, V. ROMERO, F. BERTAILS-DESCOUBES, A. LAZARUS, S. WUHRER, J.-S. FRANCO. *Estimating friction in cloth, using simulation and machine learning*, in "American Physical Society March Meeting 2019", Boston, United States, March 2019, https://hal.inria.fr/hal-01982257

[11] N. VERMA, E. BOYER, J. VERBEEK. *FeaStNet: Feature-Steered Graph Convolutions for 3D Shape Analysis*, in "CVPR - IEEE Conference on Computer Vision & Pattern Recognition", Salt Lake City, United States, June 2018, pp. 1-9, https://arxiv.org/abs/1706.05206 , https://hal.inria.fr/hal-01540389

[12] J. YANG, J.-S. FRANCO, F. HÉTROY-WHEELER, S. WUHRER. *Analyzing Clothing Layer Deformation Statistics of 3D Human Motions*, in "ECCV 2018 - European Conference on Computer Vision", Munich, Germany, September 2018, pp. 1-17, https://hal.inria.fr/hal-01763706