IN PARTNERSHIP WITH:
Institut polytechnique de Grenoble

Activity Report 2019

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, until Mar 2019]
Stefanie Wuhrer [Inria, Researcher, HDR]

Faculty Members

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

PhD Students

Matthieu Armando [Inria, PhD Student]
Jean Basset [Inria, PhD Student]
Sanae Dariouche [Univ Grenoble Alpes, PhD Student, from Nov 2019]
Victoria Fernandez Abrevaya [Univ Grenoble Alpes, PhD Student]
Claude Goubet [Inria, PhD Student, until Sep 2019]
Roman Klokov [Inria, PhD Student]
Vincent Leroy [Inria, PhD Student, until Mar 2019]
Mathieu Marsot [Inria, PhD Student, from Nov 2019]
Di Meng [Inria, PhD Student]
Abdullah Haroon Rasheed [Inria, PhD Student]
Nitika Verma [Univ Grenoble Alpes, PhD Student]
Boyao Zhou [Inria, PhD Student]
Pierre Zins [Inria, PhD Student, from Oct 2019]
Jinlong Yang [Inria, PhD Student, until Mar 2019]

Technical staff

Eymeric Amselem [Inria, Engineer, until Jun 2019]
Laurence Boissieux [Inria, Engineer]
Julien Pansiot [Inria, Engineer, from Oct 2019]
Tomas Svaton [Inria, Engineer]

Interns and Apprentices

Jiabin Chen [Inria, from Mar 2019 until Aug 2019]
Robin Magnet [Inria, until Jul 2019]

Administrative Assistant

Nathalie Gillot [Inria, Administrative Assistant]

Visiting Scientist

Stephane Durocher [Université du Manitoba (Canada), from Jul 2019]

2. Overall Objectives

2.1. Overall Objectives

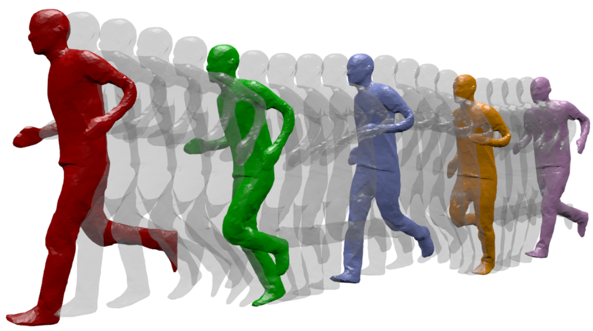

Figure 1. Dynamic Geometry Modeling

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 is 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 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, 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 emerge 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

5.1.1. Awards

BEST PAPER AWARD:

[15]
R. KLOKOV, J. VERBEEK, E. BOYER. *Probabilistic Reconstruction Networks for 3D Shape Inference from a Single Image*, in "BMVC 2019 - British Machine Vision Conference", Cardiff, United Kingdom, September 2019, pp. 1-15, https://arxiv.org/abs/1908.07475 - Awarded with Best Science Paper Honourable Mention Award at BMVC'19., https://hal.inria.fr/hal-02268466

6. New Software and Platforms

6.1. 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/

6.2. 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

6.3. 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/

6.4. 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/

6.5. Adaptive mesh texture

KEYWORDS: 3D - Geometry Processing - Texturing

FUNCTIONAL DESCRIPTION: Tool for computing appearance information on a 3D scene acquired with a multi-view stereo (MVS) pipeline. Appearance information is sampled in an adaptive way so as to maximize the entropy of stored information. This is made possible through a homemade representation of appearance, different from the more traditional texture maps. This tool laso includes a compression module, so as to optimize disk space.

RELEASE FUNCTIONAL DESCRIPTION: 1st version

- Authors: Matthieu Armando, Edmond Bover, Jean-Sébastien Franco and Vincent Leroy
- Partner: Microsoft
- Contact: Matthieu Armando
- URL: https://gitlab.inria.fr/marmando/adaptive-mesh-texture

7. New Results

7.1. Surface Motion Capture Animation Synthesis

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 2).

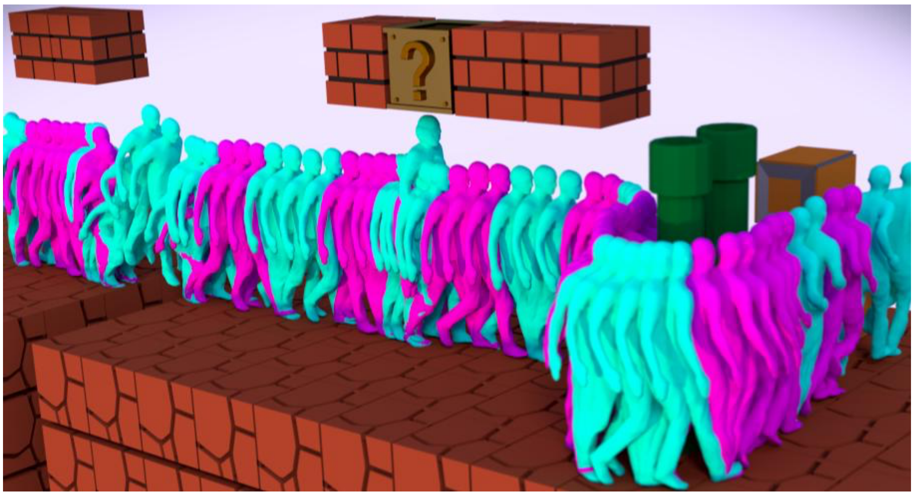

Figure 2. Animation Synthesis

This result was published in a prominent computer graphics journal, IEEE Transactions on Visualization and Computer Graphics [7].

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

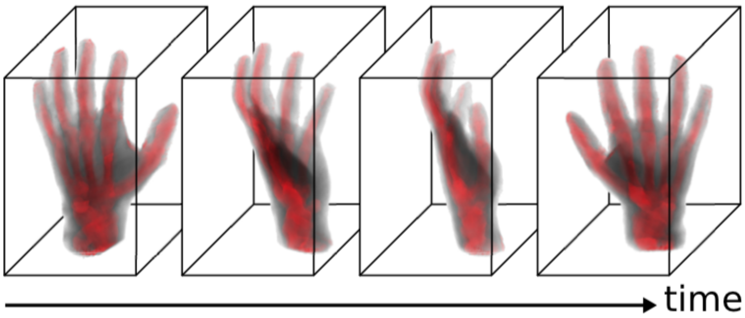

Figure 3. Dense volumetric attenuation reconstruction from a rigidly moving sample captured by a single planar X-ray imaging device and a surface motion capture system. Higher attenuation (here bone structure) is highlighted in red.

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 3.

This result was published in a prominent medical journal, IEEE Transactions on Medical Imaging [9].

7.3. Learning and Tracking the 3D Body Shape of Freely Moving Infants from RGB-D sequences

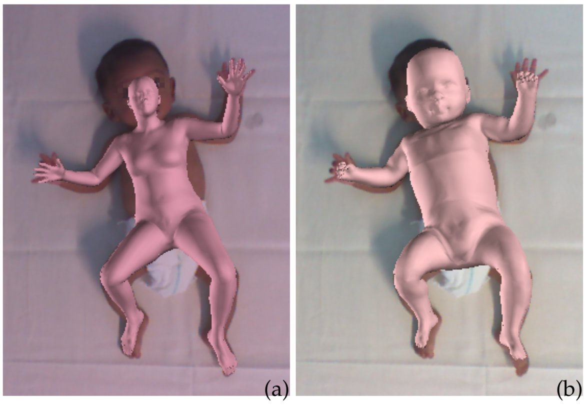

Figure 4. (a) Simply scaling a generic adult body model and fitting it to an infant does not work as body proportions significantly differ. (b) The proposed SMIL model properly captures the infants' shape and pose

Statistical models of the human body surface are generally learned from thousands of high-quality 3D scans in predefined poses to cover the wide variety of human body shapes and articulations. Acquisition of such data requires expensive equipment, calibration procedures, and is limited to cooperative subjects who can understand and follow instructions, such as adults. We presented a method for learning a statistical 3D Skinned Multi-Infant Linear body model (SMIL) from incomplete, low-quality RGB-D sequences of freely moving infants. Quantitative experiments show that SMIL faithfully represents the RGB-D data and properly factorizes the shape and pose of the infants. To demonstrate the applicability of SMIL, we fitted the model to RGB-D sequences of freely moving infants and show, with a case study, that our method captures enough motion detail for General Movements Assessment (GMA), a method used in clinical practice for early detection of neurodevelopmental disorders in infants. SMIL provides a new tool for analyzing infant shape and movement and is a step towards an automated system for GMA. This result was published in a prominent computer vision journal, IEEE Transactions on PAMI [8].

7.4. The Virtual Caliper: Rapid Creation of Metrically Accurate Avatars from 3D Measurements

Creating metrically accurate avatars is important for many applications such as virtual clothing try-on, ergonomics, medicine, immersive social media, telepresence, and gaming. Creating avatars that precisely represent a particular individual is challenging however, due to the need for expensive 3D scanners, privacy issues

Figure 5. Using the wand controllers of the HTC Vive, the Virtual Caliper produces a rigged 3D model with exactly the dimensions of the measured person.

with photographs or videos, and difficulty in making accurate tailoring measurements. We overcome these challenges by creating “The Virtual Caliper”, which uses VR game controllers to make simple measurements. First, we establish what body measurements users can reliably make on their own body. We find several distance measurements to be good candidates and then verify that these are linearly related to 3D body shape as represented by the SMPL body model. The Virtual Caliper enables novice users to accurately measure themselves and create an avatar with their own body shape. We evaluate the metric accuracy relative to ground truth 3D body scan data, compare the method quantitatively to other avatar creation tools, and perform extensive perceptual studies. We also provide a software application to the community that enables novices to rapidly create avatars in fewer than five minutes. Not only is our approach more rapid than existing methods, it exports a metrically accurate 3D avatar model that is rigged and skinned.

This result was published in a prominent computer graphics journal, IEEE Transactions on Visualization and Computer Graphics [10].

7.5. Adaptive Mesh Texture for Multi-View Appearance Modeling

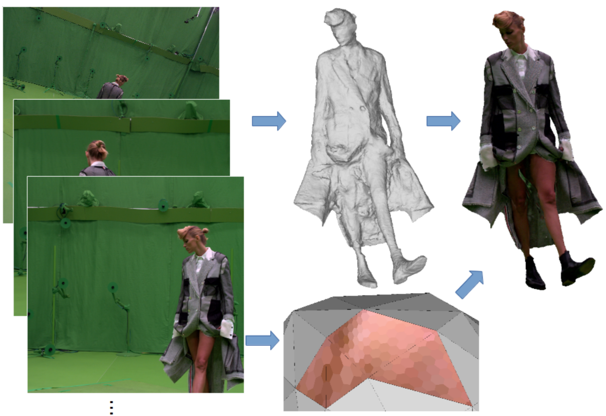

Figure 6. Texturing 3D models: given a set of input photographs (left), a geometric mesh is computed (top), along with an appearance function stored within the surface mesh structure (bottom).

Most applications in image based 3D modeling resort to texture maps, a 2D mapping of shape color information into image files. Despite their unquestionable merits, in particular the ability to apply standard image tools, including compression, image textures still suffer from limitations that result from the 2D mapping of information that originally belongs to a 3D structure. This is especially true with 2D texture atlases, a generic 2D mapping for 3D mesh models that introduces discontinuities in the texture space and plagues many 3D appearance algorithms. Moreover, the per-triangle texel density of 2D image textures cannot be individually adjusted to the corresponding pixel observation density without a global change in the atlas mapping function. To address these issues, we have proposed a new appearance representation for image-based 3D shape modeling, which stores appearance information directly on 3D meshes, rather than a texture atlas. We have shown this representation to allow for input-adaptive sampling and compression support. Our experiments demonstrated that it outperforms traditional image textures, in multi-view reconstruction contexts, with better visual quality and memory foot- print, which makes it a suitable tool when dealing with large amounts of data as with dynamic scene 3D models.

This result was published in the international conference on 3D Vision (3DV'19) [11].

7.6. Contact Preserving Shape Transfer for Motion Retargeting

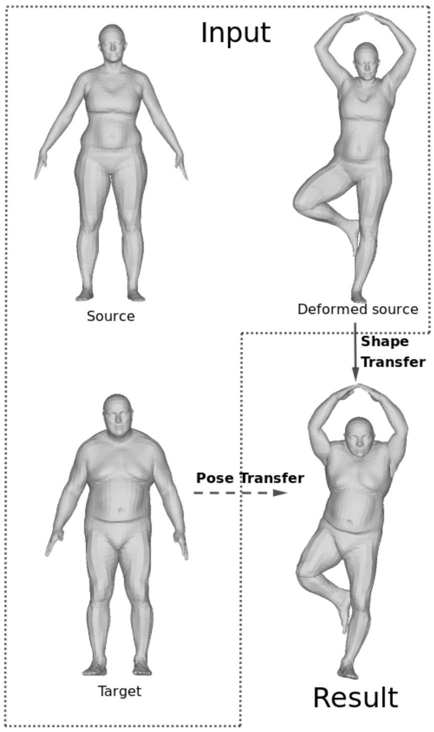

Figure 7. Motion Retargeting: Instead of transferring the pose from a source to a target shape, we propose to transfer the shape of the target to the deformed source character.

Retargeting a motion from a source to a target character is an important problem in computer animation, as it allows to reuse existing rigged databases or transfer motion capture to virtual characters. Surface based pose transfer is a promising approach to avoid the trial-and-error process when controlling the joint angles. In this work we investigated whether shape transfer instead of pose transfer would better preserve the original contextual meaning of the source pose. To this end, we proposed an optimization-based method to deform the source shape+pose using three main energy functions: similarity to the target shape, body part volume preservation, and collision management (preserve existing contacts and prevent penetrations). The results show that this strategy is able to retarget complex poses, including several contacts, to very different morphologies. In particular, we introduced new contacts that are linked to the change in morphology, and which would be difficult to obtain with previous works based on pose transfer that aim at distance preservation between body parts.

This result was published in the ACM SIGGRAPH Conference on Motion Interaction and Games (MIG'19) [12].

7.7. A Decoupled 3D Facial Shape Model by Adversarial Training

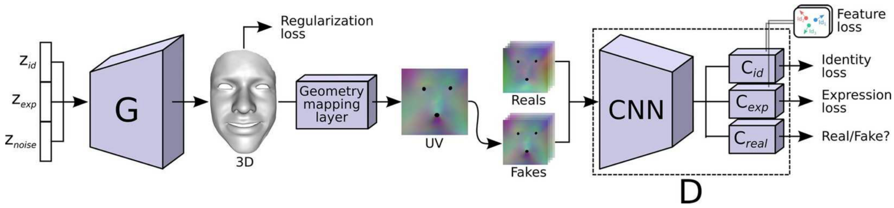

Figure 8. The face generator. Identity and expression codes zid , zexp are used to control the generator, and classification losses are added to decouple between the two. A feature loss is introduced to ensure consistency over features with fixed identities or expressions

Data-driven generative 3D face models are used to compactly encode facial shape data into meaningful parametric representations. A desirable property of these models is their ability to effectively decouple natural sources of variation, in particular identity and expression. While factorized representations have been proposed for that purpose, they are still limited in the variability they can capture and may present modeling artifacts when applied to tasks such as expression transfer. In this work, we explored a new direction with Generative Adversarial Networks and showed that they contribute to better face modeling performances, especially in decoupling natural factors, while also achieving more diverse samples. To train the model we introduced a novel architecture that combines a 3D generator with a 2D discriminator that leverages conventional CNNs, where the two components are bridged by a geometry mapping layer. We further presented a training scheme, based on auxiliary classifiers, to explicitly disentangle identity and expression attributes. Through quantitative and qualitative results on standard face datasets, we illustrated the benefits of our model and demonstrate that it outperforms competing state of the art methods in terms of decoupling and diversity.

This result was published in the international conference on computer vision (ICCV'19) [13]

7.8. Non-parametric 3D Human Shape Estimation from Single Images

In this work, we tackle the problem of 3D human shape estimation from single RGB images. While the recent progress in convolutional neural networks has allowed impressive results for 3D human pose estimation, estimating the full 3D shape of a person is still an open issue. Model-based approaches can output precise meshes of naked under-cloth human bodies but fail to estimate details and un-modelled elements such as hair or clothing. On the other hand, non-parametric volumetric approaches can potentially estimate complete shapes but, in practice, they are limited by the resolution of the output grid and cannot produce detailed estimates. In this work, we propose a non-parametric approach that employs a double depth map to represent the 3D shape of a person: a visible depth map and a "hidden" depth map are estimated and combined, to reconstruct

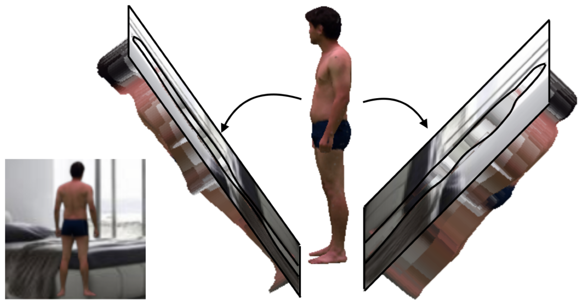

Figure 9. Given a single image, we estimate the "visible" and the "hidden" depth maps from the camera point of view. The two depth maps can be seen as the two halves of a virtual "mould".

the human 3D shape as done with a "mould". This representation through 2D depth maps allows a higher resolution output with a much lower dimension than voxel-based volumetric representations. Additionally, our fully derivable depth-based model allows us to efficiently incorporate a discriminator in an adversarial fashion to improve the accuracy and "humanness" of the 3D output. We train and quantitatively validate our approach on SURREAL and on 3D-HUMANS, a new photorealistic dataset made of semi-synthetic in-house images annotated with 3D ground truth surfaces.

This work was published in the international conference on computer vision (ICCV'19) [14]

7.9. Probabilistic Reconstruction Networks

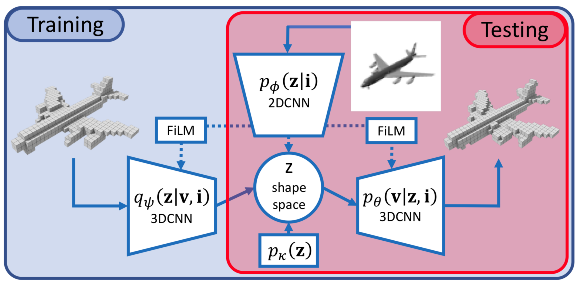

Figure 10. Probabilistic Reconstruction Networks for 3D shape inference from a single image. Arrows show the computational flow through the model, dotted arrows show optional image conditioning. The inference network q_Ψ is only used during training for variational inference

We study end-to-end learning strategies for 3D shape inference from images, in particular from a single image. Several approaches in this direction have been investigated that explore different shape representations and suitable learning architectures. We focus instead on the underlying probabilistic mechanisms involved and contribute a more principled probabilistic inference-based reconstruction framework, which we coin Probabilistic Reconstruction Networks. This framework expresses image conditioned 3D shape inference through a family of latent variable models, and naturally decouples the choice of shape representations from the inference itself. Moreover, it suggests different options for the image conditioning and allows training in two regimes, using either Monte Carlo or variational approximation of the marginal likelihood. Using our Probabilistic Reconstruction Networks we obtain single image 3D reconstruction results that set a new state of the art on the ShapeNet dataset in terms of the intersection over union and earth mover's distance evaluation metrics. Interestingly, we obtain these results using a basic voxel grid representation, improving over recent work based on finer point cloud or mesh based representations.

This work was published in the British machine vision conference (BMVC'19) [15] where it won the runner-up best paper award.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

1. The Morpheo Inria team and Microsoft research set up a collaboration on the capture and modelling of moving shapes using multiple videos. Two PhD works are part of this collaboration with the objective to make contributions on 4D Modeling. The PhDs take place at Inria Grenoble Rhône-Alpes and involve visits and stays at Microsoft in Cambridge (UK) and Zurich (CH). The collaboration is part of the Microsoft Research - Inria Joint Centre.
2. The Morpheo Inria team has another collaboration with Facebook reality lab in San Francisco. The collaboration involves one PhD who is currently at the Inria Grenoble Rhône-Alpes working on the estimation of shape and appearance from a single image. The collaboration started in 2019.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Data Driven 3D Vision

Edmond Boyer obtained a chair in the new Multidisciplinary Institute in Artificial Intelligence (MIAI) of Grenoble Alpes University. The chair entitled Data Driven 3D Vision is for 4 years and aims at investigating deep learning for 3D artificial vision in order to break some of the limitations in this domain. Applications are especially related to humans and to the ability to capture and analyze their shapes, appearances and motions, for upcoming new media devices, sport and medical applications.

9.2. National Initiatives

9.2.1. ANR

9.2.1.1. 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 and is currently handled by Tomas Svaton, recruited as an engineer in April 2018.

9.2.1.2. ANR JCJC SEMBA – Shape, Motion and Body composition to Anatomy

Existing medical imaging techniques, such as Computed Tomography (CT), Dual Energy X-Ray Absorption (DEXA) and Magnetic Resonance Imaging (MRI), allow to observe internal tissues (such as adipose, muscle, and bone tissues) of in-vivo patients. However, these imaging modalities involve heavy and expensive equipment as well as time consuming procedures. External dynamic measurements can be acquired with optical scanning equipment, e.g. cameras or depth sensors. These allow high spatial and temporal resolution acquisitions of the surface of living moving bodies. The main research question of SEMBA is: "can the internal observations be inferred from the dynamic external ones only?". SEMBA's first hypothesis is that the quantity and distribution of adipose, muscle and bone tissues determine the shape of the surface of a person. However, two subjects with a similar shape may have different quantities and distributions of these tissues. Quantifying adipose, bone and muscle tissue from only a static observation of the surface of the human might be ambiguous. SEMBA's second hypothesis is that the shape deformations observed while the body performs highly dynamic motions will help disambiguating the amount and distribution of the different tissues. The dynamics contain key information that is not present in the static shape. SEMBA's first objective is to learn statistical anatomic models with accurate distributions of adipose, muscle, and bone tissue. These models are going to be learned by leveraging medical dataset containing MRI and DEXA images. SEMBA's second objective will be to develop computational models to obtain a subject-specific anatomic model with an accurate distribution of adipose, muscle, and bone tissue from external dynamic measurements only.

9.2.1.3. ANR JCJC 3DMOVE - Learning to synthesize 3D dynamic human motion

It is now possible to capture time-varying 3D point clouds at high spatial and temporal resolution. This allows for high-quality acquisitions of human bodies and faces in motion. However, tools to process and analyze these data robustly and automatically are missing. Such tools are critical to learning generative models of human motion, which can be leveraged to create plausible synthetic human motion sequences. This has the potential to influence virtual reality applications such as virtual change rooms or crowd simulations. Developing such tools is challenging due to the high variability in human shape and motion and due to significant geometric and topological acquisition noise present in state-of-the-art acquisitions. The main objective of 3DMOVE is to automatically compute high-quality generative models from a database of raw dense 3D motion sequences for human bodies and faces. To achieve this objective, 3DMOVE will leverage recently developed deep learning techniques. The project also involves developing tools to assess the quality of the generated motions using perceptual studies. This project currently involves one Ph.D. student who was hired in November 2019.

9.2.2. Competitivity Clusters

9.2.2.1. 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.

9.3. International Research Visitors

The Morpheo team is hosting Professor Stephane Durocher during his sabbatical from July 2019 to June 2020. He is involved in the team research activities, in particular on the development of efficient algorithms to cluster a set of moving objects based on their trajectories, as obtained using the Kinovis platform. This will allow to perform motion analysis tasks, such as clustering objects into components that follow similar motions, which can help in analyzing the relative motion of body parts.

9.3.1. Visits to International Teams

9.3.1.1. Research Stays Abroad

1. Victoria Fernández Abrevaya did an internship with a British company in London, from July 2019 until September 2019.
2. Nitika Verma did an intership with Google at New York, from May 2019 until September 2019.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Selection

10.1.1.1. Chair of Conference Program Committees

- Stefanie Wuhrer was program co-chair of 3DV 2019, held in Québec City, Canada

10.1.1.2. Member of the Conference Program Committees

- Edmond Boyer was area chair for BMVC and ICCV 2019.

10.1.1.3. Reviewer

- Stefanie Wuhrer reviewed for CVPR, ICCV, and SIGGRAPH
- Jean-Sébastien Franco reviewed for CVPR, ICCV, and 3DV
- Edmond Boyer reviewed for CVPR.

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 (IJCV, Springer).

10.1.2.2. Reviewer - Reviewing Activities

- Edmond Boyer reviewed for IJCV and PAMI.
- Jean-Sébastien Franco reviewed for IJCV and Computer & Graphics.

10.1.3. Invited Talks

- Stefanie Wuhrer was invited to present at Dagstuhl seminar 19102
- Edmond Boyer gave invited talks at: ETH Zurich (February), Prague university (April), Naverlabs Grenoble (May), Lille University (June), Microsoft Zurich (October).

10.1.4. Scientific Expertise

- Jean-Sebastien Franco was a member of the Ensimag Engineering school - Grenoble INP steering commitee (Conseil d'École) in 2019.
- Jean-Sebastien Franco was a member of the recruiting commitee of Ensimag - Grenoble INP Engineering school for temporary research and teaching associates (ATER) in 2019.

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 : Sergi Pujades, Algorithmique et programmation fonctionnelle, 41H équivalent TD, L1, Université Grenoble Alpes, France.

Licence : Sergi Pujades, Modélisation des structures informatiques: aspects formels, 45H équivalent TD, L1, Université Grenoble Alpes, France.

Master : Sergi Pujades, Numerical Geometry, 15H équivalent TD, M1, Université Grenoble Alpes, France

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

Licence: Jean-Sébastien Franco, C Programming project, 20h, 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, 45h, 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), 11h, Ensimag 3rd year, Grenoble INP.

Master: Edmond Boyer, 3D Modeling, 23h, M2R Mosig GVR, Grenoble INP.

Master: Edmond Boyer, Introduction to Visual Computing, 42h, M1 MoSig, Université Grenoble Alpes.

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

10.2.2. Supervision

PhD: Jinlong Yang, Learning shape space of dressed human in motion, Université Grenoble Alpes (France), March 2019, Franck Hétroy-Wheeler and Stefanie Wuhrer.

PhD: Vincent Leroy, 4D shape reconstruction from photoconsistency cues, Université Grenoble Alpes, October 2019, supervised by Edmond Boyer and Jean-Sébastien Franco.

PhD in progress : Sanae Dariouche, Anatomic Statistical Models of adipose, bone and muscle tissue, 01/11/2019, Sergi Pujades and Edmond Boyer

PhD in progress : Marilyn Keller, Ribs motion models for personalized 3D printed implants, 23/09/2019, Sergi Pujades and Michael Black

PhD in progress : Di Meng, , 23/09/2019, Edmond Boyer and Sergi Pujades

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: Boyao Zhou, Augmenting User Self-Representation in VR Environments, Université Grenoble Alpes (France), started 01/02/2019, supervised by Edmond Boyer and Jean-Sébastien Franco.

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: Pierre Zins, Learning to infer human motion, Université Grenoble Alpes (France), started 01/10/2019, supervised by Edmond Boyer, Tony Tung (Facebook) and Stefanie Wuhrer.

PhD in progress: Mathieu Marsot, Generative modeling of 3D human motion, Université Grenoble Alpes (France), started 01/11/2019, supervised by Jean-Sébastien Franco and Stefanie Wuhrer.

10.2.3. Juries

- Stefanie Wuhrer was jury member for the PhD defense of Alban Fichet, Université Grenoble Alpes, December 2019.
- Edmond Boyer was president of the PhD jury of Romain Rombourg, Université Grenoble Alpes, December 2019.
- Edmond Boyer was jury member for the PhD defense of Arthur Crenn, Université de Lyon, December 2019.

10.3. Popularization

10.3.1. Articles and contents

- Edmond Boyer was interviewed by the journal *Vivre à Grenoble* en février 2019. An article mentioning his work with the Kinovis platform was published in april 2019.

11. Bibliography

Major publications by the team in recent years

[1] A. BOUKHAYMA, E. BOYER. *Surface Motion Capture Animation Synthesis*, in "IEEE Transactions on Visualization and Computer Graphics", June 2019, vol. 25, n^{o} 6, pp. 2270-2283 [*DOI :* 10.1109/TVCG.2018.2831233], https://hal.inria.fr/hal-01781164

[2] V. LEROY, J.-S. FRANCO, E. BOYER. *Shape Reconstruction Using Volume Sweeping and Learned Photoconsistency*, in "European Conference on Computer Vision", Munich, Germany, Lecture Notes in Computer Science, Springer, September 2018, vol. 11213, pp. 796-811 [*DOI :* 10.1007/978-3-030-01240-3_48], https://hal.archives-ouvertes.fr/hal-01849286

[3] J. PANSIOT, E. BOYER. *CBCT of a Moving Sample from X-rays and Multiple Videos*, in "IEEE Transactions on Medical Imaging", February 2019, vol. 38, n^{o} 2, pp. 383-393 [*DOI :* 10.1109/TMI.2018.2865228], https://hal.inria.fr/hal-01857487

[4] 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, IEEE, 2018, pp. 2598-2606 [*DOI :* 10.1109/CVPR.2018.00275], https://hal.inria.fr/hal-01540389

Publications of the year

Doctoral Dissertations and Habilitation Theses

[5] V. LEROY. *Fast and Accurate 4D Modeling of Large Multi-Camera Sequences*, Université Grenoble Alpes (2016-2019) ; Inria Grenoble, October 2019, https://hal.archives-ouvertes.fr/tel-02435385

[6] J. YANG. *Learning shape spaces of dressed 3D human models in motion*, Université Grenoble Alpes, March 2019, https://tel.archives-ouvertes.fr/tel-02091727

Articles in International Peer-Reviewed Journals

[7] A. BOUKHAYMA, E. BOYER. *Surface Motion Capture Animation Synthesis*, in "IEEE Transactions on Visualization and Computer Graphics", June 2019, vol. 25, n^{o} 6, pp. 2270-2283 [*DOI :* 10.1109/TVCG.2018.2831233], https://hal.inria.fr/hal-01781164

[8] N. HESSE, S. PUJADES, M. J. BLACK, M. ARENS, U. HOFMANN, S. SCHROEDER. *Learning and Tracking the 3D Body Shape of Freely Moving Infants from RGB-D sequences*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", June 2019, 12 p. [*DOI :* 10.1109/TPAMI.2019.2917908], https://hal.inria.fr/hal-02162171

[9] J. PANSIOT, E. BOYER. *CBCT of a Moving Sample from X-rays and Multiple Videos*, in "IEEE Transactions on Medical Imaging", February 2019, vol. 38, n^{o} 2, pp. 383-393 [*DOI :* 10.1109/TMI.2018.2865228], https://hal.inria.fr/hal-01857487

[10] S. PUJADES, B. MOHLER, A. THALER, J. TESCH, N. MAHMOOD, N. HESSE, H. H. BÜLTHOFF, M. J. BLACK. *The Virtual Caliper: Rapid Creation of Metrically Accurate Avatars from 3D Measurements*, in "IEEE Transactions on Visualization and Computer Graphics", May 2019, vol. 25, n^{o} 5, pp. 1887-1897 [*DOI :* 10.1109/TVCG.2019.2898748], https://hal.inria.fr/hal-02162159

International Conferences with Proceedings

[11] M. ARMANDO, J.-S. FRANCO, E. BOYER. *Adaptive Mesh Texture for Multi-View Appearance Modeling*, in "3DV 2019 - 7th International Conference on 3D Vision", Quebec City, Canada, IEEE, September 2019, pp. 1-9, https://hal.inria.fr/hal-02284101

[12] J. BASSET, S. WUHRER, E. BOYER, F. MULTON. *Contact Preserving Shape Transfer For Rigging-Free Motion Retargeting*, in "MIG 2019 - ACM SIGGRAPH Conference Motion Interaction and Games", Newcastle Upon Tyne, United Kingdom, ACM, October 2019, pp. 1-10 [*DOI :* 10.1145/3359566.3360075], https://hal.archives-ouvertes.fr/hal-02293308

[13] V. FERNÁNDEZ ABREVAYA, A. BOUKHAYMA, S. WUHRER, E. BOYER. *A Decoupled 3D Facial Shape Model by Adversarial Training*, in "ICCV 2019 - International Conference on Computer Vision", Seoul, South Korea, October 2019, pp. 1-10, https://hal.archives-ouvertes.fr/hal-02064711

[14] V. GABEUR, J.-S. FRANCO, X. MARTIN, C. SCHMID, G. ROGEZ. *Moulding Humans: Non-parametric 3D Human Shape Estimation from Single Images*, in "ICCV 2019 - International Conference on Computer Vision", Seoul, South Korea, October 2019, pp. 1-10, https://hal.inria.fr/hal-02242795

[15] *Best Paper*
R. KLOKOV, J. VERBEEK, E. BOYER. *Probabilistic Reconstruction Networks for 3D Shape Inference from a Single Image*, in "BMVC 2019 - British Machine Vision Conference", Cardiff, United Kingdom, September 2019, pp. 1-15, https://arxiv.org/abs/1908.07475 - Awarded with Best Science Paper Honourable Mention Award at BMVC'19., https://hal.inria.fr/hal-02268466.

[16] 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 "APS 2019 - American Physical Society March Meeting", Boston, United States, March 2019, 1 p. , https://hal.inria.fr/hal-01982257

Other Publications

[17] B. EGGER, W. A. P. SMITH, A. TEWARI, S. WUHRER, M. ZOLLHOEFER, T. BEELER, F. BERNARD, T. BOLKART, A. KORTYLEWSKI, S. ROMDHANI, C. THEOBALT, V. BLANZ, T. VETTER. *3D Morphable Face Models - Past, Present and Future*, September 2019, https://arxiv.org/abs/1909.01815 - working paper or preprint, https://hal.inria.fr/hal-02280281