Master 2 Internship subject:

Representing and Learning over 3D Evolving Shapes

Hosting institute

— ICube laboratory: Le laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie (The Engineering science, computer science and imaging laboratory), http://icube.unistra.fr/

— University of Strasboug, France, http://www.unistra.fr/

Work place and salary

Place de l'hôpital, Strasbourg (67), France.

Salary: 500€/month approximately for a duration of 6 months.

Supervisors

— director: Hyewon Seo (ICube, Univ. Strasbourg), seo@unistra.fr

— co-supervisor: Frederic Cordier (Univ. Haut Alsace), fcordier@unistra.fr

Starting date

December 2019 ~ January 2020.

Context

Reconstructing, characterizing, and understanding the shape and motion of individuals or groups of people have many important applications, such as ergonomic design of products, rapid reconstruction of realistic human models for virtual worlds, and an early detection of abnormality in predictive clinical analysis. Naturally, the capture and analysis of people's shape and motion have a long tradition in disciplines such as computer vision, computer graphics, and virtual reality. This is evidenced by the large amount of research done on shape reconstruction from images and 3D scans, on subspace construction with multiple shapes, on motion capture and action recognition from video inputs. However, most current techniques treat shape and motion independently, with devoted techniques for either shape or motion in isolation. This is largely due to the difficulty of acquiring proper observations on full moving shapes: Traditional systems have been devoted to capture either static shapes, e.g. 3D scanners, or motion only, e.g. motion capture.

Recent evolutions in the technology for capturing moving shapes have changed this paradigm with new multi-view acquisition systems that enable now full 4D models of human shapes including geometry, motion and appearance, as in Microsoft [CCS+15], Inria [Kino], MPI [DAST+08], or more recently with commercial platforms deployed by Intel [Ints], 8i [8i] or Microsoft [Holo], among others (see Figure 1). Such data open new possibilities and challenges for the analysis and the synthesis of human shapes in motion that are yet largely unexplored but would be of benefit to a wide range of applications in virtual and augmented reality, or in the sport and medical domains, among others. This is especially true with the rapidly growing VR/AR immersive applications based on head mounted displays, which require realistic and detailed models to improve the immersive experience. Magic leap, Microsoft Hololense, Facebook Oculus Rift, Sony PS4 HMD and the HTC Vive, among others, are clear examples of this recent and rapid evolution and the associated need to produce adapted realistic contents. In the future we will be able to make digital copies of moving persons using a handy imaging device, send them over the network, and make customized compositions of the retrieved 4D human data in our daily life.

Figure 1: Moving shapes can now be captured with recent multi-view shape acquisition systems.

In the framework of a French national project[1] Human4D (Acquisition, Analysis and Synthesis of Human Body Shape in Motion), **we aim at contributing to this evolution with objectives that can profoundly improve the reconstruction, transmission, and reuse of digital human data, by unleashing the power of recent deep learning techniques and extending it to 4D human shape modeling.** This project is timely, and gathers French research teams with long-standing expertise in the field, associated to one of the only acquisition platforms worldwide able yet to produce the required data on moving human bodies [Kino].

Detailed goals

In the last decade, research in computer vision and artificial intelligence has achieved disruptive results in the recognition and synthesis of objects in the image by means of large annotated datasets, deep learning (DL) algorithms and adequate GPU resources. However, with most existing architectures and algorithms having developed for 2D images, their adaptation to 3D data (point clouds or meshes) is less obvious, where a regular structure is not directly available. While DL CNNs have been used in some 3D contexts, e.g. face modeling [TZK+17][TL18] or shape classification, their interest in live modeling of complex, articulated shapes like human body has not yet been fully explored, and extending the learning ability to 4D context remains an unexplored area. Not surprisingly, the ability to model, analyze, and synthesize 4D (3D+time) human models as fully dynamic models is limited in the current techniques, where geometry and appearance are modeled with static poses in a frame-by-frame manner: They do not model the 4D human shape in a true sense, with the temporal evolution of 3D data neglected in the model.

Human4D project takes up new challenges in the context of flagship human body modeling, aiming at a new, efficient 4D shape modelling of human body under motion. Our ambition is to go beyond existing shape space representations that mostly focus on static shape poses, as in the seminal work [ASKT+05], and seldom consider the continuous dynamic of body shapes. Although all these ideas exist already and may not novel per say, no existing work has achieved a similar goal so far, especially on the moving human body shapes. Indeed, the complexity of human body shape and motion as well as the persistent change in geometry and topology of the time-varying data make most traditional shape analysis algorithms unsuitable.

In this study, we will propose a composite deep convolutional neural network that learns to predict both the shape and the motion of the human body, based on compact representations of 4D human atlas which models the commonality and variability over motions and individuals. While atlases have already been successfully used to model static human poses [ACPH06], the case of dynamic motions (e.g. running, jumping) remains mostly

[1] http://anr.fr

unexplored. We will carry out co-analysis of multiple dataset, based on the per-subject analysis that is done independently from each other. This will require a spatio-temporal correspondence established across multiple 4D dataset, which is particularly challenging since 4D data is typically of high dimension both spatially and temporally [SC16]. We will rely on the compact mathematical representations of individual dynamic shapes, combined with good levels of abstraction based on our previous studies [MSC15][LCS16], in order to find the spatio-temporal correspondence efficiently and reliably. Another critical issue is the nonlinearity of such shape spaces and that of temporal shape evolutions, for which standard linear statistics, e.g. [ASKT+05] are not appropriate. To this end, we set our goal to define nonlinear statistical models of spatiotemporal variability among multiple 4D shapes and infer the parameters of these models from our datasets. Finally, we will demonstrate the power of 4D atlases by developing an illustrative application that will exploit the 4D atlases' ability on learning and inference.

Requirements

— Master degree in Computer Science or in (Applied) Mathematics

— Working knowledge of programming in Python/Matlab

— Good notion of statistics and kinematic modeling

— Experience in machine learning and/or Tensorflow is a plus

References

[8i] https://8i.com.

[ACPH06] B. Allen, B. Curless, Z. Popotić, and A. Hertzmann: Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. Proc. ACM SIGGRAPH/Eurographics symposium on Computer animation, pp.147-156, Vienna, Austria, 2006.

[ASKT+05] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis.: SCAPE: shape completion and animation of people. ACM Trans. Graphics (Proc. SIGGRAPH 2005).

[CCS+15] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe, A. Kirk, S. Sullivan: High-quality streamable free-viewpoint video. ACM Trans. Graphics (Proc. ACM SIGGRAPH 2015).

[DAST+08] E. De Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, S. Thrun: Performance capture from sparse multi-view video, ACM Trans. Graphics (Proc. ACM SIGGRAPH 2008).

[Holo] https://www.microsoft.com/en-us/mixed-reality/capture-studios.

[Ints] Intel Studios, Los Angeles.

[Kino] http://kinovis.inrialpes.fr

[LCS16] G. Luo, F. Cordier, and H. Seo: Spatio-temporal Segmentation for the Similarity Measurement of Deforming Meshes, The Visual Computer, Vol.32, No.2, pp.243-256, Springer, 2016.

[MSC15] V. Mykhalchuk, H. Seo, F. Cordier: On Spatio-Temporal Feature Point Detection for Animated Meshes, The Visual Computer, Springer, 31(11), pp. 1471-1486, 2015.

[SC16] H. Seo, F. Cordier: Spatial Matching of Animated Meshes, Computer Graphics Forum (Proc. Pacific Graphics), Vol. 35, Num. 7, pp.21-32, Wiley, 2016.

[TL18] L. Tran, X. Liu: Nonlinear 3D Face Morphable Model. IEEE Conf. IEEE Conf. Computer Vision and Pattern Recognition, 2018

[TZK+17] A. Tewari, M. Zollhofer, H. Kim, P. Garrido, F. Bernard, P. Perez, C. Theobalt: MoFA: Model-Based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction. Int'l Conf. Computer Vision 2017.