



Live Digital Human Generation using Dense Pose

Project Management and Software Development
for Medical Applications

General Info

Contact Person: Mahdi Saleh

Contact Email: m.saleh@tum.de

Project Abstract

A Digital human is a virtual 3D double of a human from the real world. Such an avatar representation can be visualized and used in many applications. Standard human pose estimation algorithms provide a coarse and simplified understanding of the body. In recent years there has been considerable interest in bringing high-latency avatars and modelling the human body. In this project, we want to combine state-of-the-art human pose estimation methods with known 3D captures of the human body to animate a high-definition avatar given the predicted human pose.

Background and Motivation

Human pose estimation has been an active field of research in computer vision. With the advent of deep learning models, neural networks could predict a set of predefined body joints in the image frame [1,2]. By seeing millions of image frames, the models can estimate the human skeleton in normalized 3D space [3]. A simplified body skeleton is functional in solutions where an understanding of human gestures or joints is needed. However, such models provide a weak potential for visualization purposes.

Human pose estimation is a significant building block in many AR and VR applications. For instance, medical telepresence requires a high definition of digital content and reliable visualization of the patient and the medical staff.

Derived from the computer graphics field, there have been efforts to model the human body and pose in a simpler form than dense mesh or point clouds. SMPL [4], for example, provides a

parametrized human model able to capture pose and shape features. Recent dense human pose estimation networks can predict human pose as an SMPL model [5,6,7].

In this project, we want to embody an SMPL model with a human avatar captured from real scans. This way, we can augment a high-definition avatar in AR applications.

Student's Tasks Description

The project aims to register a 3D avatar mesh with real-time human pose estimation. To do this, the student first needs to make a reference scan or a 3D avatar. By using multiple RGBD scans of a person and fusion techniques, a mesh is generated in canonical T-Pose.

The student then needs to run a dense human pose estimation algorithm. A pretrained deep neural network can regress an SMPL model [4] from a live capture.

Next, the software registers the SMPL model to the reference T-Pose of the avatar. This would help to model and update the digital human given the predicted pose estimation.

Finally, the software should visualize the result of a live digital human as an augmentation over the live feed using common libraries.

Technical Prerequisites

Recommended:

Good knowledge on computer vision and deep learning

Good knowledge in python programming

Optional:

Experience in computer graphics and visualization frameworks

Experience with 3D scanning and data

Please send the completed proposal to ardit.ramadani@tum.de, lennart.bastian@tum.de and tianyu.song@tum.de. Please note that this proposal will be evaluated by the BMC coordinators and will be assigned to a student only in case of acceptance.



References

- [1] Toshev, Alexander, and Christian Szegedy. "DeepPose: Human pose estimation via deep neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
- [2] Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7291-7299).
- [3] Mehta, D., Sotnychenko, O., Mueller, F., Xu, W., Elgharib, M., Fua, P., ... & Theobalt, C. (2020). XNect: Real-time multi-person 3D motion capture with a single RGB camera. *Acm Transactions On Graphics (TOG)*, 39(4), 82-1.
- [4] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: A skinned multi-person linear model. *ACM transactions on graphics (TOG)*, 34(6), 1-16.
- [5] Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., & Black, M. J. (2016, October). Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. In *European conference on computer vision* (pp. 561-578). Springer, Cham.
- [6] Yuan, Y., Wei, S. E., Simon, T., Kitani, K., & Saragih, J. (2021). Simpose: Simulated character control for 3d human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7159-7169).
- [7] Rempe, D., Birdal, T., Hertzmann, A., Yang, J., Sridhar, S., & Guibas, L. J. (2021). Humor: 3d human motion model for robust pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 11488-11499).