Real-Time Human Motion Transfer System for Holographic Displays

Wenkang Shan[1], Haopeng Lu[2], Chuanmin Jia[1], Xinfeng Zhang[3,*], Siwei Ma[1], Yaqiang Wu[4], Wen Gao[1]
[1]Peking University
[2]Shanghai Jiao Tong University
[3]University of Chinese Academy of Sciences
[4]Lenovo (Beijing) Co., Ltd
{wkshan, cmjia, swma, wgao}@pku.edu.cn
luhp2018@sjtu.edu.cn, xfzhang@ucas.ac.cn, wuyqe@lenovo.com

***Abstract*—This paper presents a human motion transfer system designed for holographic displays. It utilizes a single RGB camera to capture real-time human movement data. The captured data is then processed through off-the-shelf models for 2D and 3D human pose estimation. Subsequently, the estimated poses are retargeted to a pre-defined 3D avatar using forward kinematics, enabling high-fidelity motion transfer. Demonstrations on both a holographic cabin and Microsoft HoloLens 2 showcase the system's ability to enhance the expressiveness and authenticity of digital humans, thus paving the way for more intuitive interactions within the metaverse.**

***Index Terms*—Human pose estimation, Motion retargeting, Holographic, Digital human, Metaverse**

I. Introduction

The metaverse is an emerging concept that envisions a virtual world where people can interact with each other across various platforms. To achieve a high level of immersion and social presence in the metaverse, users need realistic and expressive digital humans that can mimic human beings in appearance and behavior. However, constructing convincing digital human motion is a difficult problem that requires a significant amount of manpower and computational resources.

The traditional methods for creating digital human motion include handcrafted animation and motion capture systems. Handcrafted animation is time-consuming, labor-intensive, and often limited by the artistic skills of the animators. Meanwhile, motion capture systems suffer from high costs, complex setups, and restricted environments. These methods are unsuitable for real-time applications due to their lack of immediacy and interaction. Additionally, traditional displays like monitors and head-mounted devices limit the spatial and perspective effects of digital human motion.

To alleviate the existing problems, this paper demonstrates a real-time human motion transfer system for holographic displays. The proposed system aims to map the movements of a real person to a virtual character in real time. To achieve a more natural and intuitive presentation of digital humans in the metaverse, holographic displays, which can project 3D models into the air, are a promising alternative. Specifically, we use a single RGB camera to capture the movements of a real person, synthesize realistic and consistent virtual character animations, and present them on a holographic cabin and Microsoft HoloLens 2. The proposed system shows a novel way of displaying human movements while reducing production costs and complexity.

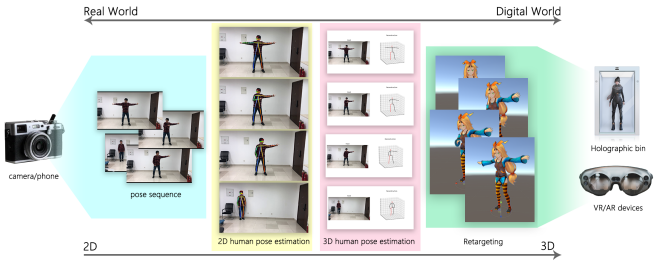

Fig. 1. The framework of the proposed system.

II. Framework

The proposed framework, illustrated in Fig. 1, is described step-by-step below.

Video capture. The system can accept image data from various devices such as smartphones and cameras. To ensure accurate tracking, we recommend using a clean background and capturing videos where all the user's limbs are fully visible within the frame.

2D human pose estimation. We use an off-the-shelf 2D keypoint detector [1] to yield 2D poses from images. This method crops and resizes the input image, and outputs the 2D coordinates of 17 keypoints.

3D human pose estimation. We use P-STMO [2] to estimate 3D poses from 2D keypoints and TCN [3] to regress the 3D global trajectory. These estimators take 2D keypoints in COCO format as input and output 3D joint positions in Human3.6M format. We transform the network weights of the 2D and 3D pose estimators into onnx files, which are read by Barracuda in Unity3D.

Retargeting. To create a more immersive experience, we adapt an existing tool [1] to retarget estimated poses to 3D

* Corresponding author

[1]https://github.com/digital-standard/ThreeDPoseUnityBarracuda

Fig. 2. Demonstration results in the holographic cabin and HoloLens 2. (a) System setup. (b)-(c) The results of the 'Rock Hippie' and 'Biomechanical Mutant' avatars. (d) The results captured directly through a mobile phone. (e) The results captured using the built-in recorder.

avatars for holographic displays. This involves upgrading the tool with state-of-the-art models (P-STMO) for better pose estimation accuracy. Additionally, optimizations are made to ensure seamless transfer of content originally designed for on-screen displays to the holographic cabin and HoloLens 2. Utilizing rigged 3D models sourced from TurboSquid as avatars, we employ forward kinematics to align the avatar's skeleton vectors with the direction of extracted 3D pose vectors, ensuring a natural and accurate motion transfer.

Display. This system showcases results through the holographic cabin and Microsoft HoloLens 2, which are introduced in detail below.

The holographic cabin consists of a 4K LCD screen with an enclosed transparent cabin. By matching the on-screen content display with the background lighting, the holographic cabin creates an illusion that the 3D avatar is standing inside the cabin. As shown in Fig. 2 (a), the proposed system can be easily assembled by connecting a standalone PC and a high frame rate camera to the holographic cabin. Fig. 2 (b) and Fig. 2 (c) show the subjective results of the proposed system. The user can adjust their standing position by observing the real-time camera view in the upper right corner.

Microsoft HoloLens 2 is a leading mixed reality (MR) headset that seamlessly blends the physical and digital worlds by projecting high-resolution holograms directly onto the user's field of view. The 3D avatars are dynamically rendered based on captured data, ensuring an accurate representation of user movements and gestures. Fig. 2 (d) and Fig. 2 (e) demonstrate the actual and simulated viewing effect of the proposed system on HoloLens 2, respectively.

III. Conclusion

This paper presents a novel real-time human motion transfer system for holographic displays. Human movements are captured and efficiently translated into animations for a predefined 3D avatar, enabling a more natural and interactive user experience. Compatibility with both a holographic cabin and Microsoft HoloLens 2 showcases the system's versatility for various holographic display applications.

While the proposed human motion transfer system demonstrates promising results, it's essential to acknowledge its limitations, particularly when operating in complex or dynamic backgrounds. In such scenarios, the accuracy of pose estimation may degrade due to occlusions, fast movements, or cluttered environments, which impacts the fidelity of motion transfer. Addressing these challenges requires further research in robust pose estimation algorithms tailored for real-time applications in dynamic settings, ensuring consistent performance across diverse environmental conditions.

Acknowledgment

This work was supported by the National Key R&D Program of China 2022YFF0902400, NSFC 62025101, NSFC U21B2012, and the Xplore Prize, which is gratefully acknowledged.

References

[1] Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, and Cewu Lu, "Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[2] Wenkang Shan, Zhenhua Liu, Xinfeng Zhang, Shanshe Wang, Siwei Ma, and Wen Gao, "P-stmo: Pre-trained spatial temporal many-to-one model for 3d human pose estimation," in *European Conference on Computer Vision*. Springer, 2022, pp. 461–478.

[3] Dario Pavllo, Christoph Feichtenhofer, David Grangier, and Michael Auli, "3d human pose estimation in video with temporal convolutions and semi-supervised training," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 7753–7762.