3D Human Body Pose Estimation in Virtual Reality: A survey

1st Taravat Anvari
dept. of Computer Software (of Chung-Ang university)
Seoul, South Korea
taravat26anvari@cau.ac.kr

2nd Kyoungju Park
dept. of Computer Software (of Chung-Ang university)
Seoul, South Korea
kjpark@cau.ac.kr

***Abstract*—In Virtual Reality (VR), human pose estimation refers to the approximation of the location of body parts of a real human in a physical world. During the past decades, virtual reality has a widespread application in many areas such as the video game industry, user interfaces, and visual simulation. Hence the human pose estimation has recently attracted increasing attention in the VR community. However, it is challenging due to the variety of human movements. This paper discusses different methods for human motion estimation that have been investigated for years. We cover a variety of approaches, including motion capture techniques, inverse kinematics (IK), and the learning-based techniques that are the primary solutions to estimate the body joints in real-time in this scope.**

***Index Terms*—Virtual Reality; Human Body; Pose Estimation**

I. Introduction

Presence or self-location refers to the sensation of physically being at a virtual place while knowing that one is not there [AMTK83], [SBW17]. In addition, embodiment describes the sensation that our self is located inside a virtual body; we control this body and that this body belongs to us [KGS12]. Utilizing a Virtual Reality (VR) system, it is possible to induce a sense of embodiment toward virtual bodies [SPMESV08]. A virtual reality system can display immersive virtual environments where participants feel fully present and immersed in the virtual environment [Sla18]. In other words, localizing an avatar in a virtual world from the user in a physical world creates the feeling of embodiment.

The visual impression of having a virtual body often suffices to induce a sense of ownership [VMR*04], [KKBS16], [BLYK20]. Furthermore, multisensory feedback cues provide means to manipulate the strength of the presence. Several factors contribute to the induction of virtual embodiment [KMKS15]: an accurate human pose estimation is the most significant factor. Generally, to achieve a pose estimation of a real user, studies fulfill two conditions. First, the virtual body appears in the same place as the corresponding real body. Second, the virtual avatar moves synchronously with the real user in real-time. Therefore, a precise full-body pose estimation increases the feeling of embodiment and presence.

In recent decades, following the human pose of a real user in a physical world has been a popular topic. An accurate solution to follow and predict a user's movements is required to increase the feeling of embodiment and presence. The human motion prediction has been extensively studied to imitate the user's body movements in VR. Yet a full-body avatar requires a variety of motions to affect the sense of embodiment for a VR user. Human motion prediction is a task where we anticipate a future motion based on past observation. To this aim, researchers study different approaches such as motion capture, inverse kinematics (IK), and learning-based methods. In this survey, we focus on surveying the proposed approaches for human motion prediction and discuss their contributions.

This survey investigates the most recent and successful studies on human pose estimation in VR. In section 2, we begin our study by introducing the common solutions, including motion capture, Inverse Kinematics (IK), and deep learning-based models. Then, in section 3, we conclude the contributions and applications of each approach. The contribution of our paper is as follows: (a) We provide a comprehensive survey of current 3D human pose estimation methods. (b) We cover the existing methods in three categories: motion capture, Inverse Kinematics (IK), and learning-based frameworks. (c) We evaluate the contributions of each technique.

II. Full-body pose estimation in Virtual Reality

In VR, we describe the pose of the avatar from the body joint positions or rotations of a user in a physical world. Compelling solutions such as motion capture systems, IK, and neural networks predict the body joints' positions or rotations and help to render the corresponding avatar in VR.

A. Motion Capture Techniques

Motion capture (mocap) systems help to track and record human motions at a high resolution. The early form of motion capture is called Rotoscoping techniques. In 1915, using rotoscoping, Max Fleischer recorded an actor's movement with a camera and then moved and drew one by one for each frame [HBY22]. For years, researchers have utilized motion capture data in various fields such as data-driven computer animation, sports, healthcare, games, and the movie industry. We classify the motion capture technologies into optical marker-based [ZCX12], marker-less [ZM13], [SJMS17], [TSLP14] and inertial measurement methods (IMUs) [TLS*17].

A marker-based motion capture system is a technique for extracting specific information from the sensors attached to a human's body. This technique tracks full-body motions very accurately and reliably [CAZT22]. In 2010, Spanlang et al. [SCS10] suggested a marker-based motion capture system to measure the sense of ownership. Furthermore, in 2014, Spanlang et al. [SNB*14] discussed the technical infrastructure necessary to achieve virtual embodiment. They introduce a real-time motion capture, a simple haptics system that integrates physiological and brain electrical activity recordings. In 2021, Pergar et al. [PTX*21] collected a motion capture dataset using 23 Optitrack Prime and 22 cameras to capture 57 retro-reflective markers placed around the subject's body. They conclude that their method generates high-fidelity embodied poses.

An optical markerless system can analyze a target using a camera, infrared, and depth sensor. Many research publications use a low-cost marker-less solution, such as the Kinect device [K*10]. In 2012, Obdrvzalek et al. [OKH*12] presented a real-time algorithm for human pose tracking from vision-based 3D data (markerless) and its application. Their algorithm helps to visualize two remote users with real-time pose evaluation from stereo cameras. In 2017, Usama et al. [IMA*17] proposed a system that can monitor human body parts movement and the accuracy of different yoga poses. They use Microsoft Kinect to detect joint points of the human body in real-time. From those joint points, they calculate various angles to measure a user's accuracy of specific yoga poses.

While one of the main features of VR is to allow the users to look and move in all directions, a marker-less approach is not suitable for many immersive VR experiences [JLT*15], [CGAG19]. The sensor cannot handle occlusion when the limbs are out of sight or when one user stands between the sensor and another user. In 2012, Obdrzalek et al. [OKO*12] compared the Kinect pose estimation with more established techniques relying on motion capture data. They observed the Kinect skeleton tracking struggles with occluding body parts or objects in the scene. In 2015, Gao et al. [GYZD15] leveraged two Kinect sensors to acquire more information about human movements for a full-body presentation to obtain an accurate estimation even when significant occlusion occurs. Eubanks et al. [ELM15], in 2015 suggest using multiple Kinect cameras for a complete 360 tracing. However, with numerous devices, tracking problems can occur due to various IR sources for depth recognition. Moreover, in the marker-less approach usually, the motion data are inaccurate. Hoang et al. [HRVT16], in 2016, measure the significant error rate for lower body tracking of a marker-less approach.

The IMUs are wearable sensors that consist of an accelerometer (measures acceleration), a gyroscope (measures orientation), and a magnetometer (measures magnetism) [VdKR18]. Despite marker-based motion capture, IMUs are used for outdoor shooting due to their high portability. However, when the sensors are used for a long time, the error range of the data increases. Many research publications show that IMUs can provide accurate tracking of the upper and lower body [VdKR18], [KND*18], [PVS*16]. In 2018, Marcard et al. [VMHB*18] proposed a method that combines a single hand-held camera and a set of Inertial Measurement Units (IMUs) attached to the body limbs to estimate accurate 3D poses in the wild. They validate their method with specific 3D poses in challenging sequences, including walking in the city, going upstairs, having coffee, or taking the bus.

With tremendous advances in motion capture devices for several decades, it is now possible to capture a human motion and to animate a virtual character in real-time without modifications. Real-time motion captures and animations allow the use of motion capture devices for VR experiences. Therefore, a character in a virtual world is animated following the motion of the user in a physical world using motion capture devices. In 2022, Ha et al. [HBY22] propose a VR remote collaboration system that uses IMUs to improve immersion in a multi-user environment. They implement a technique for synchronizing the character's size according to the user's body. These motion capture techniques produce immersive experiences in VR. Motion capture devices, however, require a significant space of about at least 10 by 10 m space to set up and an assistant to operate the motion capture devices. Therefore, VR applications with motion capture devices are available, especially in theme parks.

B. *Inverse Kinematics Techniques*

The inverse kinematics (IK) approaches, including numerical solutions and analytical methods, have been extensively studied in robotics. Numerical methods are such as FABRIK [AL11], and the most popular methods are based on the Jacobian matrix [Bus04], [Ken12]. Jacobian matrix methods aim to find a linear approximation of the problem to move the end-effector to the desired target gradually [Bus04]. The numerical methods provide accurate results; however, complex mathematical calculations lead to higher computational costs. In 2016, Jiang et al. [JYF16] presented a novel real-time motion reconstruction and recognition method using the positions and orientations of the user's head and two hands. They divide the whole body into the upper body and the lower body. The upper body reconstruction algorithm is based on inverse kinematics which is more accurate, and the lower body is based on animation blending, which only needs a small number of prepared animations. In 2019, Caserman et al. [CAG19] applied some of the most popular Jacobian IK methods to estimate the full-body pose; the Damped Least Squares (DLS) outperformed the other methods.

Reconstructing the body pose of the avatar in VR using IK methods has been investigated widely. Many studies render the user's motions by tracking the position and orientation of the head, hands, feet, and pelvis. Seele et al. [SMB*17] suggest only tracking the hands with HTC Vive controllers. From a minimum number of trackers [JYF16], considering the user's comfort, the system can estimate the skeleton of the upper body by solving the IK problem. In 2013, Kenwright [Ken13] proposed a realistic, robust, and computationally fast solution to the IK problem using the Gauss-Seidel iterative method for a biped character. In 2017, Tan et al. [THX17] presented a VR-based telepresence system to construct the full-body, including the facial expression and finger movements of the users, using FABRIK and motion capture gloves. In 2018, Parger et al. [PMSS18] presented a heuristic IK method to render the upper body of a VR avatar. They demonstrate that their method can not only be used to increase embodiment but can also support interaction involving arms or shoulders.

C. Learning-based Techniques

Neural Networks play a significant role in extracting and learning features from extensive, high-dimensional datasets and the network's structure determines the features' quality. To predict human motion from sparse joint information in VR, researchers have investigated two approaches in terms of significant network structures: Recurrent Neural Networks (RNN)-based methods and Convolutional Networks (CN)-based methods.

RNN-based methods such as Long Short-Term Memory (LSTM) [HS97] and Gated Recurrent Units [CGCB14] have been demonstrated to be successful for sequence prediction tasks [Gra13]. In 2018, Huang et al. [HKA*18] developed a bi-directional RNN to learn the temporal pose priors and reconstructed the human pose from 6 IMUs worn on the body. They first trained the network with synthetic IMU data and then fine-tuned it with a real dataset. In 2018, Deep Motion [LH18] introduced a 3-point tracking system (HMD and controllers) with a physics-based character model, trained to satisfy physical and kinematic constraints using deep reinforcement learning. The resulting motion may take unnatural steps, different from the user's actual motion. In 2019, Lin [LO19] presented a Temporal IK model using an LSTM-based method from HMD and controllers. Their solution produces natural-looking motions for the upper body. However, it fails to predict the movement of the legs accurately. In 2021, Bashirov et al. [BII*21] present an RNN model for real-time RGBD-based estimation of 3D human pose. They use a parametric 3D deformable human mesh model (SMPL-X) to estimate the parameters for the body pose, hands pose, and facial expression from the Kinect Azure RGB-D camera in real-time. These works show that RNN models can learn the dependencies between spatially correlated data such as human joints‘ rotation.

CNN-based methods have their inherent advantage in capturing spatial dependencies. In 2016, Linna et al. [LKR16] presented a method for real-time multi-person human pose estimation from video by utilizing convolutional neural networks. Mehta et al. [MSS*17] combine a fully convolutional CNN that regresses 2D and 3D joint positions and a kinematic skeleton fitting method, producing a real-time temporally stable 3D reconstruction of the motion. Temporal Convolutional Networks (TCNs) are more acceptable for solving sequential problems [SSCC19]. Therefore, In 2020, Zang et al. [ZPK21] propose a novel approach named Motion Prediction Network (MoPredNet), including three modules: Deformable Spatio-Temporal Convolution Network (DSTCN), a sequence mask module, and a parameter generation module to predict a few-short human motion. Their model MoPredNet can select the correlated motion frames of past motion to provide explicit guidance for long-term pose prediction. In 2020, Liu et al. [LYL*20] designed Trajectory-CNN [15] to predict future poses. The proposed CNN captured spatio-temporal dependencies of the data and predicted the following poses.

III. Conclusion

A survey of the human motion prediction methods for VR avatars is presented in this paper. We review motion capture approaches from marker-based, marker-less, and IMUs. In addition, we discuss IK solutions divided into analytical and numerical solvers. Next, we discuss the recent proposed learning-based models that render a sequence of human poses.

A motion capture system using markers is especially suitable when developing multiplayer VR experiences. Due to a large number of markers and cameras, multiple users will not disturb the tracking of one another. Indeed, the state-of-the-art research also revealed that most VR applications enable multiplayer mode using marker-based systems [CGAG19]. A technology using multiple markers and cameras provides good tracking results even under partial occlusions. Thus, the user can freely move in a setup room. Although the captured data has high accuracy, the process is time-consuming and expensive [PSAS13] because it can only be shot in a fixed indoor studio [HBY22]. Motion capture suits with retro-reflective markers are accurate. However, tight suits often create discomfort [LSB*17], [HBY22]. Furthermore, a specialist is required to process the complicated setup of the motion capture studio such as hardware setup and careful calibration.

A markerless motion capture system is owing to its low cost, flexibility, high precision, and fast response. On the contrary, the image processing method is inaccurate in the marker-less motion capture systems and IMUs because there is no clear reference point. Moreover, when the user notices these inaccuracies from the first-person perspective, this can decrease immersion and reduce the sense of body ownership. In addition, a marker-less solution necessitates many manual initialization steps, such as room scaling, and initial position estimation. Therefore, a specialist needs to take care of the system. This technology often cannot handle occlusion.

Previous studies on upper body pose estimation of a VR avatar with IK approaches achieve accurate results [PMSS18], [GYLW21]. To estimate an avatar poses in VR, IK-based techniques are easy-to-apply, stable, and fast compared to the motion capture technologies. IK solutions avoid any occlusion issues without any additional costs. Numerical solvers of IK solutions create natural movements at the cost of an iterative process that requires more computational effort than an analytical solution. In addition, other end targets can easily be added to the numerical optimization. Analytical solvers of IK solutions are fast to compute without any convergence problems that numerical solvers have. Still, it might be challenging to implement constraints or multiple tasks to influence which of all possible solutions is computed. However, IK failed to provide accurate lower body pose prediction of a VR avatar from only HMD and controllers.

Machine learning and deep learning techniques are top-rated as they can provide high-quality, specialized solutions depending on the quality and amount of the training dataset. Pose estimation of a VR avatar from learning-based approaches is easy to apply. They are fast in a real-time pose prediction system; however, the training time is time-consuming. The disadvantages of learning-based methods are the expensive data acquisition, wrong solutions for poses that are not covered well in the training data, and errors caused by low-quality data [PMSS18]. Furthermore, full-body pose estimation from a learning-based approach is achievable. However, it requires more development in the case of complex lower body motions.

Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT). (No. 2021R1A2C1014210)

References

[AL11] Aristidou A., Lasenby J.: Fabrik: A fast, iterative solver for the inverse kinematics problem. *Graphical Models 73*, 5 (2011), 243–260.

[AMTK83] Akin D. L., Minsky M., Thiel E., Kurtzman C.: *Space Applications of Automation, Robotics and Machine Intelligence Systems (ARAMIS), phase 2. Volume 2: Telepresence project applications*. Tech. rep., NASA, 1983.

[BII*21] Bashirov R., Ianina A., Iskakov K., Kononenko Y., Strizhkova V., Lempitsky V., Vakhitov A.: Real-time rgbd-based extended body pose estimation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2021), pp. 2807–2816.

[BLYK20] Burin D., Liu Y., Yamaya N., Kawashima R.: Virtual training leads to physical, cognitive and neural benefits in healthy adults. *Neuroimage 222* (2020), 117297.

[Bus04] Buss S. R.: Introduction to inverse kinematics with jacobian transpose, pseudoinverse and damped least squares methods. *IEEE Journal of Robotics and Automation 17*, 1-19 (2004), 16.

[CAG19] Caserman P., Achenbach P., Göbel S.: Analysis of inverse kinematics solutions for full-body reconstruction in virtual reality. In *2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH)* (2019), IEEE, pp. 1–8.

[CAZT22] Chatzitofis A., Albanis G., Zioulis N., Thermos S.: A low-cost & real-time motion capture system. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 21453–21458.

[CGAG19] Caserman P., Garcia-Agundez A., Göbel S.: A survey of full-body motion reconstruction in immersive virtual reality applications. *IEEE transactions on visualization and computer graphics 26*, 10 (2019), 3089–3108.

[CGCB14] Chung J., Gulcehre C., Cho K., Bengio Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).

[ELM15] Eubanks J. C., Lai C., McMahan R. P.: Portable virtual reality: Inertial measurements and biomechanics. In *2015 IEEE 1st Workshop on Everyday Virtual Reality (WEVR)* (2015), IEEE, pp. 1–4.

[Gra13] Graves A.: Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850* (2013).

[GYLW21] Gu L., Yin L., Li J., Wu D.: Human upper-body inverse kinematics for consumer grade virtual reality. In *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)* (2021), vol. 5, IEEE, pp. 2092–2095.

[GYZD15] Gao Z., Yu Y., Zhou Y., Du S.: Leveraging two kinect sensors for accurate full-body motion capture. *Sensors 15*, 9 (2015), 24297–24317.

[HBY22] Ha E., Byeon G., Yu S.: Full-body motion capture-based virtual reality multi-remote collaboration system. *Applied Sciences 12*, 12 (2022), 5862.

[HKA*18] Huang Y., Kaufmann M., Aksan E., Black M. J., Hilliges O., Pons-Moll G.: Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time. *ACM Transactions on Graphics (TOG) 37*, 6 (2018), 1–15.

[HRVT16] Hoang T. N., Reinoso M., Vetere F., Tanin E.: Onebody: remote posture guidance system using first person view in virtual environment. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (2016), pp. 1–10.

[HS97] Hochreiter S., Schmidhuber J.: Long short-term memory. *Neural computation 9*, 8 (1997), 1735–1780.

[IMA*17] Islam M. U., Mahmud H., Ashraf F. B., Hossain I., Hasan M. K.: Yoga posture recognition by detecting human joint points in real time using microsoft kinect. In *2017 IEEE Region 10 humanitarian technology conference (R10-HTC)* (2017), IEEE, pp. 668–673.

[JLT*15] Joo H., Liu H., Tan L., Gui L., Nabbe B., Matthews I., Kanade T., Nobuhara S., Sheikh Y.: Panoptic studio: A massively multiview system for social motion capture. In *Proceedings of the IEEE International Conference on Computer Vision* (2015), pp. 3334–3342.

[JYF16] Jiang F., Yang X., Feng L.: Real-time full-body motion reconstruction and recognition for off-the-shelf vr devices. In *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry-Volume 1* (2016), p. 309–318.

[K*10] Kar A., et al.: Skeletal tracking using microsoft kinect. *Methodology 1*, 1 (2010), 11.

[Ken12] Kenwright B.: Real-time character inverse kinematics using the gauss-seidel iterative approximation method. In *International Conference on Creative Content Technologies* (2012), vol. 4, pp. 63–68.

[Ken13] Kenwright B.: Real-time reactive biped characters. In *Transactions on Computational Science XVIII*. Springer, 2013, pp. 155–171.

[KGS12] Kilteni K., Groten R., Slater M.: The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments 21*, 4 (2012), 373–387.

[KKBS16] Kokkinara E., Kilteni K., Blom K. J., Slater M.: First person perspective of seated participants over a walking virtual body leads to illusory agency over the walking. *Scientific reports 6*, 1 (2016), 1–11.

[KMKS15] Kilteni K., Maselli A., Kording K. P., Slater M.: Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in human neuroscience 9* (2015), 141.

[KND*18] Kishore S., Navarro X., Dominguez E., De La Peña N., Slater M.: Beaming into the news: a system for and case study of tele-immersive journalism. *IEEE computer graphics and applications 38*, 2 (2018), 89–101.

[LH18] Liu L., Hodgins J.: Learning basketball dribbling skills using trajectory optimization and deep reinforcement learning. *ACM Transactions on Graphics (TOG) 37*, 4 (2018), 1–14.

[LKR16] Linna M., Kannala J., Rahtu E.: Real-time human pose estimation from video with convolutional neural networks. *arXiv preprint arXiv:1609.07420* (2016).

[LO19] Lin J., O'Brien J.: Temporal ik: Data-driven pose estimation for virtual realiity.

[LSB*17] Leoncini P., Sikorski B., Baraniello V., Martone F., Luongo C., Guida M.: Multiple nui device approach to full body tracking for collaborative virtual environments. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics* (2017), Springer, pp. 131–147.

[LYL*20] Liu X., Yin J., Liu J., Ding P., Liu J., Liu H.: Trajectorycnn: a new spatio-temporal feature learning network for human motion prediction. *IEEE Transactions on Circuits and Systems for Video Technology 31*, 6 (2020), 2133–2146.

[MSS*17] MEHTA D., SRIDHAR S., SOTNYCHENKO O., RHODIN H., SHAFIEI M., SEIDEL H.-P., XU W., CASAS D., THEOBALT C.: Vnect: Real-time 3d human pose estimation with a single rgb camera. *Acm transactions on graphics (tog) 36*, 4 (2017), 1–14.

[OKH*12] OBDRŽÁLEK Š., KURILLO G., HAN J., ABRESCH T., BAJCSY R.: Real-time human pose detection and tracking for tele-rehabilitation in virtual reality. In *Medicine Meets Virtual Reality 19*. IOS Press, 2012, pp. 320–324.

[OKO*12] OBDRŽÁLEK Š., KURILLO G., OFLI F., BAJCSY R., SETO E., JIMISON H., PAVEL M.: Accuracy and robustness of kinect pose estimation in the context of coaching of elderly population. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2012), IEEE, pp. 1188–1193.

[PMSS18] PARGER M., MUELLER J. H., SCHMALSTIEG D., STEINBERGER M.: Human upper-body inverse kinematics for increased embodiment in consumer-grade virtual reality. In *Proceedings of the 24th ACM symposium on virtual reality software and technology* (2018), pp. 1–10.

[PSAS13] PECK T. C., SEINFELD S., AGLIOTI S. M., SLATER M.: Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition 22*, 3 (2013), 779–787.

[PTX*21] PARGER M., TANG C., XU Y., TWIGG C. D., TAO L., LI Y., WANG R., STEINBERGER M.: Unoc: Understanding occlusion for embodied presence in virtual reality. *IEEE Transactions on Visualization and Computer Graphics* (2021).

[PVS*16] PODKOSOVA I., VASYLEVSKA K., SCHOENAUER C., VONACH E., FIKAR P., BRONEDERK E., KAUFMANN H.: Immersivedeck: A large-scale wireless vr system for multiple users. In *2016 IEEE 9th Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS)* (2016), IEEE, pp. 1–7.

[SBW17] SKARBEZ R., BROOKS JR F. P., WHITTON M. C.: A survey of presence and related concepts. *ACM Computing Surveys (CSUR) 50*, 6 (2017), 1–39.

[SCS10] SPANLANG B., COROMINAS D., SLATER M.: A virtual whole body system. In *CHI Whole Body Interaction Workshop* (2010).

[SJMS17] SIMON T., JOO H., MATTHEWS I., SHEIKH Y.: Hand keypoint detection in single images using multiview bootstrapping. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (2017), pp. 1145–1153.

[Sla18] SLATER M.: Immersion and the illusion of presence in virtual reality. *British Journal of Psychology 109*, 3 (2018), 431–433.

[SMB*17] SEELE S., MISZTAL S., BUHLER H., HERPERS R., SCHILD J.: Here's looking at you anyway! how important is realistic gaze behavior in co-located social virtual reality games? In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (2017), pp. 531–540.

[SNB*14] SPANLANG B., NORMAND J.-M., BORLAND D., KILTENI K., GIANNOPOULOS E., POMÉS A., GONZÁLEZ-FRANCO M., PEREZ-MARCOS D., ARROYO-PALACIOS J., MUNCUNILL X. N., ET AL.: How to build an embodiment lab: achieving body representation illusions in virtual reality. *Frontiers in Robotics and AI 1* (2014), 9.

[SPMESV08] SLATER M., PÉREZ MARCOS D., EHRSSON H., SANCHEZ-VIVES M. V.: Towards a digital body: the virtual arm illusion. *Frontiers in human neuroscience* (2008), 6.

[SSCC19] SHAO J., SHEN H., CAO Q., CHENG X.: Temporal convolutional networks for popularity prediction of messages on social medias. In *China Conference on Information Retrieval* (2019), Springer, pp. 135–147.

[THX17] TAN Z., HU Y., XU K.: Virtual reality based immersive telepresence system for remote conversation and collaboration. In *International Workshop on Next Generation Computer Animation Techniques* (2017), Springer, pp. 234–247.

[TLS*17] TEE K. S., LOW E., SAIM H., ZAKARIA W. N. W., KHIALDIN S. B. M., ISA H., AWAD M., SOON C. F.: A study on the ergonomic assessment in the workplace. In *AIP conference proceedings* (2017), vol. 1883, AIP Publishing LLC, p. 020034.

[TSLP14] TOMPSON J., STEIN M., LECUN Y., PERLIN K.: Real-time continuous pose recovery of human hands using convolutional networks. *ACM Transactions on Graphics (ToG) 33*, 5 (2014), 1–10.

[VdKR18] VAN DER KRUK E., REIJNE M. M.: Accuracy of human motion capture systems for sport applications; state-of-the-art review. *European journal of sport science 18*, 6 (2018), 806–819.

[VMHB*18] VON MARCARD T., HENSCHEL R., BLACK M. J., ROSENHAHN B., PONS-MOLL G.: Recovering accurate 3d human pose in the wild using imus and a moving camera. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 601–617.

[VMR*04] VOGELEY K., MAY M., RITZL A., FALKAI P., ZILLES K., FINK G. R.: Neural correlates of first-person perspective as one constituent of human self-consciousness. *Journal of cognitive neuroscience 16*, 5 (2004), 817–827.

[ZCX12] ZHAO W., CHAI J., XU Y.-Q.: Combining marker-based mocap and rgb-d camera for acquiring high-fidelity hand motion data. In *Proceedings of the ACM SIGGRAPH/eurographics symposium on computer animation* (2012), pp. 33–42.

[ZM13] ZHAO J. Z. W., MIN J.: Robust realtime physically based motion control for human grasping. *ACM SIGGRAPH Asia* (2013).

[ZPK21] ZANG C., PEI M., KONG Y.: Few-shot human motion prediction via learning novel motion dynamics. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence* (2021), pp. 846–852.