Train Me: Exploring Mobile Sports Capture and Replay for Immersive Sports Coaching

Mehran Rastegar Sani
University of Otago
Dunedin, New Zealand

Tobias Langlotz
University of Otago
Dunedin, New Zealand

Stefanie Zollmann
University of Otago
Dunedin, New Zealand

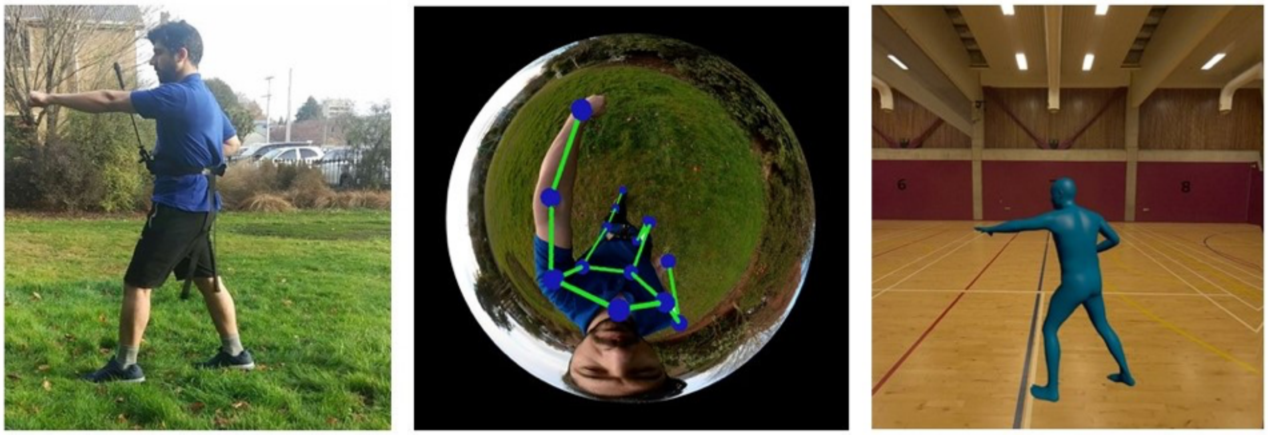


Figure 1: Mobile Sports Capture and Replay concept. (Left) Capturing setup with a body-attached camera recording the instructor's karate movements. (Middle), captured footage with extracted skeleton data overlaid. (Right) Visualizing the extracted avatar in an Augmented Reality replay for training purposes.

ABSTRACT

In recent years, a wide variety of instructional materials and applications have been developed to enhance athletes' learning in different sports. The amount of instructional videos, in particular, has increased significantly, though these often impose a high cognitive load as users must map displayed instructions to their actions and body movements. Mobile Augmented Reality (AR) interfaces can reduce this burden by presenting instructional information directly where it is needed. This paper explores a mobile sports capture and replay approach for immersive self-training, aiming to help users improve their skills without needing coaches on-site. We investigate different capturing methods, including an Egocentric capturing method, to enhance instructor mobility and flexibility. Using the captured data, we visualize sports training instructions in an immersive 3D environment on an AR headset. We propose three visualization methods and create a first prototype that allows us to explore the approach's feasibility across different sports.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **Mobile devices**.

MOBILEHCI Adjunct '24, September 30-October 3, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-0506-9/24/09
https://doi.org/10.1145/3640471.3680239

KEYWORDS

Augmented Reality, Sport self-training, Virtual Reality, Self-Coaching

ACM Reference Format:
Mehran Rastegar Sani, Tobias Langlotz, and Stefanie Zollmann. 2024. Train Me: Exploring Mobile Sports Capture and Replay for Immersive Sports Coaching. In *26th International Conference on Mobile Human-Computer Interaction (MOBILEHCI Adjunct '24), September 30-October 3, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3640471.3680239

1 INTRODUCTION

In sports, beginners, as well as recreational athletes, often find it difficult to develop their skills and abilities without the instruction and guidance of experienced coaches. As a result, virtual fitness classes with video-on-demand and online classes have gained some interest in recent years [5, 22]. In addition, video-based training and Virtual Reality (VR)-based training have shown some potential for athletes, for instance, in soccer and basketball [2, 4]. However, these options come with the disadvantage of a disconnection between the current action and movement of an athlete and the displayed content. Often, virtual fitness classes are shown from a third-person perspective, and VR-based tutorials have to be watched before a training session, creating a gap between instructions and user action and creating an additional mental workload. Mobile Augmented Reality (AR) interfaces have the potential to bridge that gap by presenting instructional information directly where it is needed and when it is needed [1].

In this project, we aim to investigate the suitability of AR for sports tutorials, self-training, and self-coaching. The goal is to introduce a concept that helps users improve their skills in different sports without on-site coaches or attending classes. For this purpose, we designed a framework that captures an expert providing instructions for skill improvement in different sports, such as how to swing a golf or how to throw a basketball. Our framework extracts the motion and position of each person in these videos and creates an avatar that animates the movements. The avatar with the movement pattern is then replayed in AR using a head-mounted display (HMD). This allows users to see the differences between the posture of the instructor and their own movements, helping them to improve their skills.

While there is some previous work on using remote coaching systems with annotation tools to assist trainers in an immersive environment [16, 31], they often require specific capturing setups or have a fixed capturing area making them less suitable for mobile activities such as sports coaching. As a result, our goal is to develop a comprehensive system capable of supporting a range of sports. This work described the concept and provides initial feedback including:

- Integrating different methods for casual mobile capture of motion capturing data for later replaying in mobile immersive Virtual Reality.
- Exploration of a set of AR visualization methods.
- Preliminary feedback on the feasibility and applicability of the proposed method across 7 different types of sports together with preliminary feedback from a domain expert.

2 RELATED WORK

Our research covers aspects of motion capture and pose estimation, body-worn capturing, and sports training and performance analysis.

2.1 Sports Training and Performance Analysis

Recent advancements in video feedback and augmented reality systems have significantly advanced sports training and athletic performance. Scientists and engineers have developed systems for sports like rock climbing, golf, volleyball, and basketball, using detailed visual analysis and immersive virtual environments to enhance skills. Kosmalla et al. [11] introduced a video recording and replay system that supports climbers by providing an augmented third-person view with an overlay of an experienced climber, shown on Google Glass or projected on the climbing wall. Liao et al. [13] developed a golf swing analysis method that compares 3D poses to help users understand the differences between their swings and those of professionals. Sato et al. [19] designed a system to help volleyball beginners predict ball landing positions by using a projector.

Yu et al. [30] utilized visual feedback to assist amateur tennis players in improving their skills, employing motion-tracking sensors to capture shot data during practice and instructor demonstrations. Their findings indicated that beginners could enhance abilities such as timing shots, controlling wrist and racket angle, and maintaining balance by comparing their performance metrics with those of experts. Augmented Coach [27] implemented an immersive annotation system for remote sports coaching, reconstructing athletes in 3D to enhance spatio-temporal understanding through HMD visualizations, validated through user studies involving coaches and athletes. Meanwhile, Liao et al. [14] presented AI Coach, a self-training system for golf swing motor skills, integrating motion synchronization, neural network-driven discrepancy detection, and real-time motion visualization via HMD feedback, demonstrating superior training outcomes in enhancing motor skill acquisition through precise and actionable feedback.

In summary, despite significant recent advancements in self-training, there are still limitations and gaps that this work aims to address. Many researchers [3, 7, 21, 31] commonly use sensors like IMUs to capture the motion of athletes. However, such setups require special facilities, which affects the generalizability of the system. Therefore, since the target audience is the general public who cannot afford private coaches, this work decided to use computer vision-based methods (e.g., OpenPose) for estimating the athletes' motions and positions from videos. Furthermore, most prior research [15, 17, 21, 27] has focused on creating models tailored to a single sport that restrict athletes to a limited range of motion and make applying their methods across different sports impractical. This customization limited the ability of their systems to handle multiple sports. As a result, developing a comprehensive system capable of supporting a range of sports remains a challenge in this field.

To address these challenges, we propose a capturing and replay concept that provides flexibility for mobile capturing and mobile replay. The concept contains three components: a capturing component, a pose estimation component and a live AR replay visualization component.

The capturing component captures video footage from an expert trainer. Then, the pose estimation component analyzes these videos to estimate the motions and body poses of the individuals captured. Finally, the ARReplay visualization component visualizes the extracted avatars representing the learner and the expert on a head-mounted display (HMD). By displaying these avatars side-by-side, learners can observe the differences between their form and the expert's form, enabling them to identify areas for improvement. The final goal is to help ordinary people enhance their sports skills without the need for personal coaches or attending formal classes. We will discuss each part in the following with details.

2.2 Motion Capture and Pose Estimation

Researchers have used IMUs and pose estimation approaches to capture and visualize users' skeletons and movements, creating immersive virtual avatars. Cha et al. [3] developed a real-time egocentric 3D human body motion reconstruction system by combining visual and inertial information from egocentric views and IMUs to tackle challenges like inconsistent limb visibility and pose ambiguity. Zhang et al. [31] proposed a 3D reconstruction model, using IMUs to estimate the client's 3D pose and mapping key points on a Skinned Multi-Person Linear Model (SMPL) 3D human shape to generate a mesh. They also used AR HMD footage to reconstruct the client's appearance and combined it with IMU-estimated poses to create a novel view of the user, displaying it in a VR HMD for teleconsultation.

Studies have employed motion capture systems with 360° cameras and OpenPose to enhance various applications. Yu et al. [29] captured people in an environment to generate virtual skeletons for motion generation. Besides, Xu et al. [28] used these technologies to improve network collaboration in the Metaverse by capturing multiple people's actions simultaneously.

Wang et al. [24, 25] introduced several innovative approaches for egocentric 3D human pose estimation using head-mounted cameras. They have developed a range of methods addressing various challenges in this domain, such as occlusions, fisheye distortion, and real-time processing. One notable contribution is the EgoPW dataset [25], which leverages external cameras for weak supervision, enabling accurate pseudo-label generation for training an egocentric pose estimation network that bridges the synthetic-real domain gap, significantly outperforming previous methods. Additionally, they proposed a spatiotemporal optimization framework utilizing a convolutional VAE and a global pose optimizer for stable egocentric pose estimation [26].

2.3 Body-worn Capturing

The potential of body-worn cameras has been demonstrated by capturing users for different applications and, in particular, for telepresence. For instance, the JackIn framework by Kasahara et al. connects two individuals through first-person view video streaming using a head-mounted display and camera [9]. This system allows a person (the "Body") to share their activities with another (the "Ghost"), who can then provide guidance or assistance. Applications include cooking lessons, craft-work education, and live event sharing such as sports [10]. An out-of-body view feature lets the Ghost virtually navigate the body's surroundings, enhancing spatial understanding and communication. Huang et al. create how-to videos by using a head-mounted 360° action camera to capture immersive first-person perspectives [6]. Viewers can watch these videos through a VR headset, gaining an eye-level viewpoint. Evaluations showed that immersivePOV reduces cognitive load and facilitates better task learning compared to traditional third-person perspectives, proving effective for both learners and content creators. Kratz et al. used an attached camera to allow users to explore remote locations in their polling system [12].

Researchers have also proposed different types of setups for egocentric capture. Tome et al. [23] placed a fish-eye camera on the rim of a VR headset, pointing downwards towards the user's body, with the camera lens being only 2 cm away from the user's face, to capture an egocentric/first-person view of the user's body. Kang et al. [8] presented Ego3DPose, a highly accurate binocular egocentric 3D pose reconstruction system. The binocular setup offers practicality and usefulness in various applications. Rhodin et al. [18] introduced a markerless motion capture system using a lightweight stereo pair of fisheye cameras attached to a helmet or virtual reality headset. The cameras capture the user's full-body skeleton pose from an egocentric perspective.

These approaches demonstrate how body-worn cameras can transform interactive experiences by providing immersive and informative first-person perspectives. These advancements are beneficial in education, training, and daily activities, enhancing both understanding and learning.

3 CONCEPT DESIGN

3.1 Capturing

To enhance the generalizability of the system, we integrated three different video capture methods:

(1) **Standard Video**: This involves utilizing readily available video sources such as YouTube videos or footage recorded on a mobile phone. This approach allows users to leverage instructional videos from platforms like YouTube, featuring experts demonstrating tips and techniques for various sports, as input for the system.
(2) **Exocentric 360° Video**: In this setup, 360-degree cameras mounted on tripods are employed to capture the instructions (Figure 3, left, middle left). The 360-degree view allows the instructor to move freely within the environment without concern for going out of the camera's field of view.
(3) **Egocentric 360° Footage**: For some activities like skateboarding, climbing, or surfing, where using a tripod-mounted camera is impractical, we use an alternative capture setup. We can consider using a camera attached to the client's cap to accomplish this. However, for our first prototype, we used 360 cameras mounted in front of the user's face, with a downward-facing view capturing the person's body (Figure 3, middle right, right). This configuration enables instructors to capture footage of themselves covering a large area of movement without the need to occupy their hands.

3.2 Pose and Shape Estimation

To estimate the pose and shape of the human subjects in the videos, I utilized Wham [20], which is the leading state-of-the-art method in this area. This algorithm provides the motion and key joint positions of the person in standard videos, which is suitable for the first two setups (Simple Video and Exocentric 360 Footage). However, since the Wham algorithm was not originally developed to handle egocentric-vision footage, it failed to accurately extract the global movement of the person in the real-world environment. As a result, we decided to explore alternative pose and shape estimation methods. For this purpose, we employed the method proposed in [24], a human pose estimation approach specifically designed for fisheye camera videos. The algorithm is developed to operate on spherical videos, similar to the view of our Egocentric setup (Figure 3, right).

3.3 Visualization

Replaying the extracted pose and avatar of a person on a head-mounted display (HMD) can help learners experience and observe a tutorial in an immersive, three-dimensional environment. This way, they can view and analyze the avatar and form of both the trainer and themselves from different angles, allowing them to correct their form as needed. We developed three different visualization methods and investigated them to determine their feasibility for AR self coaching:

(1) **External-View Visualization**: This method displays the extracted avatars of the learner and the expert trainer side-by-side in the virtual environment (Figure 4, a). We hypothesize that with this method, the learner can move around and

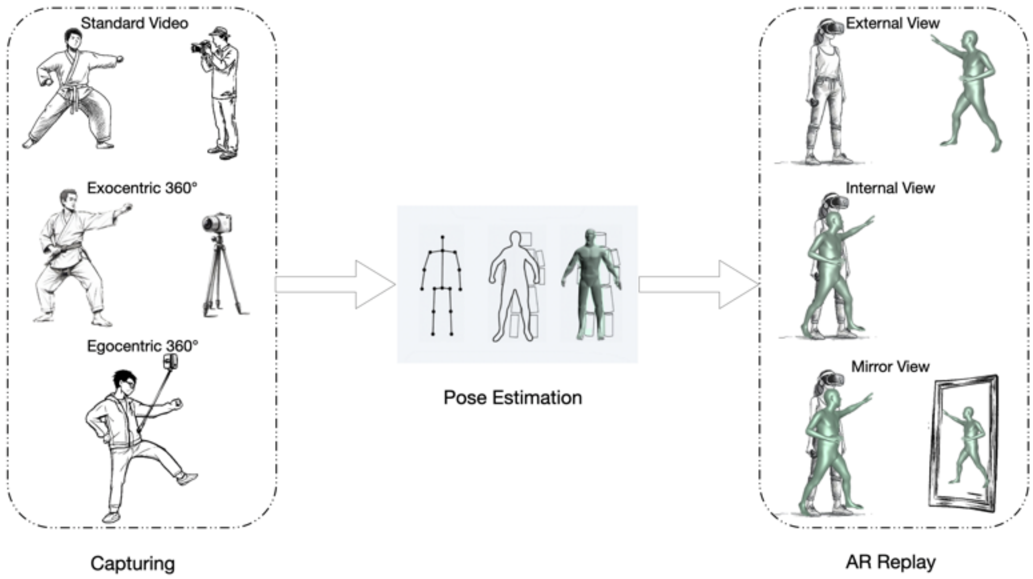


Figure 2: Train Me concept. Capturing component uses different video inputs to capture instructor's movements. Pose Estimation component extracts instructor's body posture and creates 3D avatar. AR Replay component overlays instructor's avatar on the view of the user. (DALL-E was used for generating subparts of this figure.)

observe the avatars from various viewpoints, facilitating a comparison of their form with the trainer's in a 3D space.

(2) **Internal-View Visualization**: For this approach, we align the extracted avatar of the professional trainer with the learner's body, providing a down-facing(first-person) view of the movement. The avatar moves in sync with the learner's body as they wear the HMD (Figure 4, b). We hypothesize that for activities like golf or baseball swings, having an internal view of the expert's body can be beneficial for learners to imitate the same movements accurately.

(3) **Mirror Visualization**: This method combines the Internal-View with a mirror reflection, enabling the learner to see the entire avatar in a mirror while simultaneously experiencing it from a first-person viewpoint. We hypothesize that this visualization is particularly suitable for sports that occur in a fixed location, such as golf swings or basketball free throws (Figure 4, c).

3.4 Implementation

For capturing exocentric (third-person) footage, we use an Insta360 One X2, a 360-degree camera mounted on a tripod. Additionally, we utilized the Insta360 "Backpack Mount" attached to the client's chest along with the One X2 camera to capture egocentric (first-person) footage.

To estimate poses and extract avatars, we leveraged two state-of-the-art open-source models: the "WHAM" algorithm [20] for the exocentric footage, and the "Egocentric Whole Body" [24] for the egocentric videos. To generate avatar representations from the estimated poses, we use a custom Python code converter.

For visualizing the extracted avatars, we utilized Unity 2023 as the development platform.We then use an Oculus Quest 3 headset in passthrough mode to display the avatars within the real-world environment of the user.

4 FINDINGS/LESSONS LEARNED

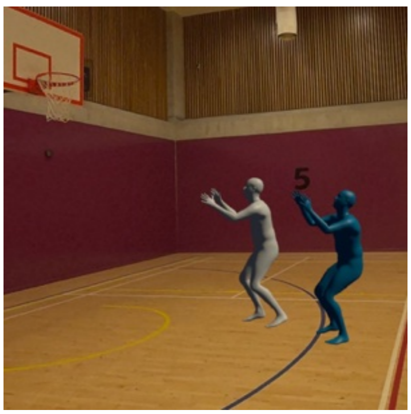
4.1 Preliminary Feasibility Analysis

We explored the feasibility and applicability of our concept across seven different sports: golf, basketball, Gym exercise (e.g. squats and push-ups), and sports activities that need more mobility, like skateboarding, soccer, and Karate kata. Our preliminary findings (Table 1) show that the system works sufficiently well and could be beneficial for sports like soccer, skateboarding, and squats that require a lot of movement and positioning of the entire body. However, it faces challenges detecting correct body position and movement in certain situations where views are obstructed, or body parts go above the head.

For example, the system missed tracking arm movements when they were raised above the head, which happens frequently in overhead sports like basketball (free throws) and kata routines with high arm movements. In other cases like the golf swing or push-up exercises, the system struggled to accurately detect proper lower body form and angles like knee bend when the legs were obstructed from view by the upper body parts like arms.

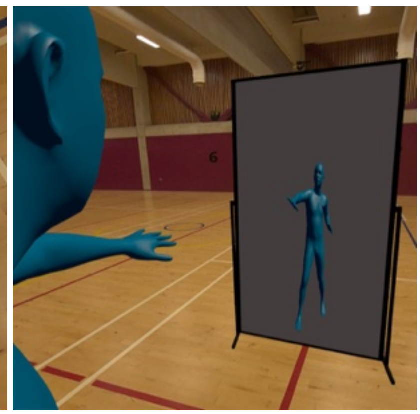
Despite the issues with obstructed views, we found the system was useful overall for capturing casual sports, and see potential in particularly for those that require more mobility, such as surfing and skateboarding, where a fixed capturing setup may not be practical and action cameras are already used. Further refinements may be needed to improve the visibility and tracking of the entire body.

Figure 3: Settings used for capturing Exocentric 360° and Egocentric 360° footage. Left) Exocentric Capturing Setup. Middle Left) Captured exocentric view. Middle Right) Egocentric Capturing Setup. Right) Egocentric Capturing View.



(a) External-View Visualization

(b) Internal-View Visualization



(c) Mirror Visualization

Figure 4: Example of visualization modes: (a) Representing both the trainer and learner avatars side by side to make the comparison easier for the learner. (b) Putting the trainer's avatar into the learner's actual body, allowing them to experience the trainer's first-person view. (c) Combining the internal visualization with a mirror to facilitate following the avatar's movements.

4.2 Preliminary Expert Feedback

To collect preliminary feedback on the proposed approach, we demonstrated the concept to a sports and exercise expert (academic from the Department of *removed for review*) who also has a background in motion capture and motion analysis for sports. We used the use case "Karate kata" as an example and asked the expert user to use our prototypical implementation. We showed them the three visualization methods on the AR HMD (External View, Internal View, Mirror View) each for approximately 2 minutes. After the trial, we collected feedback in a semi-structured interview.

In the mode **External-View**, the user attempted to imitate the movement sequences, which appeared visually similar. However, they mentioned it being tricky and confusing to determine which arm to move.

In the **Internal-View** mode, the user also tried to imitate the movements. This mode was noted as the favorite option of the user.

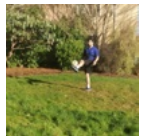
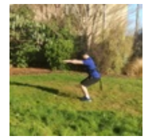
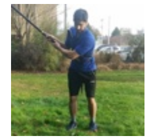
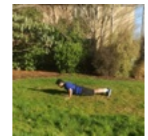
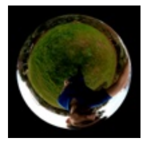
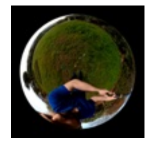
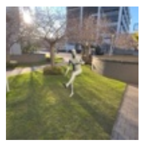
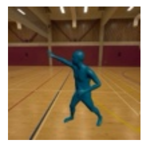
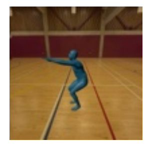
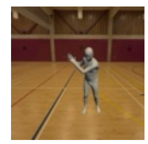
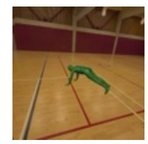
In **Mirror Visualization**, the user imitated the movements and provided feedback. They found the combination of Mode Ego and Mode Mirror quite helpful.

Despite challenges with the accuracy of the captures, the expert user mentioned the potential for beginners. In particular, they mentioned the potential of gamification and described the need for support for engagement vs correct techniques. They also mentioned room for improvement with regard to the quality, such as realism and placement. For instance, the avatar appeared to be floating on the ground. We are currently improving the prototype based on the expert feedback to prepare a full user study.

5 CONCLUSIONS AND FUTURE WORK

The main objective of this work is to create a first system that helps individuals enhance their sports skills without needing dedicated coaches. To make the system widely applicable, we integrated three different capture methods: Simple Video (e.g. YouTube tutorials), Exocentric 360° Footage (for improved and as a baseline condition), and Egocentric 360° Footage (ideal for sports requiring greater mobility, such as climbing and surfing). Additionally, we investigated a set of visualization techniques to understand their respective advantages and disadvantages.

Table 1: Summary of Preliminary Findings

	Basketball	Skateboarding	Soccer	Kata	Squat	Golf	Push-up
Mobile Setup							
Captured Data							
AR Replay							
Findings	Facing difficulties in detecting the arms' form when they move higher than the camera's view.	Suitable for capturing skateboarding movement and techniques.	Suitable for capturing selected techniques in soccer (e.g., kick-ups). Headers and fast movements pose challenges.	Suitable for performing kata in martial arts excluding overhead arm movements.	Suitable for capturing the human body form during squat exercises.	Cannot detect the correct lower body poses (e.g. knee angle) when it is occluded by the upper body (e.g. arms).	While the upper body form can be detected, the lower body is often not visible in the camera view.

In future work, we plan to conduct a user study to gather feedback and analyze users' experiences with the system. Based on the insights gained, we can refine and improve the system to better align with user needs and priorities. Subsequently, we aim to develop advanced techniques for automatically identifying and recognizing specific actions captured within the videos. By implementing such action recognition capabilities, the system could intelligently extract only the most relevant video segments of interest to each user.

ACKNOWLEDGMENTS

This work is supported by an MBIE Endeavour Smart Ideas grant. We thank Dr. Peter Lamb for feedback on the research prototype.

REFERENCES

[1] Mark Billinghurst, Adrian Clark, Gun Lee, et al. 2015. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction* 8, 2-3 (2015), 73–272.

[2] Pierre-Michel Bernier Caleb Pagé and Maxime Trempe. 2019. Using video simulations and virtual reality to improve decision-making skills in basketball. *Journal of Sports Sciences* 37, 21 (2019), 2403–2410. https://doi.org/10.1080/02640414.2019.1638193 arXiv:https://doi.org/10.1080/02640414.2019.1638193 PMID: 31280685.

[3] Young-Woon Cha, Husam Shaik, Qian Zhang, Fan Feng, Andrei State, Adrian Ilie, and Henry Fuchs. 2021. Mobile. Egocentric human body motion reconstruction using only eyeglasses-mounted cameras and a few body-worn inertial sensors. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 616–625.

[4] Leonardo S. Fortes, Sebastião S. Almeida, Gibson M. Praça, José R.A. Nascimento-Júnior, Dalton Lima-Junior, Bruno Teixeira Barbosa, and Maria E.C. Ferreira. 2021. Virtual reality promotes greater improvements than video-stimulation screen on perceptual-cognitive skills in young soccer athletes. *Human Movement Science* 79 (2021), 102856. https://doi.org/10.1016/j.humov.2021.102856

[5] Jiajing Guo and Susan R. Fussell. 2022. "It's Great to Exercise Together on Zoom!": Understanding the Practices and Challenges of Live Stream Group Fitness Classes. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 71 (apr 2022), 28 pages. https://doi.org/10.1145/3512918

[6] Kevin Huang, Jiannan Li, Mauricio Sousa, and Tovi Grossman. 2022. immersivePOV: Filming How-To Videos with a Head-Mounted 360° Action Camera. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>New Orleans</city>, <state>LA</state>, <country>USA</country>, </conf-loc>) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 91, 13 pages. https://doi.org/10.1145/3491102.3517468

[7] Atsuki Ikeda, Dong-Hyun Hwang, Hideki Koike, Gerd Bruder, Shunsuke Yoshimoto, and Sue Cobb. 2018. AR based Self-sports Learning System using Decayed Dynamic TimeWarping Algorithm.. In *ICAT-EGVE*. 171–174.

[8] Taeho Kang, Kyungjin Lee, Jinrui Zhang, and Youngki Lee. 2023. Ego3dpose: Capturing 3d cues from binocular egocentric views. In *SIGGRAPH Asia 2023 Conference Papers*. 1–10.

[9] Shunichi Kasahara and Jun Rekimoto. 2014. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference* (Kobe, Japan) *(AH '14)*. Association for Computing Machinery, New York, NY, USA, Article 46, 8 pages. https://doi.org/10.1145/2582051.2582097

[10] Shunichi Kasahara and Jun Rekimoto. 2015. JackIn head: an immersive human-human telepresence system. In *SIGGRAPH Asia 2015 Emerging Technologies* (Kobe, Japan) *(SA '15)*. Association for Computing Machinery, New York, NY, USA, Article 14, 3 pages. https://doi.org/10.1145/2818466.2818486

[11] Felix Kosmalla, Florian Daiber, Frederik Wiehr, and Antonio Krüger. 2017. Climbvis: Investigating in-situ visualizations for understanding climbing movements by demonstration. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. 270–279.

[12] Sven Kratz, Don Kimber, Weiqing Su, Gwen Gordon, and Don Severns. 2014. Polly: " being there" through the parrot and a guide. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. 625–630.

[13] Chen-Chieh Liao, Dong-Hyun Hwang, and Hideki Koike. 2022. Ai golf: Golf swing analysis tool for self-training. *IEEE Access* 10 (2022), 106286–106295.

[14] Chen-Chieh Liao, Dong-Hyun Hwang, Erwin Wu, and Hideki Koike. 2023. Ai coach: A motor skill training system using motion discrepancy detection. In *Proceedings of the Augmented Humans International Conference 2023*. 179–189.

[15] Tica Lin, Rishi Singh, Yalong Yang, Carolina Nobre, Johanna Beyer, Maurice A Smith, and Hanspeter Pfister. 2021. Towards an understanding of situated ar visualization for basketball free-throw training. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[16] Peter Mohr, Shohei Mori, Tobias Langlotz, Bruce H. Thomas, Dieter Schmalstieg, and Denis Kalkofen. 2020. Mixed Reality Light Fields for Interactive Remote Assistance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376289

[17] Ho-Jun Park, Jang-Woon Baek, and Jong-Hwan Kim. 2020. Imagery based parametric classification of correct and incorrect motion for push-up counter using OpenPose. In *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*. IEEE, 1389–1394.

[18] Helge Rhodin, Christian Richardt, Dan Casas, Eldar Insafutdinov, Mohammad Shafiei, Hans-Peter Seidel, Bernt Schiele, and Christian Theobalt. 2016. Egocap: egocentric marker-less motion capture with two fisheye cameras. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–11.

[19] Koya Sato, Yuji Sano, Mai Otsuki, Mizuki Oka, and Kazuhiko Kato. 2019. Augmented recreational volleyball court: Supporting the beginners' landing position prediction skill by providing peripheral visual feedback. In *Proceedings of the 10th Augmented Human International Conference 2019*. 1–9.

[20] Soyong Shin, Juyong Kim, Eni Halilaj, and Michael J Black. 2023. WHAM: Reconstructing World-grounded Humans with Accurate 3D Motion. *arXiv preprint arXiv:2312.07531* (2023).

[21] Shuya Suda, Yasutoshi Makino, and Hiroyuki Shinoda. 2019. Prediction of volleyball trajectory using skeletal motions of setter player. In *Proceedings of the 10th Augmented Human International Conference 2019*. 1–8.

[22] Walter Thompson. 2021. Worldwide Survey of Fitness Trends for 2021. *ACSM'S Health & Fitness Journal* 25 (01 2021), 10–19. https://doi.org/10.1249/FIT.0000000000000631

[23] Denis Tome, Patrick Peluse, Lourdes Agapito, and Hernan Badino. 2019. xr-egopose: Egocentric 3d human pose from an hmd camera. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7728–7738.

[24] Jian Wang, Zhe Cao, Diogo Luvizon, Lingjie Liu, Kripasindhu Sarkar, Danhang Tang, Thabo Beeler, and Christian Theobalt. 2023. Egocentric Whole-Body Motion Capture with FisheyeViT and Diffusion-Based Motion Refinement. *arXiv preprint arXiv:2311.16495* (2023).

[25] Jian Wang, Lingjie Liu, Weipeng Xu, Kripasindhu Sarkar, Diogo Luvizon, and Christian Theobalt. 2022. Estimating egocentric 3d human pose in the wild with external weak supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13157–13166.

[26] Jian Wang, Lingjie Liu, Weipeng Xu, Kripasindhu Sarkar, and Christian Theobalt. 2021. Estimating egocentric 3d human pose in global space. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 11500–11509.

[27] Jiqing Wen, Lauren Gold, Qianyu Ma, and Robert LiKamWa. 2024. Augmented Coach: Volumetric Motion Annotation and Visualization for Immersive Sports Coaching. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 137–146. https://doi.org/10.1109/VR58804.2024.00037

[28] Jingtao Xu, Wei Shi, and Yoshihiro Okada. 2022. Design and Implementation of an Immersive Network Collaborative Environment Using OpenPose with 360VR Camera and WebXR. In *Computational Intelligence in Security for Information Systems Conference*. Springer, 557–568.

[29] Bai Yu, Wei Shi, and Yoshihiro Okada. 2021. Action input interface of IntelligentBox using 360-degree VR camera and OpenPose for multi-persons' collaborative VR applications. In *Complex, Intelligent and Software Intensive Systems: Proceedings of the 15th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2021)*. Springer, 747–757.

[30] Chih-Hung Yu, Cheng-Chih Wu, Jye-Shyan Wang, Hou-Yu Chen, and Yu-Tzu Lin. 2020. Learning tennis through video-based reflective learning by using motion-tracking sensors. *Journal of Educational Technology & Society* 23, 1 (2020), 64–77.

[31] Qian Zhang, Akshay Paruchuri, YoungWoon Cha, Jiabin Huang, Jade Kandel, Howard Jiang, Adrian Ilie, Andrei State, Danielle Albers Szafir, Daniel Szafir, et al. 2023. Reconstruction of Human Body Pose and Appearance Using Body-Worn IMUs and a Nearby Camera View for Collaborative Egocentric Telepresence. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 96–97.