From Sports Videos to Immersive Training: Augmenting Human Motion to Enrich Basketball Training Experience

Yihong Wu
wuyihong@zju.edu.cn
State Key Lab of CAD&CG
Zhejiang University, China

Xiao Xie*
xxie@zju.edu.cn
Department of Sports Science
Zhejiang University, China

Lingyun Yu
Lingyun.Yu@xjtlu.edu.cn
Xi'an Jiaotong-Liverpool University
Suzhou, Jiangsu, China

Xinyi Ruan
ruanxy24@mails.tsinghua.edu.cn
Tsinghua University Shenzhen, China

Runzhou Li
darkpaper2024@gmail.com
State Key Lab of CAD&CG
Zhejiang University, China

Liqi Cheng
lycheecheng@zju.edu.cn
State Key Lab of CAD&CG
Zhejiang University, China

Shuainan Ye*
sn_ye@zju.edu.cn
State Key Lab of CAD&CG
Zhejiang University, China

Dazhen Deng
dengdazhen@zju.edu.cn
School of Software Technology
Zhejiang University, China

Hui Zhang
zhang_hui@zju.edu.cn
Department of Sports Science
Zhejiang University, China

Yingcai Wu
ycwu@zju.edu.cn
State Key Lab of CAD&CG
Zhejiang University, China

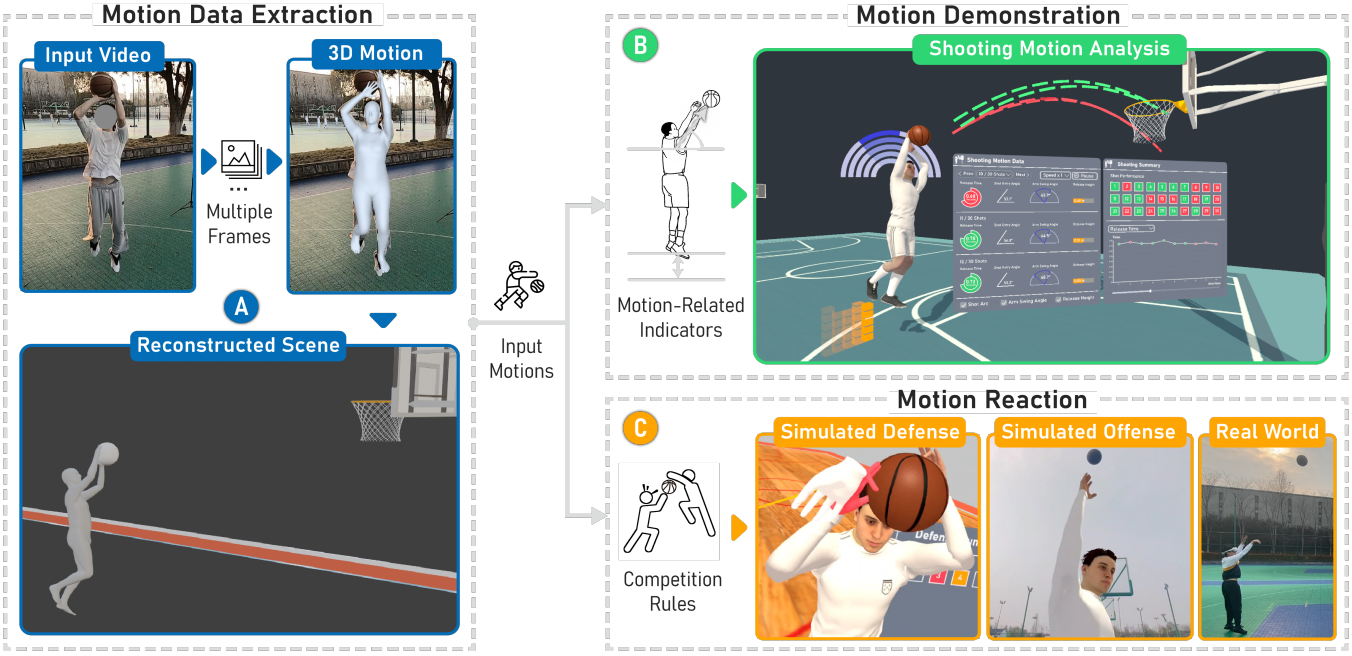


Figure 1: Overview of the immersive video training. (A) Reconstructs 3D motion from video. (B) Visualizes key indicators and provides quantitative evaluations. (C) Simulates one-on-one competitions, bridging virtual practice and real world.

*Co-corresponding authors: Xiao Xie and Shuainan Ye.

UIST '25, Busan, Republic of Korea

ACM ISBN 979-8-4007-2037-6/2025/09
https://doi.org/10.1145/3746059.3747762

Abstract

Video plays a crucial role in sports training, enabling participants to analyze their movements and identify opponents' weaknesses. Despite the easy access to sports videos, the rich motion data within them remains underutilized due to the lack of clear performance indicators and discrepancies from real-game conditions. To address this, we employed advanced computer vision algorithms to reconstruct human motions in an immersive environment, where users can freely observe and interact with the movements. Basketball shooting was chosen as a representative scenario to validate this framework, given its fast pace and extensive physical contact. Collaborating with experts, we iteratively designed motion-related visualizations to improve the understanding of complex movements. A one-on-one matchup simulating real games was also provided, allowing users to compete directly with the reconstructed motions. Our user studies demonstrate that this method enhances participants' movement comprehension and engagement, while insights derived from interviews inform future immersive training designs.

CCS Concepts

• **Human-centered computing** → **Visualization**; *Visualization design and evaluation methods*;

Keywords

SportsXR, Immersive Training, Mixed Reality

ACM Reference Format:

Yihong Wu, Xiao Xie, Lingyun Yu, Xinyi Ruan, Runzhou Li, Liqi Cheng, Shuainan Ye, Dazhen Deng, Hui Zhang, and Yingcai Wu. 2025. From Sports Videos to Immersive Training: Augmenting Human Motion to Enrich Basketball Training Experience. In *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25), September 28-October 1, 2025, Busan, Republic of Korea.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3746059.3747762

1 Introduction

Video, a vital medium for information in modern sports, has become integral to how people engage in physical activity. Sports enthusiasts often review their own training footage to identify areas for improvement, while athletes analyze match videos to recognize opponents' behavioral patterns and better prepare for future competitions. As such, the effective use of sports video holds great potential for enhancing athletic performance.

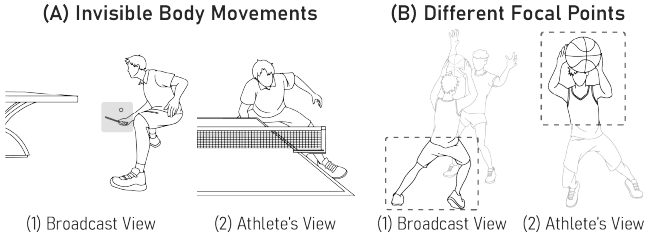


Figure 2: Potential differences between the third-person view in sports videos and the athlete's first-person view.

In recent years, video-based analysis has provided valuable insights into team strategies [19]. Besides strategic planning, individual athletes' technical skills are also key determinants of game outcomes. While these skills and movements are embedded within game videos, fully extracting and utilizing this information presents significant challenges. **First, the performance of movements is difficult to present in a quantifiable manner.** Beyond basic statistics (e.g., race time), motion analysis relies on sport-specific indicators [26], such as stride length. These indicators vary by sport and context (e.g., offense vs. defense), often targeting specific body parts. Determining them demands substantial domain knowledge. Moreover, manually recording motion-related indicators from videos is labor-intensive, which further raises the bar for analysis. **Second, fixed camera angles may make it hard to observe all aspects of movements.** As shown in Fig. 2(A1), some body parts (e.g., right arm) may be occluded. Traditional videos also lack depth information, affecting the accurate perception of 3D aspects like an arm swing's trajectory. While fixed viewpoints remain widely used, complementary methods can capture more details of complex movements. **Third, movements observed in videos often differ from those perceived in real matches.** Insights from pre-match video analysis are not always directly applicable in real games. For example, a stroke clearly visible in broadcast view (Fig. 2(A1)) may be obscured from the athlete's perspective (Fig. 2(A2)), making the stroke technique harder to judge. Likewise, from a third-person perspective, viewers can detect the ball handler's intent to fake a shot through subtle leg cues (Fig. 2(B1)). Yet in head-to-head competition, inexperienced defenders are easily distracted by the ball (Fig. 2(B2)), and often miss these cues. Athletes must adapt to this perceptual gap to react effectively in real games.

Recognizing the limitations of traditional sports videos, we turn to immersive training [12], which features realistic scenario reconstruction and in-situ interaction, often supported by technologies such as VR and MR. Although multi-angle videos can help mitigate perspective limitations, the effort required for capturing and the need for careful synchronization and management limits their use in regular training—especially for individuals. Furthermore, immersive environments offer unique benefits: they support depth and spatial perception, which are essential for accurately judging and responding to fast, dynamic actions (e.g., hitting a fastball)—capabilities not available in standard 2D video.

Despite these advantages, acquiring motion data for immersive environments often relies on wearable motion capture devices, which are expensive and impractical for capturing opponents. In contrast, sports videos are rich in motion data and widely accessible, yet remain underutilized as a resource for immersive training. To bridge this gap, we propose a proof-of-concept approach that immerses users into sports videos, enabling first-person interaction with opponents reconstructed from video. Recent advances in video-based 3D motion reconstruction [22, 49, 64] make it possible to combine the accessibility of video with the perceptual advantages of immersive environments, expanding training opportunities.

In this work, we present an immersive sports training approach that enables trainees to analyze and interact with video-captured human motions. Focusing on basketball shooting, we collaborate with domain experts to identify key performance indicators and

design immersive visualizations that connect motion data with performance insights. We also simulate one-on-one matchups, allowing users to practice both offense and defense against video-based opponent actions. User studies validate the effectiveness of our visualizations and simulations. Our main contributions are:

- A design study with domain experts to identify key challenges and tasks in the current use of video for sports training.
- A proof-of-concept training method that immerses users in sports video scenes, enabling motion analysis through visualizations or match simulation via interaction with virtual characters.
- User studies assessing the effectiveness of immersive basketball training, along with insights derived from interviews.

2 Related Work

In this section, we discuss relevant studies, including motion-related visualization, immersive training, and motion reconstruction.

2.1 Motion-Related Visualization

As motion capture technology advances, interpreting motion data becomes increasingly important [62]. Due to their intuitive nature, motion-related visualization have proven effective for understanding complex movements [43, 47]. We categorize these visualizations by the motion attributes they emphasize.

Trajectory Guidance. Trajectories intuitively represent direction and position, and are widely used in immersive sports analytics [15]. For example, AvatAR [46] embedded 3D trajectories alongside virtual avatars in immersive environments to enhance the understanding of complex movements. Similarly, Reactive Video [16] used trajectories to provide real-time feedback, enabling users to compare their arm movements against standards and make immediate adjustments. Various techniques also visualize directional data [52, 57, 62]. PoseCoach [35], for instance, uses glyphs to highlight key joint angles and positions for running posture coaching.

Body Part Augmentation. In physical activities, certain body parts are often augmented to highlight key muscle actions [9, 44, 55]. Semeraro et al. [47] highlighted major muscle groups in fitness videos to help beginners learn movements. Showcasing skeletal joints is also a common visualization [13, 18, 20]. To analyze continuous skiing movements, Wang et al. [59] applied pose detection to enhance posture and body contours in key frames. Some approaches use metaphors—for example, sharply extending parts of the body that need to be stretched during dance [52]. Similarly, our work highlights the virtual opponent's body parts involved in simulated physical contact during head-to-head competition.

2.2 Immersive Physical Training

Immersive training enhances user engagement by providing realistic experiences and real-time feedback [29, 34, 50]. Many immersive sports systems do not use real equipment, instead focusing on equipment-free activities such as running [11] and martial arts [21, 27]. For example, Pastel et al. [43] developed a Karate learning system where trainees use motion capture to compare their movements with standard ones in VR. For equipment-based sports, immersive methods often recreate game scenarios, allowing trainees to recognize and react without physical gear [23]. In baseball, head-mounted displays [37, 66] and 3D screens [41] help players identify different pitches from the batter's view. To improve spatial awareness, Tsai et al. [56] placed users in a VR court to learn basketball strategies, using location tracking to assess tactical awareness.

Immersive training can be further enhanced by incorporating real sports equipment [40, 51], often through modified or sensor-embedded gear. In racquet sports, SpinPong [61] used a sensor-equipped racket to capture physical data during strokes. In addition, ski simulators combined with Vive Trackers [60] enhance physical sensations and support detailed motion analysis. Some approaches use imperceptible data collection for real-time feedback. For instance, Lin et al. [33] applied computer vision algorithms to capture and visualize the in-situ trajectory of basketball free throws.

Current sports coaching systems often trade off ease of data collection against the complexity of training tasks (Table 1). Commercial tools like Catapult [7] and SkyCoach [8] primarily use broadcast videos, limiting their applications to performance analysis with minimal interaction involving athlete motion. In contrast, immersive avatar-based systems focus more on motion demonstration. For example, VIRD [32] uses monocular badminton videos for in-situ motion playback, but its core analysis relies on expert-collected shot data rather than detailed motion analysis. More complex tasks typically require additional hardware—avaTTAR [39], for instance, supplements video data with IMU sensors, enabling real-time side-by-side action comparisons. Still, most existing systems focus on visual guidance or comparison, without supporting direct physical interaction with avatars. Therefore, our work establishes a cost-efficient framework, leveraging advanced computer vision algorithms to convert accessible 2D video into 3D motion representations. Through a proof-of-concept, we aim to demonstrate that even simple data sources like video can support more sophisticated interactive tasks (e.g., simulated basketball shooting defense), opening new possibilities for immersive training.

2.3 Human Motion Reconstruction

Human motion capture has traditionally been a challenging task due to its reliance on specialized equipment and controlled environments [17, 24]. However, with the growing availability of video data, recent advances have significantly improved motion reconstruction from monocular video [31, 36, 48]. Many methods build on end-to-end human mesh recovery frameworks [28], often using SMPL [38] parameters to reconstruct motion from video. For example, CLIFF [30] incorporates full-frame location information to enhance global motion awareness, while SLAHMR [63] extends reconstruction to the world frame by decoupling video and motion, enabling multi-person tracking in global coordinates.

Despite these advances, monocular methods still struggle with occlusions and extreme poses [54, 58]. To address these challenges, recent work has focused on improving robustness in such conditions. MAED [65] introduces a multi-level attention mechanism that learns spatial, temporal, and joint-level cues to better reconstruct occluded body parts. 4DHumans [22] adopts a transformer-based architecture for reliable reconstruction and tracking under partial visibility. Given their effectiveness, such methods have been successfully applied to sports video, yielding reliable results [14, 32]. In our work, we leverage these advances in human mesh recovery algorithms to extract motion from monocular video.

	Input Source	Interaction Mode	Visualization Types	Feedback	Sport & Task Specificity
Catapult [7]	monocular video	video replay	co-related visualizations	visual	football performance report
VIRD [32]	monocular video	motion demo	situated 3D visualizations	visual	badminton game analysis
VR Karate [43]	mocap system	guided demonstration	on-body visualization	visual	karate movement learning
VR Basketball [56]	synthetic animation	guided demonstration	on-field trajectory	visual & audio	basketball tactical training
avaTTAR [39]	video with sensor	real-time comparison	on-body & detached cues	visual, real-time	table tennis training

Table 1: Comparison of representative sports coaching and immersive avatar-based training systems.

3 Formative Study

In this section, we present expert interviews, summarize challenges in video-based sports training, and outline tasks for different goals.

3.1 Expert Interview

Different types of sports emphasize distinct aspects of movement. For example, *performance-focused sports* (e.g., gymnastics) prioritize precision and speed, while *head-to-head sports* (e.g., basketball, table tennis) focus on competitive interaction. Similarly, the role of video varies across sports contexts: *training recordings* are often used to capture running postures, while *match broadcasts* support tactical analysis of opponents in table tennis.

To gain generalizable insights into video applications across diverse sports contexts, we collaborated with four seasoned experts (E1-E4) with extensive cross-sport experience: 1) E1 is a trainer for the national table tennis team and a professor of sports science, known for his leadership in international sports academia; 2) E2, a former professional table tennis player, is now a researcher specializing in the tactical analysis of racket sports, including tennis and badminton; 3) E3 is a university professor focusing on invasion team sports, such as basketball and soccer, and data-driven training methods. 4) E4 is a university lecturer who teaches fundamental sports skills, such as basketball, orienteering, and yoga, with a focus on coaching beginners. This collaboration aimed to validate our research motivation, identify typical challenges of using videos for sports training, and inform a potential improvement framework.

Informed by expert profiles, we conducted semi-structured, one-on-one, in-person interviews, each lasting approximately 40 minutes. The interviews were organized into two main sessions:

Sample Video Presentation and Documentation. First, we constructed a sample set of 15 sports videos, organized into three categories—*skill instruction*, *match broadcasts*, and *training recordings*—with each category containing videos from five sports: basketball, table tennis, orienteering, gymnastics, and volleyball. Experts reviewed videos by category and described how they incorporated these videos into their workflows, sharing practical examples. The responses revealed varied roles for each video type:

- E1 values *match broadcasts* most for analyzing and explaining opponents' tactical characteristics to team members. He also uses *training recordings* for targeted exercises with each player. As he works primarily with high-level players, *skill instruction* is rarely used except when explaining concepts to non-professionals.
- E2 works in a similar context to E1 but places even greater emphasis on *match broadcasts*. He systematically breaks down each rally, documenting the details of every stroke technique to support data-driven tactical insights.
- E3 mainly uses *training recordings* to monitor participants' physical capabilities (e.g., reaction time), combining video with diagrams. Other types mainly serve as illustrative tools in teaching.
- E4 specializes in creating *skill instruction* videos for beginner teaching. During the COVID-19 pandemic, she leveraged these videos for online courses and required students to submit *training recordings* for review and grading.

Semi-Structured Questioning and Brainstorming. Next, we began with open-ended questions tailored to each expert's work context, aiming to explore their views on current video usage. Example questions included: *"Have you encountered any problems when using match broadcasts for pre-match preparation?"* and *"If training recordings could be enhanced, what specific features would you find most beneficial?"* Their responses revealed crucial limitations like technical constraints (e.g., lack of multi-angle views) and practical challenges (e.g., differing perspectives on opponent movements). Following this, we held a brainstorming session, inviting suggestions for potential improvements without immediate evaluation. The suggestions included: *"Introducing multi-angle playback for skill instruction videos."* and *"Expanding videos with customizable content (e.g., real-time speed presentation) and interactive features."* After the interviews, we conducted content analysis: responses were transcribed and coded by research objective (e.g., "*common challenges*", Sec. 3.2), and grouped into themes like "*motion clarity*," "*interactivity*," and "*customization needs*." This analysis clarified commonalities and differences in video use, detailed in later sections.

3.2 Problem Characterization

A key commonality is that videos generally serve as effective alternatives to live demonstrations and instructions from coaches, supporting participants at all skill levels (from beginners to professionals). Their primary applications can be categorized into two distinct purposes: 1. **Motion Demonstration** ("*clarity*") - Helping trainees understand complex movements and evaluate their performance, with the ultimate goal of refining their skills. 2. **Motion Reaction** ("*interactivity*") - Guiding trainees to adapt to opponents' movements and practice appropriate responses, aimed at enhancing their ability to counter those movements. Additionally, videos should be tailored to the specific requirements of each sport ("*customization needs*"). For example, table tennis videos need to fully capture the ball's trajectory and landing points on the table. Based on these two purposes, we categorized the relevant challenges:

P1 Video often fails to fully capture movement details (Demonstration). Many sports videos do not always capture every movement detail, and sometimes even lose crucial information. E2 noted, *"When annotating table tennis videos, we often encounter situations where the player's lower body is obscured by*

the table, forcing us to infer their footwork based on our experience." Although filming from multiple angles can alleviate this issue, it also complicates data collection and increases the time required to analyze matches.

P2 Evaluating movement quality in videos is complex (Demonstration). Straightforward metrics for certain movements may only partially reflect the movement's overall quality. For instance, an athlete's high shooting percentage in training does not guarantee good performance in a game. Factors such as slow shooting speed, which makes it easier for opponents to interfere, must be considered when giving training advice. Thus, assessing a movement requires an understanding of various aspects of the sport, which can be challenging for those without relevant knowledge. E4 stated, *"In teaching, my primary task is to clearly explain the key points of the movement so that beginners can grasp it better. A simple demonstration is not enough."*

P3 The perspectives in real matches and videos differ significantly (Reaction). Analyzing opponents' technical movements is regular in pre-match preparation. E1 and E2 shared that they often have athletes watch representative plays from their opponents to gain a deeper understanding of their technical characteristics. However, they noted that learning from videos alone is insufficient, as the videos usually offer a third-person perspective, which greatly differs from the first-person view experienced in real games. This discrepancy can make it challenging for athletes to quickly recognize the intent behind their opponents' movements.

P4 The movements in videos are not reactive (Reaction). In current video training methods, athletes often find themselves merely observing the movement without any chance to actively engage with it. This one-way process especially poses a great challenge for adversary sports like basketball, which emphasize the interplay between offense and defense. When athletes cannot have their choices and actions directly measured, a disconnect between training and actual competition arises, affecting the effectiveness of the training. E2 suggested, *"If videos could respond to athletes like a game does, I think training would be a lot more fun and effective."*

Based on the challenges identified in expert interviews, we outline three tasks to improve current training methods:

T1 Reproducing 3D human motions from videos (P1).
T2 Immersing viewers in sports scenes within the video (P1, P3).
T3 Providing sport-specific visualization and interactive experiences for movements (P2, P4).

3.3 Framework Overview

In summary, sports training with videos fulfills two complementary needs: 1. *Analyzing Motions* - Helping users identify flaws in their movements and refine their skills. 2. *Simulating Matches* - Allowing users to practice against opponents in videos and better prepare for real competitions. These two needs align directly with the core purposes of video usage in sports training: **Motion Demonstration** and **Motion Reaction**.

To overcome the limitations of 2D videos, we propose an immersive video training framework that offers unique advantages such as in-situ interaction. This approach effectively fulfills both analytical and reactive demands. Our framework consists of three key components: 1) **Motion Data Extraction** (Sec. 4): using advanced computer vision algorithms to extract 3D motion data from videos (Fig. 1(A)), providing input for both training needs; 2) **Motion Demonstration** (Sec. 5): presenting key motion-related indicators and providing quantitative evaluation to enhance clarity (Fig. 1(B)); 3) **Motion Reaction** (Sec. 6): highlighting the interactive nature between players in head-to-head sports, creating a simulated virtual practice environment aligned with competition rules (Fig. 1(C)).

After consulting with experts, we selected basketball shooting as a representative scenario to validate this framework. The choice was motivated by its need to replicate realistic sports situations, address complex evaluation metrics (e.g., *shooting motion analysis*), and emphasize physical interactions between players (e.g., *simulated one-on-one competition*).

4 Motion Data Extraction

In this section, we detail the process of extracting information about athletes' movements and equipment from sports videos of diverse sources and formats (**T1**). We also demonstrate this process using a monocular video sample (Fig. 3).

4.1 Pre-Settings: Hardware Requirements

To optimize video processing efficiency, we use a server equipped with an Intel Xeon Gold 6226R CPU, 64GB of memory, and an RTX3090 GPU. Given the variance in data quality among sports videos, we focus on two validated video types: *Multi-camera fixed-view videos* using EasyMocap [3, 49] for human motion reconstruction, and *Single-camera fixed-view videos* using 4DHumans [22], both achieving state-of-the-art performance in their respective tasks. We have completed motion data extraction for two validated video types. For unfixed-view videos, such as players' first-person perspectives, frequent changes in the field of view and significant camera parameter estimation errors in poor reconstruction quality. Given the widespread use of monocular sports videos and the greater challenges they present compared to multi-view reconstruction, our pipeline focuses on the monocular video input.

4.2 Pipeline

We present an extensible process (Fig. 3) tailored for the validated video inputs, with the following steps:

Data Preprocessing. Given the excessive computational resources of 3D motion reconstruction, it is necessary to remove irrelevant segments (e.g., picking up the ball) from the video to streamline the process. In our example video, we utilize an automatic segmentation technique to detect shot-making events and isolate potential shooting moments in each clip by tracking the 2D trajectory of the basketball. The preprocessing step allows for manual verification (5-10 mins) of start and end points for precise clipping. We start with a lightweight YOLO detection model [45], processing video frames at 77 FPS on our server, to locate the basketball in the 2D frames (Fig. 3(A)). The top half of the video, manually designated as a potential "ball flight area", where any detected continuous ball flight paths approaching the rim are automatically labeled as "shooting segments". Despite this automation, manual checks are

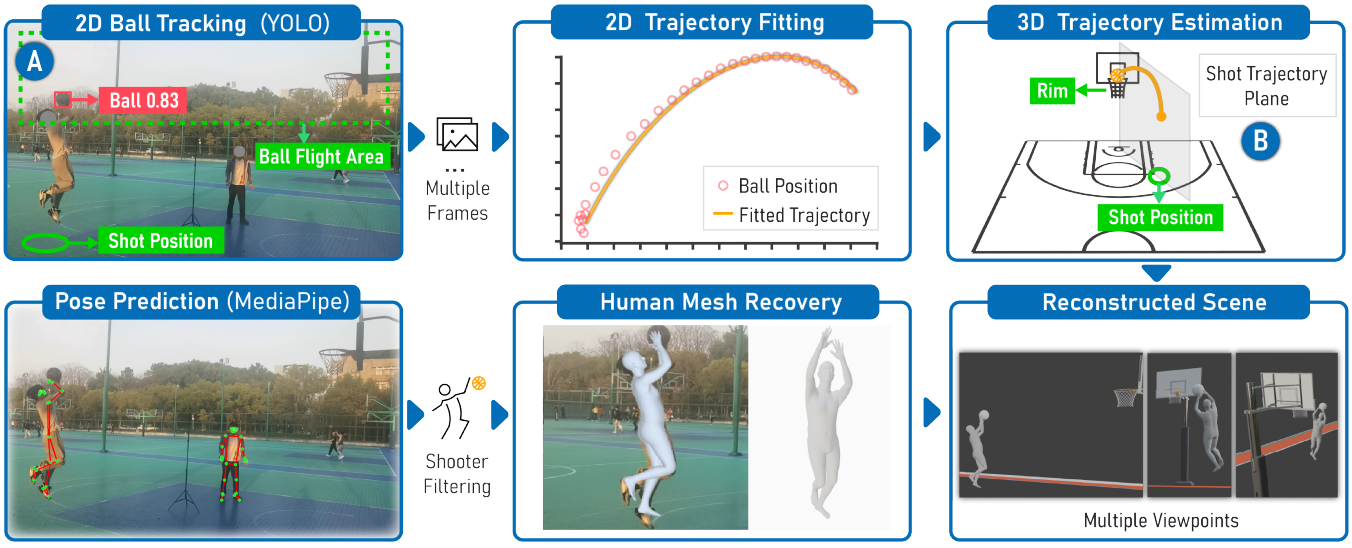


Figure 3: The pipeline for reconstructing basketball shooting scenes from our monocular video sample.

still required to correct any YOLO detection errors or interference from other basketballs. While this method greatly reduces the workload compared to manual editing, it still requires careful setup of the recording location for sports videos.

Shooter Filtering and Human Motion Reconstruction. To distinguish the shooter from others in the video, we employ YOLO for human detection, and MediaPipe [4] for skeletal joint detection of visible individuals (Pose Prediction), achieving a processing speed of 33 FPS. By matching the basketball's flight trajectory with the detected hand joints, we locate the individual performing the shooting action and the corresponding timepoint. Next, we isolate the approximate region of the shooter's complete motion and apply masks to other individuals, providing input for 3D motion reconstruction (Human Mesh Recovery). Here we utilize 4DHumans [22] for this process as the example video is monocular. Subsequently, we import the reconstructed results into Blender to review the motion quality from multiple perspectives (25-30 mins). Due to factors such as lighting and shadows, potential deviations along the z-axis may occur, such as the athlete floating or the feet sinking below the ground in few cases. Therefore, it is essential to check the z-axis of human motion properly to ensure a stable stance is maintained while stationary on the court.

3D Ball Trajectory Estimation. Accurately restoring 3D ball trajectories with multi-camera setups requires processes such as camera calibration, time synchronization, 2D tracking, and triangulation. As for single-camera setups, the lack of depth information in 2D views forces us to rely on estimation methods, which compromises the accuracy of trajectory reconstruction. We first extract the 2D positions of the basketball from the moment it leaves the shooter's hand until it hits the rim, and fit them into a parabolic trajectory (2D Trajectory Fitting). Next, based on the original video, we manually mark the shot position on the 3D court model and, together with the rim position, determine the vertical plane of the ball's trajectory (Fig. 3(B)). Finally, using the fitted parabolic function, we estimated a 3D shooting trajectory that lies approximately near the release-to-hoop plane, based on the ball's release position.

In the final phase, we adjust the virtual player's position coordinates by scaling and alignment, and fine-tune the basketball trajectory using rule-based methods, such as ensuring the ball is always held in hand before shooting. It is worth noting that our processing pipeline is tailored to the specific basketball training scenario addressed in this work. Based on our experience with a 5-minute example video, the total processing time is about 6 hours: around 4.5 hours for automatic image processing, roughly 1 hour for manual fine-tuning and validation, and the rest for steps like data transfer. Most of the time is consumed by image processing, especially motion reconstruction with 4DHumans. In the future, we plan to explore more efficient models or increase computational resources to speed up processing. Example processing code is published at https://github.com/qinwyh/SportsVideos.

4.3 Evaluation: Objective & Subjective Analysis

Objective Evaluation. We evaluate the accuracy of our 3D reconstruction by comparing it to more reliable multi-view reconstruction methods [3] due to the difficulty of obtaining high-quality ground-truth motion and ball data for in-the-wild basketball videos. During our data collection, we set up three or more additional cameras (not including the monocular view) to capture the same session, and use the multi-view reconstruction results as a reference for quantitative evaluation of 3D motion and ball trajectory accuracy. For our test data, the Procrustes-aligned mean per joint position error (PA-MPJPE) for motion reconstruction was 87.5 mm, and the average 3D distance error for the ball trajectory is 0.52 meters. These values indicate that, in terms of numerical accuracy, there are still errors in the reconstructed body motion and ball trajectory, and there remains considerable room for improvement. However, this evaluation is itself not absolutely rigorous because multi-view reconstruction is not entirely error-free. Considering the core focus

of our work is not to achieve the utmost reconstruction accuracy but the whole interactive training framework, we further conduct a subjective evaluation of our 3D reconstruction on user experience.

Subjective Evaluation. In addition to the quantitative validation, we further assessed the practical usability of our reconstructions from a user perspective. In Module Usability Study 2, we invited five university-level basketball players and ten regular enthusiasts (none of whom participated in other parts of the study) to evaluate five pairs of original training videos and their corresponding 3D reconstructions rendered in Blender, covering various environments and participants. Participants rated the results on a 7-point Likert scale (1 = unacceptable, 7 = acceptable), focusing on overall quality, motion fluidity, key joint articulation (e.g., elbows and hands), and ball trajectory (e.g., release angle and flight path). The enthusiast group gave an average score of 5.9 (STD = 0.88), and the player group scored 5.8 (STD = 1.10); all subcategories received average ratings above 5.5, indicating that the reconstructions were generally acceptable for practical training use.

Limitations. Despite these positive subjective outcomes, several limitations remain. First, relying on monocular video input can lead to reduced accuracy, especially in cases of fast motion or occlusion. However, our main goal is to establish a cost-effective video-driven 3D motion reconstruction framework, leveraging existing computer vision algorithms to convert easily accessible 2D video into 3D motion representations. Through a proof-of-concept system, we demonstrate that even simple data sources like monocular video can support sophisticated interactive training tasks (e.g., simulated shooting defense), opening new possibilities for immersive sports training. Should more accurate monocular reconstruction or ball trajectory models become available in the future, they can be seamlessly integrated into our pipeline. Second, while 4DHumans [22] provides a solid technical foundation, its performance for fine-grained basketball scenarios—such as hand-ball interactions and subtle joint movements—has not been systematically validated, which may affect the accuracy of technical feedback for certain skills. It is worth noting that our current data extraction pipeline is modular and designed for extensibility: more advanced pose estimation or ball tracking models can be integrated to improve reconstruction quality without modifying the core immersive training framework. Thus, while there are certain limitations, these do not fundamentally undermine the practical applicability of our approach in real-world training environments.

5 Motion Demonstration

In this section, we describe our collaboration with basketball experts to identify key indicators, the visualization design of shooting motion analysis (**T2, T3**), and a use case to verify its effectiveness.

5.1 Pre-Study: Key Performance Indicators

To better understand basketball shooting scenarios, we reviewed relevant literature [10, 53] and held discussions with E3 to identify candidate performance indicators for shooting motions. Explanatory charts were created for all indicators to facilitate evaluation. The identified indicators were categorized into two groups:

- Ball-related: *Ball Speed*, *Spin Rate*, *Release Point*, *Arc Height*, *Flight Time*, *Shot Release Angle*, and *Shot Entry Angle*.

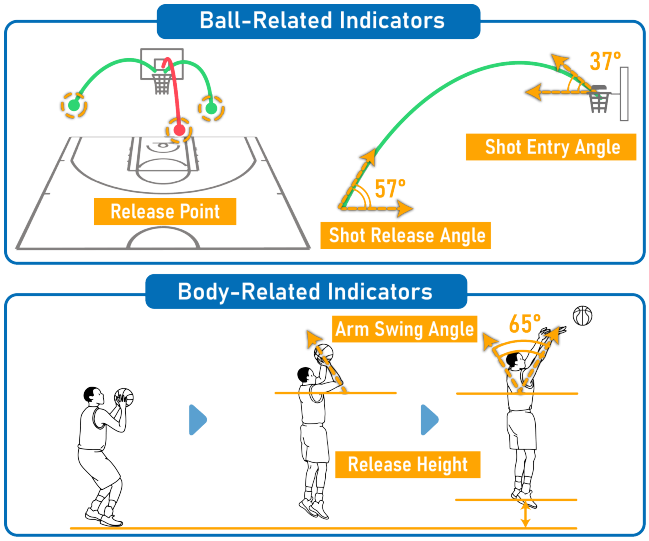


Figure 4: The explanation of several key performance indicators for shooting training.

- Body-related: *Knee Angle*, *Hip Angle*, *Elbow Angle*, *Release Angle*, *Ankle Angle*, *Shoulder Angle*, *Elbow Height*, *Release Height*, and *Release Time*. *Elbow Angle* and *Release Angle* were combined into a single indicator (*Arm Swing Angle*), resulting in 8 body-related candidate indicators.

We collaborated with six experienced participants, including two university basketball coaches (C1, C2), two university-level players (A1, A2), and two enthusiasts (F1, F2), each with over 10 years of playing experience. All participants were proficient in basketball shooting training and none overlapped with E1-E4 or experiment participants. A one-hour group discussion was conducted with the participants. During the session, we presented each candidate indicator along with its explanatory chart to ensure a clear understanding of its meaning. Participants then voted individually on whether each indicator should be included in the shooting training evaluation. To minimize peer influence, participants were asked to provide brief explanations for their votes only after completing the voting process. Indicators with more than two-thirds agreement were selected and incorporated into motion demonstrations for further analysis. In the end, three ball-related indicators (*Release Point*, *Shot Release Angle*, *Shot Entry Angle*) and three body-related indicators (*Arm Swing Angle*, *Release Height*, *Release Time*) were chosen. A brief explanation is shown in Fig. 4.

5.2 Immersive Visualization Design

Based on the indicators established in Sec. 5.1, we progressively optimized the design of the immersive video training system. This involved close collaboration with three experts, including a coach (C1), a player (A1), and an experienced enthusiast (F1). For each indicator, we showed the experts multiple design alternatives and provided an in-depth experience using an HMD. We then incorporated their feedback and made adjustments accordingly. After

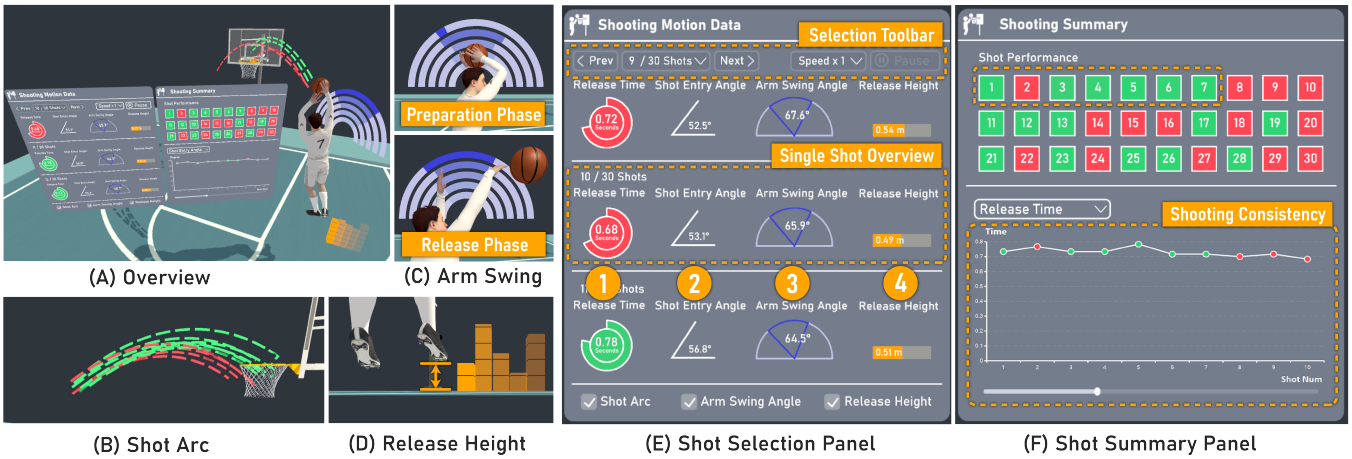


Figure 5: The visual design for shooting motion analysis. By wearing an HMD, users can explore reconstructed shooting motions from basketball videos (A). Key indicators are shown through motion-related visualizations (B), (C), and (D). Detailed values for each shot appear in (E), while the summary panel helps assess the consistency of consecutive shooting motions (F).

several iterations, we finalized the motion-related visualization design they found most intuitive and effective.

In this training scenario, trainees utilize their own shooting videos as input to thoroughly analyze their shooting techniques. To offer a comprehensive view of shooting motions (**P1, P2**), we integrate the reconstructed scenes and motion-related visualizations into an immersive environment (Fig. 5(A)). The player model is sourced from Mixamo [5].

5.2.1 *Motion-Related Visualizations.*

Shot Arc. A series of 3D parabolic trajectories visualize ball-related indicators (Fig. 5(B)). Trajectories are color-coded by outcome: green for hits and red for misses, starting from the player's *Release Point* and ending at the rim. To assess the consistency of shot trajectories, the current shot and several previous ones are displayed together. Users can adjust the number shown to identify trajectory patterns associated with successful shots.

Arm Swing Angle. We use a blue arc to represent the *Arm Swing Angle* during basketball shooting (Fig. 5(C)), from the player's preparation phase to the release phase [10]. The arc updates in real-time with the player's arm movement, positioned near the release point and aligned with the shot direction. Concentric arcs display the consistency of arm swings between consecutive shots, with inner arcs representing the angles of earlier shots.

Release Height. We employ orange cubes to depict the player's jump height at any moment (Fig. 5(D)). To minimize perspective-related errors in height perception, a block-like design is employed to quantitatively display the absolute height. The cube dynamically changes with the player's shooting motion, settling at the *Release Height* once the ball is released. Transparency levels indicate the sequence, with earlier shots appearing more transparent.

5.2.2 *Motion Data Panels.*

Based on expert feedback, while motion-related visualizations are informative, they mostly offer a relative perception of indicators. However, experts often seek specific numerical values. To fulfill this need, we design two data panels to enhance usability, which users can move by pressing the controller's button and dragging.

Shot Selection Panel. As shown in Fig. 5(E), we provide a table-like list, which displays the specific attributes of each shot, updating in real-time as the motion plays. Each row includes a set of intuitive 2D visual elements (Single Shot Overview), arranged as follows: (1) the *Release Time* indicated by the outer arc's radius and the shot success marked by colors, (2) the *Shot Entry / Release Angle*, (3) the *Arm Swing Angle* depicted in blue, and (4) the *Release Height* shown in a bar. Users can select a specific shot from the dropdown menu or adjust the playback speed in the Selection Toolbar (Fig. 5(E)). The visibility of visualizations attached to the virtual player can be toggled using the checkbox at the bottom.

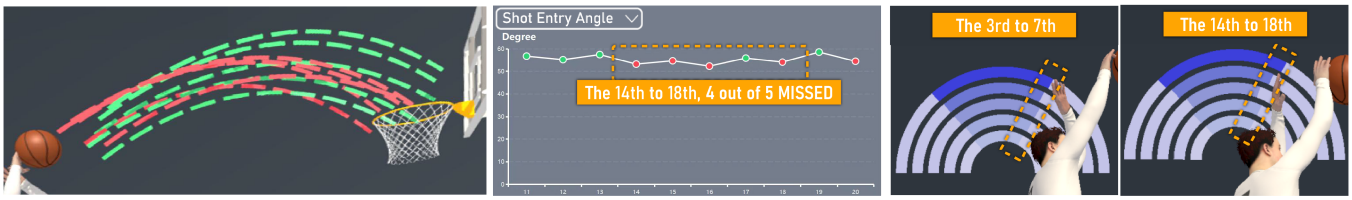
Shot Summary Panel. To offer an overview and evaluate the consistency of all shooting motions, we design a summary panel (Fig. 5(F)). The top part gives a general impression of shooting performance, while the bottom part includes a line chart to track changes in key indicators across consecutive shots.

5.3 Evaluation: A Usage Scenario

This section presents a potential usage scenario, designated as Module Usability Study 2 for shooting motion analysis. The analysis is based on an example video as the data source. As such, all insights are practical suggestions from real-world footage, further demonstrating the effectiveness of our approach.

Bob, a basketball enthusiast, is frustrated by his stagnant shooting accuracy. Despite repeatedly watching his training recordings, he cannot figure out how to improve. Bob turns to our motion analysis module, selects a training video, and recreates his shooting practice (thirty shots) in an immersive environment. By examining his movements and visualizations in detail, he hopes to identify specific areas for improvement.

Insights. At first, Bob checked the shot summary panel (Fig. 5(F)) and saw he did well in his first few attempts (making 6 out of 7 shots), but his performance dropped in later attempts (missing 13 out of 23 shots). To find out the reason, he reviewed his shooting



(A) The 11th to 20th Shot Arc (B) The 11th to 20th Shot Entry Angle (C) Arm Swing Angles

Figure 6: The use case based on our sample basketball video (Sec. 4).

motions and noticed that his shot arcs were somewhat inconsistent (**Insight1**). His first 10 shots showed concentrated arcs (Fig. 5(B)), leading to a higher success rate. However, from the 11th to the 20th shot, his arcs spread out (Fig. 6(A)), resulting in a lower accuracy. The inconsistency of his shot arcs was also evident in the shot entry angles. While exploring the shot summary panel, Bob found that the entry angles of his 11th to 20th shots showed some fluctuation (**Insight2**, Fig. 6(B)). To see if there were any problems during this phase, he compared his best five consecutive shots (3rd to 7th, all successful) against his worst (14th to 18th, missing 4 out of 5), closely examining the shooting motion and visualizations. He then noticed that his arm swing angle (Fig. 6(C)) was very stable from shot 3 to 7 but fluctuated considerably from shot 14 to 18. This led Bob to realize that his arm was not swinging consistently in the later shots, which contributed to the reduced accuracy (**Insight3**).

From this analysis, Bob concluded that he should focus on strengthening and stabilizing his arm. This would help keep his arm swing within a stable, comfortable range on successive shots, improving his overall shooting performance.

6 Motion Reaction

In this section, we detail our collaboration with experts to establish competition rules and identify relevant indicators, as well as the interaction with virtual opponents to simulate real-game competition scenarios (**T2, T3**)

6.1 Pre-Study: Competition Rules

Competition rules are crucial in designing simulated competitions, ensuring training scenarios closely replicate real-game conditions and yield valuable insights. With E3's guidance, we explored one-on-one shooting training and its rules, which emphasizes the interplay between shooter and defender. The focus shifts depending on the trainee's role: 1. On offense, trainees aim to maintain shooting accuracy and stability under defensive interference. Key indicators include shooting percentage, shot arc, arm swing angles, etc. 2. On defense, trainees focus on disrupting the ball's trajectory and applying spatial pressure (e.g., staying close or blocking vision) while avoiding fouls like direct hand contact. These principles ensure realistic offense-defense interactions.

To identify defense-related indicators (offense-related indicators align with those for individual shooting training), we followed the same process as in Sec. 5.1. Four candidate indicators were involved: *Closeout Speed*, *Anticipation Ability*, *Defensive Effectiveness*, and *Defensive Timing*. Collaborating with the same group of experts, we finalized these indicators and iteratively refined the offense-defense interaction design. The final selected indicators are:

- *Defensive Effectiveness*: In defense, there are four outcomes ranked from best to worst: blocking the shot, applying defensive pressure at release (in terms of position and height), failing to apply pressure, and committing a foul.
- *Defensive Timing*: The time gap between the opponent's shot release and the defender's jump peak, reflecting the defender's ability to anticipate the shot timing.

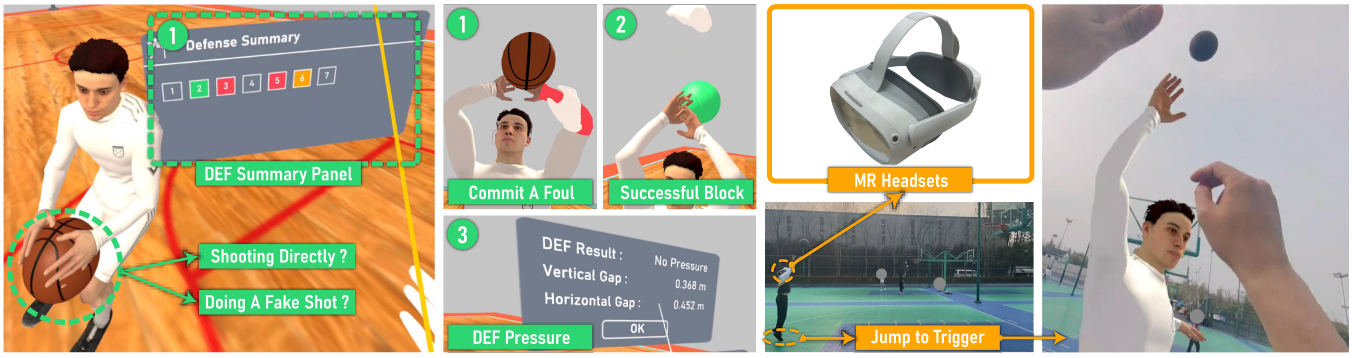
6.2 Interaction Design for Simulated Matches

To make video training in the immersive environment a more reactive process (**P3, P4**), we design role-specific interaction methods for one-on-one shooting training (Fig. 7). Trainees use their opponent's video to face their movements from a first-person perspective, simulating real-game conditions to enhance reaction abilities. This aims to enhance users' ability to recognize and respond to various technical movements in a simulated manner.

6.2.1 *Simulated Defense Practice.*

When users act as defenders, the virtual player replicates the opponents' shooting motions reconstructed from videos. This allows users to observe and defend against these motions from a first-person perspective (Fig. 7(A)). In order to simulate the complexity of a real game scenario, the virtual player's shooting motions include various types, such as quick shots (to avoid defensive interference) and fake shots (to mislead the defender's timing to jump).

Feedback on Defensive Performance. Users can lift hands and interfere with the virtual opponent's shot in a defensive manner akin to a real game. We assess the user's defensive quality in this immersive environment using key indicators (Sec. 6.1), and provide three types of feedback to support effective and intuitive training: 1) Avatar-based visual feedback: To enhance immersion, when potential physical contact occurs during defense, feedback is directly overlaid on the virtual character. For example, in the case of *Commit A Foul* (Fig. 7(B1)), when the user contacts the opponent before touching the ball, the affected body part turns red. 2) Object-based outcome feedback: If the user successfully blocks the shot without making contact with the opponent, the ball turns green and deflects, providing an immediate and clear indication of a successful block (Fig. 7(B2)). 3) Performance panels: For cases without direct contact, defensive pressure is evaluated by comparing the user's distance and height difference to the virtual player at the moment of shot release, which is shown on a performance panel (Fig.7(B3)). All defense attempts are also summarized on the Defense Summary panel (Fig.7(A1)), where green indicates a successful block, red indicates a foul, and yellow or blank denotes whether defensive pressure was exerted. Since hand position is crucial for defense,



(A) Simulated Defense Practice (B) Feedback on DEF Results (C) Real-World Setup (D) Simulated Offense Practice

Figure 7: The simulated one-on-one competition against a virtual opponent. When defending, users aim to recognize the virtual player's intentions, such as a real or fake shot (A), and adjust their defense based on performance feedback (B). When attacking, users focus on maintaining shooting accuracy and stability (C) under the visual pressure of the opponent's defense (D).

we keep the user holding the controllers throughout the defensive practice despite the fact that many head-mounted displays (HMDs) support hand tracking. This may slightly reduce immersion, but it is a compromise to ensure the accuracy of the assessment results.

6.2.2 Simulated Offense Practice.
All visualization designs are implemented in a VR environment for optimal display. However, shooting practice involving real ball interaction requires an HMD with mixed reality (MR) support. In our experiments, we used the PICO 4 [6] to enable MR functionality.

During simulated offense practice, the virtual opponent's defensive moves—extracted from real opponent videos—provide visual interference to the user's shooting attempts (Fig.7(D)). As a form of reactive opponent feedback, these defensive actions are dynamically triggered by the user: when the headset detects a jump (i.e., a change in the user's physical height), the virtual defender responds by jumping simultaneously, thereby creating visual pressure on the user (Fig.7(C)). If the user remains stationary for a period of time, the virtual defender automatically resets to a ready position about one step in front, preparing to respond to the user's next action.

7 User Studies

In this section, we present three system-level user studies (System User Study 1–3) to evaluate the effectiveness of our design for *shooting motion analysis* (Sec.5) and *simulated competition* (Sec.6).

7.1 Comparative Study 1: Motion Analysis

In this experiment, we aim to verify whether immersive observation enhances users' understanding of shooting motions. We also collected user feedback to evaluate the effectiveness of each motion visualization and data panel.

Baseline. We did not select raw video as the baseline due to uncontrollable variables, such as inherent differences from reconstructed scenes. Instead, our baseline was the 2D version of the shooting motion analysis module (Fig. 8(A)), which presents the same data and visualization forms as the immersive system. Users can adjust the camera's position (via keyboard) and angle (mouse). The only difference in presentation is the layout: the 2D system displays fixed panels on the screen, while the immersive version uses floating panels. Thus, the experimental variable is the overall presentation format—3D immersive environment vs. 2D display.

Experiment Settings. Our system runs on PICO 4, while the baseline system operates on a 16-inch laptop with a 2560x1600 resolution display (i7 12700H CPU, 16GB RAM, RTX3060 GPU). Each system contains 30 basketball shooting motions reconstructed from videos of two different players. We designed a structured questionnaire based on [42] to collect demographic data (e.g., age, gender) and subjective evaluations of the systems using a 7-point Likert scale. Participants rated statements for the learnability and usability of the visualizations on a scale from 1 (strongly disagree) to 7 (strongly agree), with all statements shown in Fig. 9(C, D). For each metric, we provided participants with a brief explanation. For example, *intuitiveness* refers to whether the motion-related visualizations matched participants' long-term experience and perception. The systems also track spatial positions (headset/camera coordinates in virtual space) and operations (number of camera angle changes for the baseline) to compare the differences in user behaviors between 2D and 3D situations. The logged spatial data is visualized with xy-coordinates in Fig. 9(A) and the z-coordinates in Fig. 9(B). The experiment took place in a 6m x 6m empty indoor room, providing enough space for free movement. After the experiment, all participants confirmed that they did not feel restricted by the space.

Procedures. We recruited 12 basketball enthusiasts from a university (Male=7, Female=5, Average Age=20.4 years, SD=2.54). Among the participants, 4 had more than 5 years of basketball experience,

(A) Baseline System (B) Immersive Environment

Figure 8: The 2D baseline and our immersive application.

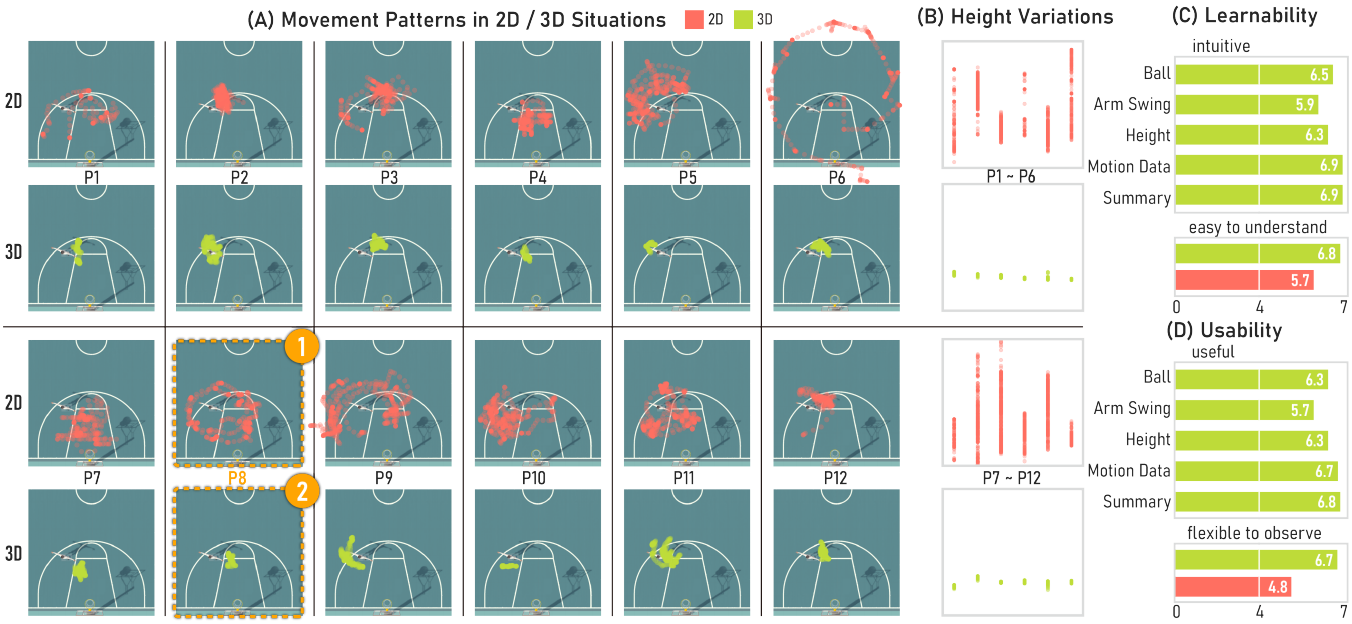


Figure 9: The results of our comparative study for shooting motion analysis. (A) displays participants' movement patterns on the XY plane while observing shooting motions on the court, (B) shows variations in participants' viewpoint height, and (C), (D) presents the participants' subjective ratings from the post-study survey.

5 had 2-5 years, and 3 had less than 2 years. All participants were new to using HMDs. Before the experiment, we taught them basic controller use, checked for VR discomfort, and familiarized them with both systems. Participants observed motions from different players using the Baseline and the HMD, respectively, and discuss their findings. The order of observation methods and player demonstrations was counterbalanced. After experiencing each system, participants immediately answered relevant questionnaire items. After experiencing both systems, we interviewed them using mostly open-ended, exploratory questions on the comparison between the two systems. Their responses were recorded and documented for analysis. The duration was approximately 30 minutes per participant, and compensation was in accordance with school standards.

Findings and Discussions. Drawing on the collected data and interviews, we summarized the following findings:

- **Participants tend to take on a different role when observing in 3D.** From the user behavior data, we noticed that in a 2D context, users changed their viewpoint much more. As a representative participant, P8 circled around the virtual player for observation in 2D (Fig. 9(A1)), whereas in 3D, he mostly stood beside the virtual player (Fig. 9(A2)). P8 mentioned in the interview: *"When looking at the screen, I felt like a spectator, so I kept moving the camera around to get different angles; but with the headset on, I felt more like a defender on the court, trying to see if I could block the shot."* This phenomenon is not isolated; in fact, similar patterns appeared across all users. Although users on 2D screens changed their viewpoint more frequently to closely examine the movements, their "easy to understand" ratings were still lower compared to the 3D motion demonstrations (Fig. 9(C)). This suggests that 3D motion demonstrations in immersive environments not only improve users' understanding of movements but also align better with their natural athletic habits.
- **Immersive environments offer a superior sense of motion by naturally aligning with the user's gaze.** As shown in Fig. 9(B), participants frequently adjusted camera height in 2D scenarios, whereas in 3D, the viewpoint remained mostly at head level. Based on user operations recorded from the 2D system, this was due to the frequent repositioning of the camera in the 2D screen to find optimal angles. In subjective feedback, 2D also scored much lower on flexibility compared to 3D (Fig. 9(C)). Therefore, we speculate that the immersive environment's perspective allows users to observe the reconstructed sports scenes as if they were present on-site, which aligns with their natural observation habits. This familiarity reduces cognitive load, allowing users to understand motions and indicators without constantly adjusting their view. As P10 mentioned: *"Viewing motions in 3D felt natural; I could see many details of the shooting motion just by following the virtual person with my gaze. But I am not used to operating the camera in 2D—it's inconvenient as the virtual player often jumps out of view when I zoom out."*
- **Visualizations directly related to key body movements may need to ideally align with the motion direction.** As shown in Fig. 9(C, D), while the Arm Swing Angle scored decently in intuitiveness and usefulness, it was the lowest-rated of all five visualizations. To seek the reasons, we asked P9, who rated Arm Swing Angle at 3 points, for her opinion: *"Although the angle moves with the arm and the body's rotation, it does not fully match the arm's swing trajectory, which confused me in the initial shooting motions."* This revealed that for complex motions like arm swings, which involve both circular (forearm rotation) and parabolic movements (player jumping), merely being close to the

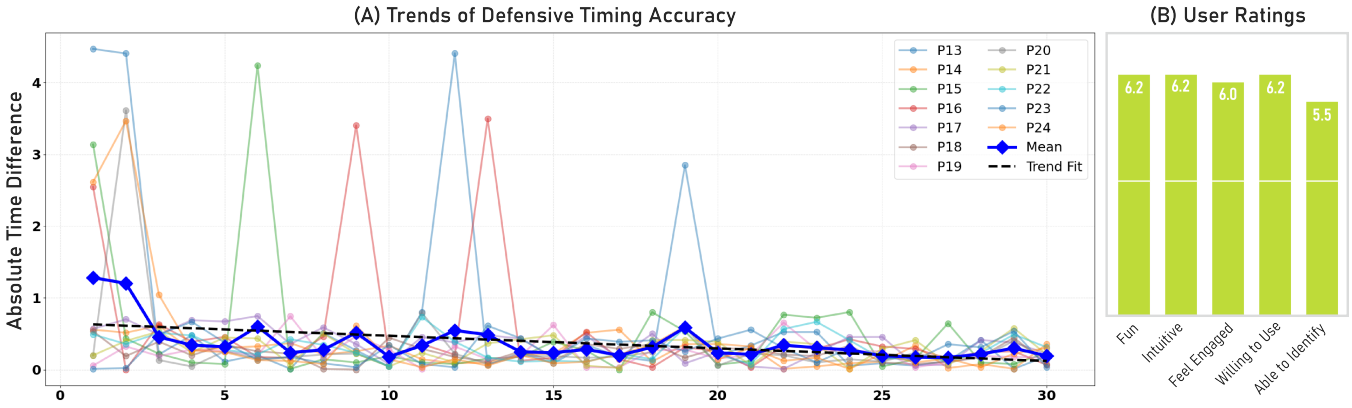


Figure 10: The results of study 2. (A) shows trends of participants' defensive timing accuracy. The blue line shows the mean trend and the black dashed line is the linear fit. (B) presents the participants' subjective ratings.

key body parts might be insufficient. Such visualizations can sometimes prevent observers from associating directly with the movement, thus complicating understanding. Moreover, the layout of visualizations directly linked to key body movements in motion demonstrations can vary based on different design considerations. Potential alternatives include displaying the visuals directly on the body, integrating them on the court, or projecting them onto a floating panel.

7.2 Effectiveness Study 2: Simulated Defense

This experiment aims to verify the effectiveness of interactions in the one-on-one immersive video training scenario, particularly in identifying the opponent's movement intentions.

Experiment Settings. To simulate the real-world process of competition preparation, our reconstruction pipeline used shooting videos from YouTube to recreate 3 fake shots [1] and 3 regular shots [2]. The virtual player's preparation time and shooting motions were randomized to prevent users from anticipating defensive actions, ensuring the training remained effective. During each defensive attempt, we recorded participants' defensive timing performance in the immersive system. The experiment was conducted indoors with enough space for free movement without interference.

Procedures. We recruited an additional 12 basketball enthusiasts from a university (Male=7, Female=5, Average Age=21.6 years, SD=3.73). Of the participants, 5 had over 5 years of basketball experience, 4 had 2-5 years, and 3 had less than 2 years. We familiarized participants with basic VR/MR operations before starting. Since the experiment involved physical movements, we followed school requirements to sign an informed consent form with each participant to ensure their voluntary participation. Participants completed 30 consecutive defense attempts against randomly played fake or direct shots reconstructed from videos. We counterbalanced the sequence of encountering defenders and gathered feedback through surveys and interviews afterward. The entire process took about 30 minutes per participant. Finally, we assessed the effectiveness of this training method by comparing changes in defensive timing judgment (Fig. 10(A)).

Findings and Discussions. Based on the recorded data and user feedback, we present the following findings:

- **The visual effects in simulated competition are effective but not sufficient.** As illustrated in (Fig. 10(A)), during the simulated defense practice, 10 users improved their defensive timing. Based on subjective ratings, they were generally better at distinguishing between fake and direct shots in head-to-head scenarios (scoring 5.5 out of 7 for 'Able to Identify'). As shown in Fig. 10(A), participants' defensive timing accuracy improved with short-term training. In P18's feedback, she mentioned, *"Holding the controller was uncomfortable during defense practice, and it didn't provide me with much physical feedback, like vibrating when I hit the hand or the ball, which made the defense outcome less perceptible to me."* Her feedback inspires us in two directions: in cases where hand tracking modules are not accurate enough, seamless devices with sensors, such as gloves equipped with position sensors, might offer a better experience; moreover, visual feedback should be integrated with physical feedback to enhance the overall training experience for users.
- **Real-time feedback on physical contact with virtual opponents enriches the defense practice experience.** From the participants' subjective ratings, most found the first-person perspective interaction with virtual characters to be easy to understand, with an average "intuitive" score of at least 6.2 out of 7 (Fig. 10(B)). The effectiveness of visual feedback methods, directly showcasing the results of physical contact on the virtual opponent's body, received particularly positive feedback. P14 mentioned in the interview: *"Marking the invaded position on the virtual player in red was very effective for my defense practice. My defensive style in basketball had always been aggressive and bold, and this method lets me know how to adjust my hand placement to avoid hitting the shooter's hand."* Overall, 4 out of 12 participants specifically emphasized its effectiveness and enjoyment during interviews, providing guidance for our future designs: interaction outcomes with virtual avatars should ideally be integrated with the data, not separated.

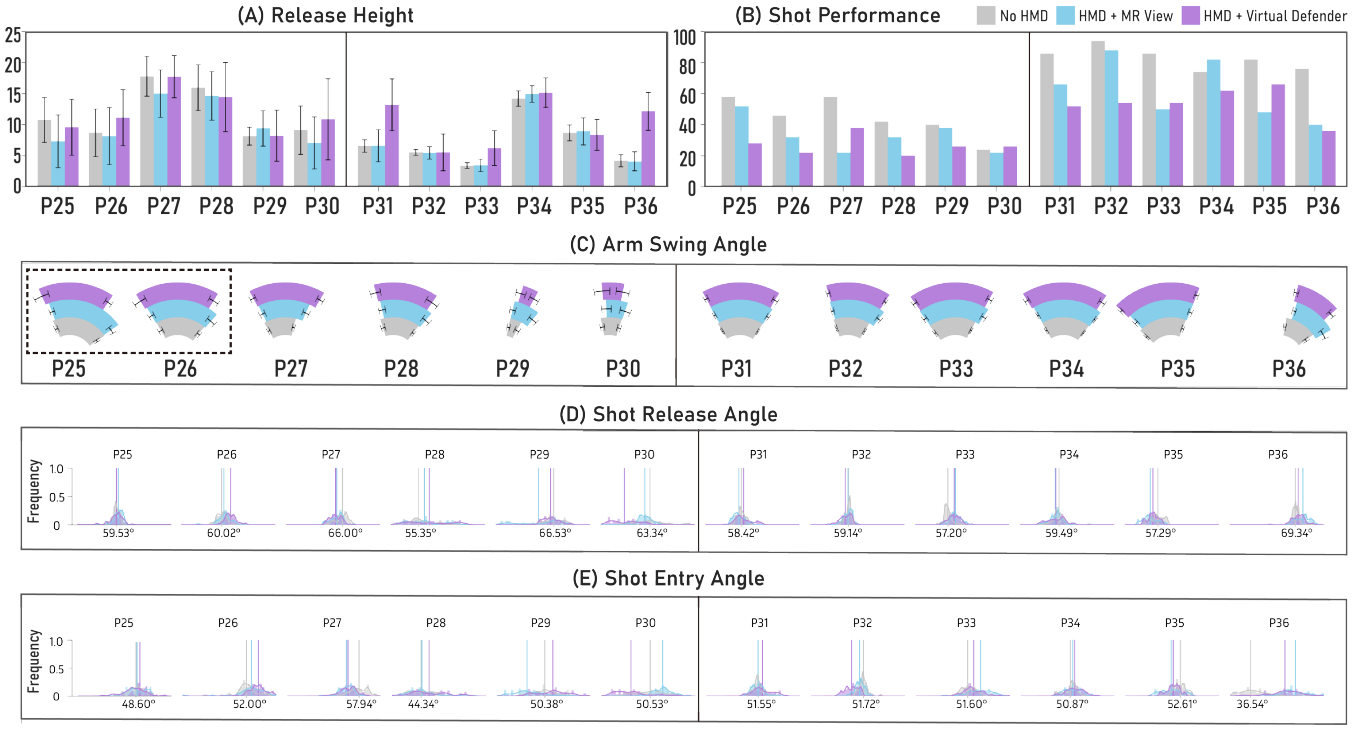


Figure 11: The results of our effectiveness study for simulated offense practice.

7.3 Effectiveness Study 3: Simulated Offense

This experiment aims to verify whether immersing users in video-based sports scenes can create a realistic competitive experience, thereby enhancing user engagement in training.

Experiment Settings. During the motion reconstruction process, we extracted several jump-defense actions from past game videos. When the system detects a user's jump (via a sudden increase in the HMD's physical height), the virtual defender jumps accordingly, replicating the defense from the video and adding visual pressure to the user's shot (Fig. 7(D)). In addition, the experiment took place on an indoor basketball court to ensure no interference.

Procedures. To ensure a diverse range of participant skill levels, we recruited 12 participants, comprising six basketball enthusiasts (Male=3, Female=3, Average Age=21.8 years, SD=1.86) and six semi-professional basketball players (Male=5, Female=1, Average Age=21.3 years, SD=2.05). They were divided into two groups, the enthusiasts (P25-P30) and the high-level players (P31-P36), for separate analysis. This grouping was essential for this experiment, which involved collecting participants' shooting motion data. For enthusiasts, performing a series of continuous shots was inherently challenging and their motions are usually unstable even without HMDs. This made it unclear whether the instability stemmed from the visual pressure of the virtual defender or their limited shooting skills. Hence, grouping participants by skill level allowed us to control for this variable, ensuring more accurate conclusions.

Prior to the experiment, participants completed a baseline task, which involved performing 50 jump shots from the same position on the court. These shots were recorded using a fixed camera setup to capture their shooting motions for analysis. After confirming sufficient recovery, participants were introduced to the MR environment via a head-mounted display (HMD) to adjust to the virtual setup. They then completed two additional sets of 50 jump shots under two conditions: (1) without a virtual defender (HMD + MR View) and (2) with a virtual defender (HMD + virtual defender). The order of conditions was randomized and counterbalanced. Sufficient rest was provided between shots to avoid fatigue and ensure consistent performance. All shooting sessions were recorded, and key motion metrics such as release height and arm swing angle were semi-automatically extracted using image processing algorithms. The distribution of shot release angle and shot entry angle across all participants' shots can be observed in Fig. 11(D, E). The total duration for each participant varied, ranging from 40 to 100 minutes. In the end, participants provided subjective feedback regarding their experience in both conditions.

Findings and Discussions. Drawing on the collected data and interviews, we summarized the following findings:

- **Both MR environment and virtual defender's reactions impacted the stability of users' shooting motions.** Based on the shot performance data (Fig. 11(B)), there was a noticeable drop in accuracy when using the MR headset, with an average decrease of 16.2%, and an additional decrease of 7.3% due to the presence of the virtual defender. Notably, the high-level group experienced more significant declines in accuracy (20.7% and 8.3%, respectively). We suspect this is because the MR environment and virtual defender's movements disrupted the more stable shooting motions of skilled players, making the impact more noticeable.

This can be seen in the standard deviations of their jump heights (Fig. 11(A)) and arm swing angles (Fig. 11(C)). For example, in normal jump shots without the headset, P31-P36 had much lower standard deviations in both jump height and arm swing angle, reflecting their consistent shooting performance (74%-94% accuracy without the HMD). However, after transitioning to the MR environment, their movements became significantly less stable, with P31-P36 showing the largest standard deviations in jump height under the HMD + virtual defender condition. In contrast, enthusiasts, who lack professional training and whose shooting form is inherently less stable, were less affected by the virtual environment and defensive actions.

- **Visual defensive pressure closely simulates real shooting competition, prompting strategic adjustments in trainees.** During player interviews, three high-level athletes (P31, P33, P34) mentioned, *"The virtual defender made me opt for a higher-arc shot."* P33 also noted, *"Just having the virtual defender there pushed me to intentionally use a higher arc, which was very effective for training."* This was reflected in their shooting release points, though the angle change was minimal for the high-level group (<1 degree) due to their consistent shooting form. In contrast, the amateur group showed greater variation. For example, P25 and P26 had noticeably earlier arm swing endpoints under the virtual defender condition compared to HMD + MR View (Fig. 11(C)). Overall, 10 out of 12 participants had earlier arm swing endpoints with the virtual defender, except for P27 and P35. These findings suggest that, even when users know the defender won't block the shot, visual pressure still induces realistic behavioral adaptation.

8 Discussion

In this section, we present the lessons learned and discuss the generalizability and limitations of our framework. Immersive techniques have been widely used to enhance training experiences by enabling intuitive perception of 3D motions. However, the motions captured in controlled lab environments are often disconnected from real-world context. Preserving this context is important for understanding the purpose of specific actions and supporting motion learning across sports. Besides basketball, this video processing pipeline, design study, and user evaluations can inform the development of motion learning systems in other domains.

8.1 Motion Related Data in Context - Motion Demonstration

Analyzing motion through video reconstruction is a growing trend [16, 35], particularly given the extensive foundational data analysis work already based on video. Compared to traditional motion capture methods, video-based techniques can integrate more seamlessly into data analysis workflows. For instance, Lin et al. [32] utilized 3D reconstruction to parse match footage and analyze badminton games on a virtual court, enhancing communication between coaches and players by providing actionable insights. A key future scenario in sports data analysis involves supporting multi-perspective, dynamic, and interactive forms of analysis that remain closely tied to the content. Additionally, some videos reference similar ones, such as professional athletes performing movements comparable to those of the user. By extracting body motion data from multiple videos and visualizing it through multiple avatars, analysts could gain richer and more detailed insights.

8.2 Interactive Experience with Virtual Opponents - Motion Reaction

Engaging in training against reconstructed virtual players is an innovative approach that allows athletes to familiarize themselves with opponents' signature moves and characteristics. For example, it can simulate how to evade a block by an NBA player with exceptional height and wingspan, to successfully complete a shot.

Our method visually simulates real-game scenarios to support motion reaction, but the lack of physical feedback still limits the sense of immersion. To address this, future work could explore integrating external devices, such as haptic feedback systems [23], to provide richer sensory experiences. Incorporating haptic feedback into immersive training or virtual confrontations has proven effective for enhancing realism [29]. In addition, we found that adopting motion generative models represents a highly promising direction for enhancing system reactivity. Such models could enable richer and more lifelike interactions by generating more realistic opponent responses—such as staggering or falling when physical contact occurs—and incorporating these into the virtual opponent's reactions. Introducing these advanced features will further enhance the training value of virtual sports scenarios.

8.3 Design Implications

Throughout our study, we gathered vital design implications.

First, motion data should ideally be displayed both as an overview and specifically linked to key body parts. In our first user experiment (Sec. 7.1), both motion data panels received high ratings for usability (over 6.7 points), outperforming other types of motion-related visualizations. This finding suggests that for complex movements (like an arm swing in shooting), visualization designs should closely follow the direction of the movement to enhance comprehension. In addition, providing specific numerical data as an overview to complement motion information could enhance perception and understanding of human motions in an immersive environment.

Second, in motion training, feedback integrated directly into the first-person view tends to be more effective than feedback presented separately. Despite the positive reception of physical contact visualizations in our second experiment, the DEF Result Panel received only average feedback during interviews. Two participants particularly noted its limited usefulness, as they typically focused on the immediate attempt and made adjustments based on recent attempts rather than overall performance. This indicates a preference for feedback that is directly relevant to the current action, suggesting that users are more engaged with immediate and actionable feedback than with a broad overview of their performance.

Third, an end-to-end framework would be more effective in minimizing the progressive accumulation of errors caused by multiple intermediate data processing steps. In the current pipeline for sports scene reconstruction, multiple steps are involved. For example, human motion and ball trajectories are processed in parallel and later aligned and adjusted during the final stage, which inevitably introduces manual errors. To reduce such cumulative errors, customized models tailored to specific scenarios can be more

effective. For instance, a model designed specifically for extracting data from basketball shooting events could integrate human motion and ball states simultaneously from a physics-based simulation perspective. This approach reduces intermediate steps, improving both consistency and accuracy.

8.4 Generalizability

We discuss the generalizability of the framework's inputs and outputs and explore ways to enhance it in future work.

Multi-modal Input: Although data extraction from monocular video can handle the majority of sports videos, this framework also supports the integration of more precise multi-modal data to enhance reconstruction results. Possible multi-modal approaches include using point cloud data from volumetric videos to improve the 3D representation of sports courts, enhancing immersion; adapting to incorporate assistive devices like position sensors during training to further improve user experience; and leveraging millimeter-wave radar for enhanced sports equipment status detection.

Broader Scenarios: The current immersive video training framework supports two scenarios: motion demonstration for observation or analysis, and motion reaction for interacting with a virtual player. This adaptability extends to both analytical and simulated competitive scenarios, highlighting the framework's robust applicability across various training contexts. In our future work, we envision remote coaching as a novel scenario. Specifically, coaches could assign various training tasks or even establish remote communication to provide real-time instructions to the trainees.

MR Performance Requirements: The suitability of MR systems in sports training is highly dependent on the specific requirements of different sports or training tasks. For instance, sports like table tennis require minimal system latency, while activities such as soccer benefit from a wider field of view. Even within the same sport, different drills may have varying requirements; for example, defensive movement training can tolerate more latency than one-touch shooting. Considering these variations, it is important for future research to address task-specific performance requirements when designing MR-based training systems.

Expanded Application: Our framework exhibits considerable scalability, effectively accommodating a wide range of sports activities, whether they involve significant use of equipment or not. For sports involving equipment, athletic performance is often reflected in the data related to the equipment, such as trajectory, speed, and rotational velocity. Our framework is already adaptable to various ball sports, such as football shooting or volleyball spiking [23]. Additionally, it supports the integration of different object detection algorithms or tracking technologies to extract equipment data. For certain racket sports, where ball status is also a key performance indicator, our framework is compatible. For instance, using TrackNet [25], we can capture badminton motion states and, after aligning them with human movement, apply the framework to immersive racket practice or even enhance shuttlecock state estimation to support feedback on striking performance. Despite its broad utility, it is important to acknowledge the specific challenges encountered in accurately reconstructing ball trajectories. These challenges suggest that adversarial sports, such as boxing—which emphasize the movement and interaction between athletes—may represent a more suitable and precise application domain for our framework.

8.5 Limitations

The limitations of our work fall into three main areas. **First, there is a lack of detail in the motion reconstruction, particularly in the capture of hand movements.** This was demonstrated in our Module Usability Study 1, where participants acknowledged the overall accuracy of the reconstructed scene, but pointed out shortcomings in the representation of fingers. This suggests our framework may face challenges in sports that emphasize fine hand movements, highlighting the need for improved motion capture techniques to achieve higher fidelity. **Second, current MR headsets present issues such as spatial distortion and physical discomfort, which can affect athletic performance, especially for high-level users.** For example, one highly experienced participant (P12) reported discrepancies between the virtual and actual positions of the basketball, with perceived distance varying during head movement, which could impact performance. The weight and fit of MR headsets can also be a burden during sports training. The significant decline in shooting performance among high-level participants after wearing the HMD in System User Study 3 provides further evidence for this limitation. While current HMDs may constrain physical performance, MR offers distinct advantages for situational awareness and visual decision-making—such as reading fakes—which are critical in sports. This presents a trade-off between immersive visual feedback and physical freedom. Future improvements in HMD technology, including lighter design and reduced latency or distortion, could help address these issues and enhance the training experience. **Third, our user study was limited in scale.** Rather than aiming for a comprehensive evaluation, our primary goal was to demonstrate the feasibility of using video-based motion analysis and immersive interactions in sports training, and to gather early but meaningful insights. Potential extensions include collaboration with sports teams for longitudinal tracking of physical metrics, enabling a more comprehensive evaluation of training effectiveness.

9 Conclusion

In conclusion, our work leverages advanced 3D motion reconstruction to create immersive training experiences, enabling users to engage with video sports scenes from a first-person perspective and interact with virtual opponents. With expert input, we identified key performance indicators and developed immersive visualizations and a simulated one-on-one matchup. User studies validated our approach, demonstrating its potential to enhance comprehension and engagement in basketball shooting. In the future, we plan to extend this framework to other sports, such as boxing and football.

Acknowledgments

This work was supported by NSFC (62421003, 62402428, U22A2032, 62402437) and partially supported by ZJU Kunpeng&Ascend Center of Excellence. The author also gratefully acknowledges the support of Zhejiang University Education Foundation Qizhen Scholar Foundation.

References

[1] 2019. How to: DEADLY Basketball FAKE Moves to Beat your Defender EVERY TIME! https://www.youtube.com/watch?v=zE4LpglnEGQ. Accessed January 15, 2024.

[2] 2020. How to: NBA Shooting Secrets to Improve Your Jump Shot! https://www.youtube.com/watch?v=BWfoUrNqKNM. Accessed January 15, 2024.

[3] 2021. EasyMoCap - Make human motion capture easier. https://github.com/zju3dv/EasyMocap. Accessed January 15, 2024.

[4] 2023. MediaPipe. https://google.github.io/mediapipe/.

[5] 2024. Mixamo. https://www.mixamo.com/.

[6] 2024. PICO. https://www.picoxr.com/sg.

[7] 2025. Catapult. https://www.catapult.com/.

[8] 2025. SkyCoach. https://www.myskycoach.com/.

[9] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: Enhancing movement training with an augmented reality mirror. In *Proceedings of the ACM symposium on User Interface Software and Technology*. 311–320.

[10] Dimitrije Cabarkapa, Damjana V Cabarkapa, Nicolas M Philipp, Chloe A Myers, Shay M Whiting, Grant T Jones, and Andrew C Fry. 2023. Kinematic differences based on shooting proficiency and distance in female basketball players. *Journal of Functional Morphology and Kinesiology* 8, 3 (2023), 129.

[11] Jacky CP Chan, Howard Leung, Jeff KT Tang, and Taku Komura. 2010. A virtual reality dance training system using motion capture technology. *IEEE Transactions on Learning Technologies* 4, 2 (2010), 187–195.

[12] David Checa and Andres Bustillo. 2020. A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications* 79, 9 (2020), 5501–5527.

[13] Hua-Tsung Chen, Yu-Zhen He, and Chun-Chieh Hsu. 2018. Computer-assisted yoga training system. *Multimedia Tools and Applications* 77 (2018), 23969–23991.

[14] Xin Chen, Anqi Pang, Wei Yang, Yuexin Ma, Lan Xu, and Jingyi Yu. 2021. Sportscap: Monocular 3d human motion capture and fine-grained understanding in challenging sports videos. *International Journal of Computer Vision* 129 (2021), 2846–2864.

[15] Xiangtong Chu, Xiao Xie, Shuainan Ye, Haolin Lu, Hongguang Xiao, Zeqing Yuan, Zhutian Chen, Hui Zhang, and Yingcai Wu. 2021. TIVEE: Visual exploration and explanation of badminton tactics in immersive visualizations. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 118–128.

[16] Christopher Clarke, Doga Cavdir, Patrick Chiu, Laurent Denoue, and Don Kimber. 2020. Reactive video: Adaptive video playback based on user motion for supporting physical activity. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. 196–208.

[17] Yann Desmarais, Denis Mottet, Pierre Slangen, and Philippe Montesinos. 2021. A review of 3D human pose estimation algorithms for markerless motion capture. *Computer Vision and Image Understanding* 212 (2021), 103275.

[18] Bhat Dittakavi, Divyagna Bavikadi, Sai Vikas Desai, Soumi Chakraborty, Nishant Reddy, Vineeth N Balasubramanian, Bharathi Callepalli, and Ayon Sharma. 2022. Pose Tutor: An explainable system for pose correction in the wild. In *Proceedings of the CVF Conference on Computer Vision and Pattern Recognition*. 3540–3549.

[19] Meng Du and Xiaoru Yuan. 2021. A survey of competitive sports data visualization and visual analysis. *Journal of Visualization* 24 (2021), 47–67.

[20] Mihai Fieraru, Mihai Zanfir, Silviu Cristian Pirlea, Vlad Olaru, and Cristian Sminchisescu. 2021. AIFit: Automatic 3d human-interpretable feedback models for fitness training. In *Proceedings of the CVF Conference on Computer Vision and Pattern Recognition*. 9919–9928.

[21] Ze Gao, Anqi Wang, Pan Hui, and Tristan Braud. 2022. Meditation in Motion: Interactive media art visualization based on ancient Tai Chi Chuan. In *Proceedings of the ACM International Conference on Multimedia*. 7241–7242.

[22] Shubham Goel, Georgios Pavlakos, Jathushan Rajasegaran, Angjoo Kanazawa, and Jitendra Malik. 2023. Humans in 4D: Reconstructing and tracking humans with transformers. In *Proceedings of the CVF International Conference on Computer Vision*. 14783–14794.

[23] Ut Gong, Hanze Jia, Yujie Wang, Tan Tang, Xiao Xie, and Yingcai Wu. 2024. VolleyNaut: Pioneering immersive training for inclusive sitting volleyball skill development. In *Proceedings of the IEEE Conference On Virtual Reality and 3D User Interfaces*. 1022–1032.

[24] William Huang, Sam Ghahremani, Siyou Pei, and Yang Zhang. 2024. WheelPose: Data Synthesis Techniques to Improve Pose Estimation Performance on Wheelchair Users. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–25.

[25] Yu-Chuan Huang, I-No Liao, Ching-Hsuan Chen, Tsì-Uí Ík, and Wen-Chih Peng. 2019. TrackNet: A Deep Learning Network for Tracking High-speed and Tiny Objects in Sports Applications. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*.

[26] Mike D Hughes and Roger M Bartlett. 2002. The use of performance indicators in performance analysis. *Journal of Sports Sciences* 20, 10 (2002), 739–754.

[27] Felix Hülsmann, Jan Philip Göpfert, Barbara Hammer, Stefan Kopp, and Mario Botsch. 2018. Classification of motor errors to provide real-time feedback for sports coaching in virtual reality—A case study in squats and Tai Chi pushes. *Computers & Graphics* 76 (2018), 47–59.

[28] Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. 2018. End-to-End Recovery of Human Shape and Pose. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[29] Mohammad Amin Kuhail, Areej ElSayary, Shahbano Farooq, and Ahlam Alghamdi. 2022. Exploring Immersive Learning Experiences: A Survey. In *Informatics*, Vol. 9. 75.

[30] Zhihao Li, Jianzhuang Liu, Zhensong Zhang, Songcen Xu, and Youliang Yan. 2022. CLIFF: Carrying location information in full frames into human pose and shape estimation. In *European Conference on Computer Vision (ECCV)*. Springer, 590–606.

[31] Kevin Lin, Lijuan Wang, and Zicheng Liu. 2021. End-to-End human pose and mesh reconstruction with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[32] Tica Lin, Alexandre Aouididi, Zhutian Chen, Johanna Beyer, Hanspeter Pfister, and Jui-Hsien Wang. 2024. VIRD: Immersive match video analysis for high-performance badminton coaching. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 458–468.

[33] 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 SIGCHI Conference on Human Factors in Computing Systems*. 1–13.

[34] Tica Lin, Yalong Yang, Johanna Beyer, and Hanspeter Pfister. 2020. SportsXR-Immersive analytics in sports. *arXiv preprint arXiv:2004.08010* (2020).

[35] Jingyuan Liu, Nazmus Saquib, Zhutian Chen, Rubaiat Habib Kazi, Li-Yi Wei, Hongbo Fu, and Chiew-Lan Tai. 2022. PoseCoach: a customizable analysis and visualization system for video-based running coaching. *IEEE Transactions on Visualization and Computer Graphics* (2022), To appear.

[36] Jingyuan Liu, Li-Yi Wei, Ariel Shamir, and Takeo Igarashi. 2024. iPose: Interactive Human Pose Reconstruction from Video. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14.

[37] Juan Liu, Yawen Zheng, Ke Wang, Yulong Bian, Wei Gai, and Dingyuan Gao. 2020. A real-time interactive Tai Chi learning system based on VR and motion capture technology. *Procedia Computer Science* 174 (2020), 712–719.

[38] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A skinned multi-person linear model. *ACM Trans. Graph.* 34, 6 (2015), 16.

[39] Dizhi Ma, Xiyun Hu, Jingyu Shi, Mayank Patel, Rahul Jain, Ziyi Liu, Zhengzhe Zhu, and Karthik Ramani. 2024. avaTTar: Table tennis stroke training with embodied and detached visualization in augmented reality. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. 1–16.

[40] Laura Marchal-Crespo, Mark van Raai, Georg Rauter, Peter Wolf, and Robert Riener. 2013. The effect of haptic guidance and visual feedback on learning a complex tennis task. *Experimental Brain Research* 231 (2013), 277–291.

[41] Dan Mikami, Mariko Isogawa, Kosuke Takahashi, Hideaki Takada, and Akira Kojima. 2015. Immersive previous experience in VR for sports performance enhancement. In *Proceedings of the International Congress on Sport Sciences Research and Technology Support*.

[42] Heather L O'Brien and Elaine G Toms. 2010. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology* 61, 1 (2010), 50–69.

[43] Stefan Pastel, Katharina Petri, Chien-Hsi Chen, Ana Milena Wiegand Cáceres, Meike Stirnatis, Carlo Nübel, Lasse Schlotter, and K Witte. 2023. Training in virtual reality enables learning of a complex sports movement. *Virtual Reality* 27, 2 (2023), 523–540.

[44] Katerina El Raheb, George Tsampounaris, Akrivi Katifori, and Yannis Ioannidis. 2018. Choreomorphy: A whole-body interaction experience for dance improvisation and visual experimentation. In *Proceedings of the International Conference on Advanced Visual Interfaces*. 1–9.

[45] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real-time object detection. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*. 779–788.

[46] Patrick Reipschläger, Frederik Brudy, Raimund Dachselt, Justin Matejka, George Fitzmaurice, and Fraser Anderson. 2022. AvatAR: An immersive analysis environment for human motion data combining interactive 3d avatars and trajectories. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–15.

[47] Alessandra Semeraro and Laia Turmo Vidal. 2022. Visualizing instructions for physical training: Exploring visual cues to support movement learning from instructional videos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1–16.

[48] Mingyi Shi, Kfir Aberman, Andreas Aristidou, Taku Komura, Dani Lischinski, Daniel Cohen-Or, and Baoquan Chen. 2020. MotioNet: 3D human motion reconstruction from monocular video with skeleton consistency. *ACM Trans. Graph.* 40, 1 (2020), 15 pages.

[49] Qing Shuai, Chen Geng, Qi Fang, Sida Peng, Wenhao Shen, Xiaowei Zhou, and Hujun Bao. 2022. Novel view synthesis of human interactions from sparse

multi-view videos. In *Proceedings of the ACM SIGGRAPH Conference*. 1–10.

[50] Pooya Soltani and Antoine HP Morice. 2020. Augmented reality tools for sports education and training. *Computers & Education* 155 (2020), 103923.

[51] Peng Song, Shuhong Xu, Wee Teck Fong, Ching Ling Chin, Gim Guan Chua, and Zhiyong Huang. 2012. An immersive VR system for sports education. *IEICE Transactions on Information and Systems* 95, 5 (2012), 1324–1331.

[52] Marina Stergiou, Katerina El Raheb, and Yannis Ioannidis. 2019. Imagery and metaphors: From movement practices to digital and immersive environments. In *Proceedings of the International Conference on Movement and Computing*. 1–8.

[53] M. Supej and J. Spörri. 2021. *Sports performance and health*. MDPI AG, Chapter 6, 73–84.

[54] István Sárándi, Timm Linder, Kai O. Arras, and Bastian Leibe. 2018. How robust is 3D human pose estimation to occlusion? arXiv:1808.09316 [cs.CV] https://arxiv.org/abs/1808.09316

[55] Richard Tang, Xing-Dong Yang, Scott Bateman, Joaquim Jorge, and Anthony Tang. 2015. Physio@Home: Exploring visual guidance and feedback techniques for physiotherapy exercises. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 4123–4132.

[56] Wan-Lun Tsai, Tse-Yu Pan, and Min-Chun Hu. 2020. Feasibility study on virtual reality based basketball tactic training. *IEEE Transactions on Visualization and Computer Graphics* 28, 8 (2020), 2970–2982.

[57] Georgios Tsampounaris, Katerina El Raheb, Vivi Katifori, and Yannis Ioannidis. 2016. Exploring visualizations in real-time motion capture for dance education. In *Proceedings of the Pan-Hellenic Conference on Informatics*. 1–6.

[58] Chongyang Wang, Siqi Zheng, Lingxiao Zhong, Chun Yu, Chen Liang, Yuntao Wang, Yuan Gao, Tin Lun Lam, and Yuanchun Shi. 2024. PepperPose: Full-Body Pose Estimation with a Companion Robot. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.

[59] Jianbo Wang, Kai Qiu, Houwen Peng, Jianlong Fu, and Jianke Zhu. 2019. AI coach: Deep human pose estimation and analysis for personalized athletic training assistance. In *Proceedings of the ACM International Conference on Multimedia*. 374–382.

[60] Erwin Wu, Takayuki Nozawa, Florian Perteneder, and Hideki Koike. 2020. VR alpine ski training augmentation using visual cues of leading skier. In *Proceedings of the CVF Conference on Computer Vision and Pattern Recognition Workshops*. 878–879.

[61] Erwin Wu, Mitski Piekenbrock, Takuto Nakumura, and Hideki Koike. 2021. SPinPong-virtual reality table tennis skill acquisition using visual, haptic and temporal cues. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2566–2576.

[62] Yihong Wu, Lingyun Yu, Jie Xu, Dazhen Deng, Jiachen Wang, Xiao Xie, Hui Zhang, and Yingcai Wu. 2023. AR-Enhanced Workouts: Exploring visual cues for at-home workout videos in AR environment. In *Proceedings of ACM Symposium on User Interface Software and Technology*.

[63] Vickie Ye, Georgios Pavlakos, Jitendra Malik, and Angjoo Kanazawa. 2021. Decoupling human and camera motion from videos in the Wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[64] Hongwen Zhang, Yating Tian, Yuxiang Zhang, Mengcheng Li, Liang An, Zhenan Sun, and Yebin Liu. 2023. PyMAF-X: Towards well-aligned full-body model regression from monocular images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).

[65] Maoqing Tian Jianbo Liu Shuai Yi Hongsheng Li Ziniu Wan, Zhengjia Li. 2021. Encoder-decoder with multi-level attention for 3D human shape and pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 13033–13042.

[66] Liyuan Zou, Takatoshi Higuchi, Haruo Noma, Lopez-Gulliver Roberto, and Tadao Isaka. 2019. Evaluation of a virtual reality-based baseball batting training system using instantaneous bat swing information. In *Proceedings of the International Conference on Virtual Reality and 3D User Interfaces (VR)*. 1289–1290.