

# Applications of motion capture technology in VR metaverse fidelity virtual human interaction systems

*Yu Pu*

School of theater film and television, Communication University of China, Beijing, China

puyu8671@gmail.com

---

**Abstract.** With the rise of Virtual Reality (VR) and metaverse concepts, high-fidelity virtual human interaction systems have become a hot research topic. Motion capture technology, as a key bridge connecting the real world and virtual environment, plays a central role in driving avatars in the VR metaverse. The purpose of this paper is to explore the application of motion capture technology in constructing a high-fidelity virtual human interaction system in the VR metaverse. By reviewing the technical classification of motion capture technology, comparing and studying its core breakthrough points, and analyzing its innovative principles in the system architecture model, this paper mainly focuses on the key features of low latency, high fidelity, and multimodality. The research results show that motion capture technology can effectively capture and reconstruct user movements, realize real-time driving of virtualized bodies, and enhance the immersion and realism of VR interaction. Meanwhile, to address the high cost of current motion capture technology, the research is committed to exploring the use of fewer sensors to realize whole-body motion reconstruction, which will provide a more economical and practical solution for future VR applications.

**Keywords:** motion capture, virtual reality, metaverse, virtual humans, human-computer interaction

---

## 1. Introduction

In recent years, emerging technologies such as Virtual Reality (VR), Augmented Reality (AR), and metaverse have been developing rapidly, and high-fidelity virtual human interaction systems have become a research area of great interest. In these virtual environments, users can interact, communicate, and collaborate with avatars for a more immersive and personalized experience. Motion capture technology, as a key bridge between the real world and virtual environments, plays a crucial role in capturing and reconstructing user movements and driving avatars in the VR metaverse. Despite the promising applications of motion capture technology in VR, there are still some research gaps and challenges. For example, how to further reduce the cost of motion capture, improve its application effect in VR environments, and how to realize a more natural and smooth virtual human interaction experience have yet to be thoroughly investigated. The purpose of this paper is to explore the application of motion capture technology to build a high-fidelity virtual human interaction system in the VR metaverse. Specifically, this paper will focus on the following issues: (1) the technical classification of motion capture technology and its advantages and disadvantages; (2) the key technologies of motion capture technology to achieve low-latency, high-fidelity, and multimodal interactions in the VR metaverse; and (3) how to optimize the motion capture system to reduce its cost and improve its performance, so as to provide users with a more economical and practical VR interaction solution. In order to solve the above problems, this paper will adopt the methods of literature review, system modeling and case study to study in depth the current status and development trend of the application of motion capture technology in the VR metaverse. By analyzing the existing research results and technical solutions, we will summarize the advantages and shortcomings of motion capture technology in VR interaction applications and put forward suggestions for improvement. The significance of this study is to provide a theoretical foundation and technical reference for the development of human-computer interaction in the VR metaverse. Through in-depth research on the application of motion capture technology in VR environments, it can provide more economical and practical solutions for future VR applications, thus promoting the popularization and development of VR technology. In addition, this study can also provide reference for researchers in related fields to promote the continuous innovation and development of VR human-computer interaction technology.

## 2. Overview of the application of motion capture technology

### 2.1. Technical classification and comparative study of motion capture technology

Motion capture technology, also known as motion capture, is a technology that records and analyzes the motion of the human body or objects. It captures motion data through a variety of sensors or vision systems and converts these data into digital information, which is used in computer animation, games, virtual reality, motion analysis and other fields.

According to the different capture principles and technical means, motion capture technology can be divided into the following categories: Optical Motion Capture: By setting markers on the human body or objects, multiple cameras are used to capture the three-dimensional position information of these markers from different angles. Then, these data are processed and analyzed by computer vision algorithms to reconstruct the motion trajectory of the human body or object. Optical motion capture has the advantages of high accuracy and good stability, but it requires specific sites and equipment, and is easily affected by occlusion and lighting conditions. There are also inertial motion capture methods, which use Inertial Measurement Units (IMUs) to capture motion data. IMUs contain sensors such as accelerometers, gyroscopes, and magnetometers, which measure the acceleration, angular velocity, and attitude information of an object. Multiple IMUs are fixed to key parts of a human body or object to capture its motion data in real time. Inertial motion capture has the advantages of being small in size, light in weight, and not requiring a specific site, but the accuracy is relatively low and susceptible to drift and cumulative errors [1]. Secondly, there is the electromagnetic motion capture method, which captures motion data by emitting electromagnetic fields and using sensors to measure the changes in the electromagnetic fields. Electromagnetic motion capture has the advantages of higher accuracy and lower latency, but it is easily interfered by metal objects and has a limited capture range.

More often, there is also acoustic motion capture, which captures motion data by emitting ultrasonic waves and using multiple microphones to receive the reflected ultrasonic signals. Acoustic motion capture has the advantages of lower cost and ease of deployment, but is less accurate and susceptible to noise and environmental factors. There is also hybrid motion capture which combines a variety of motion capture technologies to give full play to their respective advantages and improve capture accuracy and stability. For example, optical motion capture and inertial motion capture can be combined to utilize the precision of optical motion capture and the flexibility of inertial motion capture to achieve more accurate and stable motion capture. Different types of motion capture technologies have their own advantages and disadvantages and are suitable for different application scenarios. When choosing motion capture technology, it is necessary to comprehensively consider factors such as accuracy, cost, stability, and scope of application according to actual needs.

### 2.2. Core breakthrough points of motion capture technology

The core breakthroughs of motion capture technology mainly focus on the following aspects: (1) high-precision motion tracking: improving the precision of motion tracking is one of the important goals of motion capture technology. High-precision motion tracking can more realistically restore the movement details of the human body or objects and improve the immersion and realism of the virtual environment. In order to improve the accuracy of motion tracking, researchers have continuously improved sensor technology, optimized data processing algorithms, and adopted multiple sensor fusion methods [2]. (2) Low-latency data transmission: low-latency data transmission is the key to realizing real-time interaction. In applications that require real-time interaction, such as the VR metaverse, the motion capture system needs to be able to quickly transfer the captured motion data to the computer for processing and rendering to ensure that the avatar can respond to the user's actions in real time. To reduce the latency of data transmission, researchers have used high-speed wireless communication technologies, optimized data compression algorithms, and improved system architectures [3]. (3) Multimodal data fusion: multimodal data fusion refers to the fusion of data from different sensors or different modalities to obtain more comprehensive and accurate motion information. For example, visual data, inertial data, force feedback data, etc. can be fused to more realistically restore the motion state of a human body or object. Multimodal data fusion can improve the robustness and adaptability of the motion capture system so that it can adapt to a variety of complex environments and application scenarios. (4) Intelligent data processing: intelligent data processing refers to the use of artificial intelligence technology to process and analyze motion data to achieve more advanced functions. For example, machine learning algorithms can be used to classify, identify, and predict motion data to achieve functions such as motion recognition, posture estimation, and behavior analysis. Intelligent data processing can improve the degree of automation and intelligence of the motion capture system so that it can better serve various application areas.

## 3. Motion capture and high-fidelity virtual human generation

### 3.1. System architecture model

A typical high-fidelity virtual human generation system based on motion capture usually includes the following modules:

- (1) Motion Capture Module: responsible for capturing the motion data of human body or objects. This module can use various types of motion capture technology, such as optical, inertial, electromagnetic, etc.
- (2) Data transmission module: responsible for transmitting the captured motion data to a computer for processing. This module can use wired or wireless communication technologies such as USB, Ethernet, Wi-Fi, Bluetooth etc.
- (3) Data processing module: responsible for processing and analyzing the received motion data. This module usually includes steps, such as data filtering, data correction, data fusion, and attitude estimation.
- (4) Virtual human modeling module: responsible for constructing the three-dimensional model of the virtual human. This module can use a variety of 3D modeling software, such as Maya, 3Ds Max, Blender, and so on.
- (5) Animation driver module: responsible for applying the processed motion data to the virtual human model, driving the virtual human animation. This module usually uses skeleton animation technology to map the motion data to the skeleton of the virtual human, so as to control the motion of the virtual human.
- (6) VR Rendering Module: Responsible for rendering the animation of the virtual human into the VR environment and presenting it to the user. This module usually uses a VR engine, such as Unity, Unreal Engine, and so on.

### 3.2. Core innovation principle

The core innovation principle of the high-fidelity virtual human generation system based on motion capture is mainly reflected in the following aspects: Low Latency: In order to realize real-time interaction, the system needs to reduce the latency as much as possible. This requires optimization between various modules, including sensor selection, data transmission, data processing, animation driving and VR rendering. High fidelity: In order to improve the realism of the virtual human, the system needs to restore the motion details of the human body or object as much as possible. This requires optimization in motion capture, data processing, virtual human modeling and animation driving. Multimodal: To improve the robustness and adaptability of the system, the system can fuse data from different sensors or different modalities. For example, visual data, inertial data, force feedback data, etc. can be fused to more realistically restore the motion state of the human body or object. The analysis of interaction methods in virtual reality technology shows that developers are committed to providing users with complete and realistic interactive experiences, such as transmitting sensations, such as touch, smell, sight, and hearing, to the user's brain and restoring the user's perception of the real world in the first place [4]. Applied avatars can provide a variety of experiential environments and experiences that enhance the immersion of virtual environments and provide new experiences that both non-immersive users (using PCs and mobile devices) and immersive users (using Virtual Reality (VR)) can participate in and execute [5].

## 4. Application prospects and cases- VR interactive applications

Motion capture technology has a broad application prospect in VR interactive applications. The following are some typical applications:

- (1) VR games: Motion capture technology can provide a more realistic and immersive gaming experience for VR games. Players can map their movements to the virtual characters in the game through motion capture devices, thus realizing more natural and smooth game operations.
- (2) VR Social: Motion capture technology can provide a more realistic and vivid social experience for VR social. Users can transfer their expressions, movements, and gestures to the avatars in the VR environment through motion capture devices so as to communicate with other users more realistically and naturally.
- (3) VR Education: Motion capture technology can provide a more intuitive and interactive learning experience for VR education. Students can interact with virtual teachers or virtual classmates in the VR environment through motion capture devices, thus gaining a deeper understanding of the learning content.
- (4) VR Medical: Motion capture technology can provide safer and more effective medical solutions for VR medical. Doctors can perform remote surgery, rehabilitation training and other operations through motion capture devices, thus improving medical efficiency and quality. In the context of Industry 4.0, Collaborative Robots (CR) are undergoing a paradigm shift from traditional industrial manipulators to collaborative robots, with the latter serving humans more and more closely as auxiliary tools in many production processes.

In this context, continued technological advances offer new opportunities for further innovation in robotics and other areas of next-generation industry. For example, 6G can play an important role due to its human-centered industrial field perspective [6]. Immersion can be enhanced by a full-body sense-capture VR telecollaboration system that uses VR characters to mirror the user's movements, thus minimizing mismatch issues [7].

## 5. Conclusion

Motion capture technology serves as a critical bridge connecting the real world with virtual environments, playing an increasingly vital role in the VR metaverse. This paper proposes a method for generating immersive virtual environments by applying motion capture-based avatars, aiming to enhance the immersiveness of virtual environments and provide new experiences, enabling both non-immersive users (using PCs and mobile devices) and immersive users (using Virtual Reality (VR)) to participate in and execute various experiences. The research focuses on the innovative application of motion capture technology in virtual reality interactive applications, optimizing system architecture to reduce latency, enhancing the high fidelity of virtual avatars, and integrating multimodal data, with the goal of providing users with more economical and practical VR interactive solutions. As technology continues to evolve and innovate, motion capture technology will demonstrate even broader application prospects in VR interactive applications. Future research directions include improving motion tracking accuracy, reducing data transmission latency, integrating multimodal data, and achieving intelligent data processing. Through continuous technological innovation, motion capture technology will bring more realistic, immersive, and intelligent interactive experiences to the VR metaverse.

## References

- [1] Qiu, S., Zhao, H., Jiang, N., Wu, D., Song, G., & Zhao, H. (2021). Sensor network oriented human motion capture via wearable intelligent system. *International Journal of Intelligent Systems*, 37(2), 1646–1673. <https://doi.org/10.1002/int.22678>
- [2] Bai, Y., Xi, J., & Liao, W. (2021). Design and implementation of full-body motion capture system based on multi-sensor fusion. In Z. Zhang (Ed.), *Proceedings of the 2021 International Conference on Neural Networks, Information and Communication Engineering* (p. 68). SPIE.
- [3] Ha, E., Byeon, G., & Yu, S. (2022). Full-body motion capture-based virtual reality multi-remote collaboration system. *Applied Sciences*, 12(12), 5862. <https://doi.org/10.3390/app12125862>
- [4] Liu, Y. (2023). Analysis of interaction methods in VR virtual reality. *HSET*, 39, 395–407.
- [5] Park, M., Cho, Y., Na, G., & Kim, J. (2023). Application of virtual avatar using motion capture in immersive virtual environment. *International Journal of Human-Computer Interaction*, 40(20), 6344–6358. <https://doi.org/10.1080/10447318.2023.2246826>
- [6] Calandra, D., Praticò, F. G., Cannavò, A., Casetti, C., & Lamberti, F. (2024). Digital twin- and extended reality-based telepresence for collaborative robot programming in the 6G perspective. *Digital Communications and Networks*, 10(2), 315–327. <https://doi.org/10.1016/j.dcan.2023.11.002>
- [7] Liu, G., & Tsai, S. B. (2022). Online optimization of animation art design user virtual perception in mobile edge computing environment. *Scientific Programming*, 2022(1), Article 8970763. <https://doi.org/10.1155/2022/8970763>