ISSN 2616-5767 Vol.7, Issue 6: 177-181, DOI: 10.25236/AJETS.2024.070625

Innovative Design and Research on XR Interaction System Based on CV Modal Handle Tracking Algorithm

Qi Zhang[1,a], Ning Mao[1,b], Jiayi Shen[1,c], Jingxiu Li[1,d], Guwei Li[1,e,*]

[1]Artificial Intelligence College, Zhejiang Dongfang Polytechnic, Wenzhou, China
[a]364940760@qq.com, [b]2731832296@qq.com, [c]1097658283@qq.com, [d]18357655299@163.com, [e]3168897@qq.com
**Corresponding author*

Abstract: *Traditional XR controllers have problems such as low precision, poor stability and low accuracy during the miniaturization process. To solve the above problems, an XR interaction system based on the CV (Computer Vision) modal handle tracking algorithm is proposed. Research and tests show that this system not only provides new ideas for solving the tracking problems faced by traditional controllers during the miniaturization process, but also makes XR interactions more natural and smooth.*

Keywords: *CV Modal, Innovative, Tracking Algorithm*

1. Introduction

The rapid development of virtual reality technology has continuously increased users' demand for immersive experiences. And the XR (Extended Reality) interaction system, as the key to realizing immersive experiences, has made its design and optimization become the focus of current research. With the continuous progress of science and technology, traditional controllers are facing numerous challenges during the miniaturization process. On the one hand, miniaturized devices limit the size and quantity of sensors, making high-precision tracking difficult. On the other hand, when using miniaturized controllers, users may make rapid and complex gestures more frequently, which puts forward higher requirements for the stability and accuracy of the tracking system.

The XR interaction system based on the CV modal handle tracking algorithm provides new ideas for solving these problems. This system combines the optical tracking algorithm with the CV modal bare-hand tracking algorithm. When the LED lights on the handle are blocked, it can accurately restore the 6DoF (Six Degrees of Freedom) information of the handle controller by tracking the features of the bare hand, thus maintaining stable tracking of the handle[1]. This integrated solution not only solves the problem of difficult recognition of miniaturized handles under the condition of occlusion but also offers new ideas for XR interaction.

2. The Importance of Handle Tracking Technology

2.1. The Importance of Traditional Handles in XR Systems

Traditional handles play a crucial role in XR systems. As mature interactive tools, they have irreplaceable advantages in precise input and complex operation scenarios. For example, in some professional training scenarios such as painting and surgical simulation[2], traditional handles can provide more precise operation feedback, ensuring that users can accurately complete various complex tasks. Meanwhile, traditional handles can also provide haptic feedback, which is vital for creating an immersive experience. When users operate in the XR environment, haptic feedback like the vibration of the handle and the pressing sensation of the buttons can make users feel the interactions in the virtual world more realistically and enhance their sense of immersion.

2.2. Challenges and Opportunities of Handle Miniaturization

High-Precision Tracking Difficulties: Miniaturized handles face the challenge of high-precision

ISSN 2616-5767 Vol.7, Issue 6: 177-181, DOI: 10.25236/AJETS.2024.070625

tracking. Firstly, the miniaturization of the device limits the size and quantity of sensors. The reduction in sensors makes it more difficult to obtain accurate position and motion information, thus affecting the tracking precision. Secondly, since the handles are small, users may make rapid and complex gestures more frequently, which puts forward higher requirements for the stability and accuracy of the tracking system. For example, in some fast-paced game scenarios, users' hand movements are rapid and complex, and miniaturized handles need to be able to accurately capture these movements; otherwise, it will affect users' gaming experience[3]. Moreover, high-precision tracking usually requires more calculations, which also brings a heavier burden to the processors of miniaturized handles and increases the difficulty of power consumption management at the same time.

Sensor Limitations: The sensors of miniaturized handles are restricted by space and may not be equipped with as many types of sensors as traditional handles. This may result in the handles being unable to obtain sufficient information in some cases, affecting the tracking effect. For example, some miniaturized handles may not be equipped with high-precision gyroscopes or accelerometers, which will affect the detection precision of the handles on users' hand movements[4].

Power Consumption Management Challenges: Miniaturized handles usually need to use smaller batteries, while the requirements for high-precision tracking and more calculations will increase power consumption[5]. Therefore, how to achieve efficient power consumption management under the limited battery capacity is an important challenge faced by miniaturized handles. For example, if the power consumption of the handle is too high, it may lead to a short battery life and affect users' usage experience.

Enhancing User Experience: The miniaturization of handles can improve the portability for users, enabling them to carry XR devices out for use more easily. Meanwhile, miniaturized handles can also make users feel more comfortable during use and reduce the fatigue caused by long-term use. For example, users can put miniaturized XR devices into their backpacks and have immersive experiences anywhere and at any time. In addition, miniaturized handles can also bring a more natural way for XR interaction[6]. With the continuous progress of technology, miniaturized handles can be seamlessly connected with more devices to achieve more diversified interaction methods.

Driving Technological Innovation: The challenges of handle miniaturization also prompt manufacturers to continuously carry out technological innovation. In order to solve problems such as high-precision tracking, sensor limitations and power consumption management, manufacturers need to develop more advanced tracking algorithms, sensor technologies and power consumption management solutions[7].

3. Classification and Characteristics of Handle Tracking Technologies

There are various handle tracking technologies, among which optical tracking, electromagnetic positioning and ultrasonic positioning are several relatively common methods.

3.1. Advantages of Optical Tracking Algorithms

The active optical positioning and tracking scheme based on infrared light has many advantages. Firstly, it can achieve high-precision tracking. By observing the infrared light emitted from the handle through the headset camera and using the principles of spot detection and multi-view geometry to calculate the position and posture of the handle, it provides users with an accurate interactive experience. Secondly, this scheme features low cost[8]. Compared with other technologies such as electromagnetic positioning and ultrasonic positioning, its hardware cost is relatively low, which is conducive to large-scale promotion and application. In addition, low power consumption is also one of its remarkable advantages, reducing the energy consumption of the device and prolonging the battery life.

This optical tracking scheme is mainly used to achieve precise and low-latency position and motion tracking. By using the image data captured by the camera and recognizing and tracking specific visual features, such as colors, textures, shapes, etc., it calculates the position and motion of objects (such as the user's head, hands or controllers), bringing users a smooth interactive experience.

3.2. Characteristics of the CV Modal Handle Tracking Algorithm

The CV modal handle tracking algorithm has unique advantages when the handle is occluded[9]. When the LED lights on the handle are blocked, this algorithm can restore the handle information by tracking the features of the bare hand. In the case of handle occlusion, the visual features of the bare hand are

ISSN 2616-5767 Vol.7, Issue 6: 177-181, DOI: 10.25236/AJETS.2024.070625

usually not obvious, which often leads to tracking failure. To address this difficulty, the bare-hand algorithm team innovatively proposed an end-to-end 6DoF (Six Degrees of Freedom) tracking algorithm called Down-Top. By effectively utilizing the global context information of multi-view sequences, it can accurately and stably predict the pose information of the handle at one time and provide robust 6DoF poses in a timely manner when the handle tracking fails.

For example, in special action scenes such as holding the hand flat or letting it hang naturally, due to reasons like the small handle being occluded or being far away, there are many blurry areas when the hand is enlarged. The traditional Top-Down structure has great difficulty in detecting the position of the wrist point, resulting in calculation failure and handle failure. However, the Down-Top structure can determine the position of the wrist point from information such as the arm and body in the large image, achieving a higher detection rate with similar accuracy.

The characteristic of this algorithm lies in that it can accurately restore the 6DoF information of the handle controller by tracking the features of the bare hand when the handle is occluded, thus maintaining stable tracking of the handle[10]. Meanwhile, it can also assist the optical tracking algorithm. When the handle is not occluded, the optical tracking algorithm can provide high-precision tracking. When the handle is occluded, the CV modal handle tracking algorithm can provide supplementary and precise observations by tracking the features of the bare hand. The deep integration of the two provides a more stable and accurate handle tracking solution for XR interaction. The Multimodal approach is shown in Fig. 1.

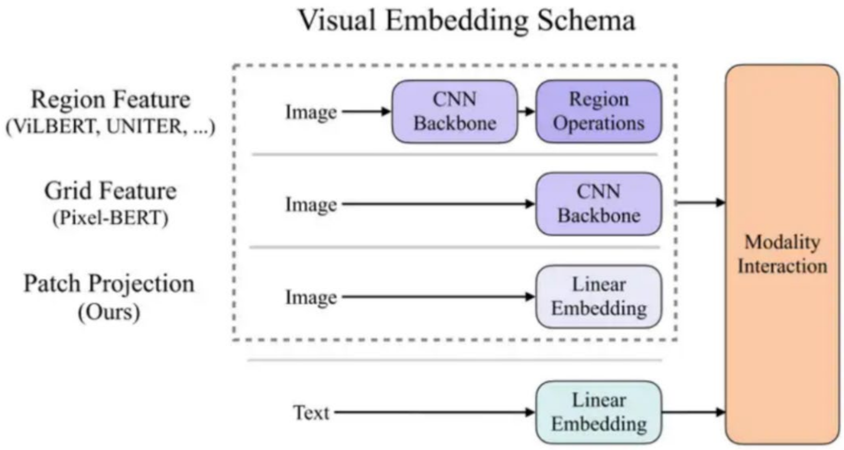

Figure 1: Multimodal approach.

The multi-modal tracking algorithm achieves high-precision handle tracking by using multi-modal data for pose prediction and estimation. The multi-modal tracking algorithm utilizes various modal data collected from the headset camera, the built-in IMU (Inertial Measurement Unit) module, the 3-DOF (Degrees of Freedom) module, the deep learning gesture detection and tracking module, and the optical positioning module to analyze hand features, handle motion information, LED positions, and global context information. In terms of data fusion, a multi-frame fusion filter is used to perform fusion calculations on all the acquired information, generating high-precision and high-frame-rate information on the position, rotation, and velocity of the handle for use by upper-layer applications. For example, in the case where the handle is occluded, bare-hand tracking can provide more accurate observations, while the handle provides more precise predictions for hand tracking, and the two are deeply integrated and mutually assist each other.

The design of the multi-modal handle tracking framework integrates the Inertial Measurement Unit (IMU), optical sensors and hand image information. Through the collaboration of an inter-departmental research team, stable tracking of the handle has been achieved.Firstly, the handle tracking algorithm team, the bare-hand tracking algorithm team and the interaction data team have determined the design scheme of the framework after numerous discussions. In the framework, the handle tracking algorithm team is responsible for improving the stability of optical tracking and building the algorithm framework. They utilize the optical tracking algorithm to observe the infrared light emitted from the handle through the headset camera and calculate the position and pose of the handle by using the principles of spot detection and multi-view geometry. Meanwhile, they also optimize the algorithm framework to enhance the efficiency and accuracy of the algorithm.

The bare-hand tracking algorithm team has designed a state-of-the-art (SOTA) end-to-end 6 Degrees of Freedom (6DoF) pose estimation model, which improves the stability of hand tracking. In the case where the handle is occluded, this model can accurately restore the 6DoF information of the handle controller by tracking the features of the bare hand. For example, in special motion scenarios such as holding the hand flat or letting it hang naturally, due to reasons like the small handle being occluded or being far away, there are many blurred areas when the hand is enlarged. The traditional Top-Down structure has difficulty detecting the position of the wrist point, resulting in calculation failure and handle malfunction. However, the Down-Top structure can determine the position of the wrist point from the information of the arm, body, etc. in the large image, achieving a higher detection rate with similar accuracy.

The interaction data team has provided a large amount of labeled data for the algorithm through self-developed multi-source three-dimensional data acquisition devices and high-precision reconstruction algorithms. These data include hand features, handle motion information, LED positions and global context information. Through the analysis and processing of these data, the multi-modal handle tracking framework can achieve accurate pose prediction and estimation.

Finally, a multi-frame fusion filter is used to perform fusion calculations on all the acquired information, generating high-precision and high-frame-rate information on the position, rotation and velocity of the handle for use by upper-layer applications. In the case where the handle is occluded, bare-hand tracking can provide more accurate observations while the handle provides more precise predictions for hand tracking. The two are deeply integrated and mutually assist each other. The application of this multi-modal data fusion not only solves the problem of "difficult recognition" of handle controllers without light rings but also provides new ideas for tracking and interaction in the XR industry.

4. Construction of a High-Precision Data Platform

4.1. Multi-View Hand Pose Labeling Technology

The platform adopts the industry-leading self-developed multi-view hand pose labeling technology based on camera arrays. This technology can label hand poses from multiple angles, providing more accurate and comprehensive hand pose data for algorithms. Through multi-view labeling, various hand movements and poses can be better captured, improving the precision and reliability of the data. For example, in different hand movement scenarios, such as grasping, rotating, stretching, etc., the multi-view labeling technology can accurately record the changes in the position and pose of the hand, providing strong support for the training and optimization of algorithms.

4.2. Pre-Reconstruction to Improve Data Precision

Pre-reconstruction is a data processing method that constructs more accurate hand pose models by analyzing and processing the original data. During the pre-reconstruction process, the hand pose data obtained by multi-source 3D data acquisition devices is processed. By analyzing hand features, motion information, and global context information, the pre-reconstruction algorithm can construct more accurate hand pose models, improving the precision and reliability of the data. For example, in some complex hand movement scenarios, such as quickly waving the hand or under low light conditions, the pre-reconstruction algorithm can predict the position and pose of the hand by analyzing historical data and global context information, improving the precision and stability of the data. This pre-reconstruction method can not only enhance the precision of hand pose data but also provide more accurate and reliable data support for the training and optimization of algorithms, further improving the performance and stability of the XR interaction system based on the CV modal handle tracking algorithm.

5. Extreme Tests and Effect Comparisons

Scenarios with extremely high requirements for tracking precision, such as music games and swimming fitness, were selected for the tests. In these scenarios, players' hand movements are rapid and complex, posing a severe challenge to the precision and stability of handle tracking. Geek players gave full play to their professional skills and ultimate operational abilities during the testing process. Through long-term and high-intensity use, the tracking performance of the new handle was comprehensively examined. The test results show that this algorithm performs excellently under various extreme circumstances. It can accurately track the position and pose of the handle, providing users with a stable

ISSN 2616-5767 Vol.7, Issue 6: 177-181, DOI: 10.25236/AJETS.2024.070625

interactive experience.

The CV modal handle tracking algorithm can restore handle information by tracking the features of the bare hand when the handle is occluded, thus maintaining stable tracking of the handle. In terms of tracking stability, the tracking algorithm fuses and calculates multiple types of data to generate high-precision and high-frame-rate information about the handle's position, rotation, and velocity for use by upper-layer applications.

Moreover, in terms of user experience, this algorithm can provide users with a more natural and smooth interactive experience. For example, in a painting scenario, users can directly perform some operations with gestures, such as adjusting colors and switching tools, without having to frequently use the buttons on the handle. This enables users to feel the interactions in the virtual world more authentically.

6. Conclusion

The XR interaction system based on the CV modal handle tracking algorithm has its unique advantages in terms of tracking precision, stability and user experience, bringing users a brand-new XR interaction experience.

Natural Interaction: Further develop the gesture tracking technology to enable users to interact with the virtual environment in a more natural way. For example, by recognizing more complex gesture movements, more interaction functions can be realized, such as grasping, rotating and scaling virtual objects.

Comfort Enhancement: Pay more attention to ergonomics in the design of handles to improve the comfort of users' grip. Meanwhile, reduce the weight and volume of the handles to relieve users' fatigue during long-term use.

Personalized Settings: Allow users to make personalized settings according to their own preferences and usage habits, such as adjusting the sensitivity and vibration intensity of the handles. This will improve users' satisfaction and usage experience with the XR interaction system.

With the continuous progress of technology, the development of handle tracking technology in the XR interaction system will bring users a more immersive, natural and comfortable virtual experience. Meanwhile, these development trends will also promote the wide application of XR technology in multiple fields such as gaming, education, medical care and design.

References

[1] Z. H. Lv and J. K. Guo, "Virtual Reality Neurorehabilitation," International Journal of Mental Health Promotion, vol. 24, no. 3, pp. 290-306, 2022.

[2] K. Kounlaxay, D. X. Yao, M. W. Ha and S. K. Kim, "Design of Virtual Reality System for Organic Chemistry," Intelligent Automation & Soft Computing, vol. 31, no. 2, pp. 1122-1128, 2022.

[3] R. Zhang and X. X. Zhao, "The Application of Folk Art with Virtual Reality Technology in Visual Communication," Intelligent Automation & Soft Computing, vol. 26, no. 4, pp. 786-792, 2020.

[4] K. Kounlaxay and S. K. Kim, "Design of Learning Media in Mixed Reality for Lao Education," CMC-Computers, Materials & Continua, vol. 64, no. 1, pp. 165-176, 2020.

[5] C. Neil, "Sportscapes in virtual reality," Journal of Cultural Geography, vol. 39, no. 3, pp. 408-411, 2022.

[6] S. Latif, H. Tarner, and F. Beck, "Talking Realities: Audio Guides in Virtual Reality Visualizations," IEEE Computer Graphics and Applications, vol. 42, no. 1, pp. 75-81, 2022.

[7] N. S. Sreekanth, N. Varghese and N. S. Chandra, "Virtual Chemistry Lab to Virtual Reality Chemistry Lab," Resonance-Journal of Science Education, vol. 27, no. 8, pp. 1381-1383, 2022.

[8] B. Bortnik, N. Stozhko, I. Pervukhina, A. Tchernysheva and G. Belysheva, "Effect of virtual analytical chemistry laboratory on enhancing student research skills and practices," Research in Learning Technology, vol. 25, no. 1, pp. 8-12, 2017.

[9] D. F. Liu, P. Valdiviezo-Díaz, G. Riofrio, Y. M. Sun, R. Barba, "Integration of virtual labs into science e-learning," Procedia Computer Science, vol. 75, no. 1, pp. 99-101, 2015.

[10] J. Piskorowski and T. Barcinsk, "Dynamic compensation of load cell response: A time-arying approach," Mechanical Systems and Signal Processing, vol. 22, no. 7, pp. 696-1702, 2008.