Proceedings of the 2nd International Conference on Computing Innovation and Applied Physics (CONF-CIAP 2023)
DOI: 10.54254/2753-8818/5/20230451

An assumption of UWB enhancement of human motion recognition

Chenyuan Zhang[1,*]

[1]Shanghai Dianji University, Shanghai, 200000, China

[*]8702610806@qq.com

Abstract. Today's VR games on the market use computers to connect a head-mounted monitor and two grips to build a virtual game world for players, providing players with a fresh sensory experience. However, the current VR game only uses the player's hand movement as a way of human-computer communication, the way is relatively single, and the player can not express their desire to play through only two hands, which cannot fully express the player's will. There are currently two types of motion capture used in consumer VR devices: optical capture, which is based on inertial sensor motion capture. These two are faced with the problem of extremely high cost and burden. The ultra-broadband system has the advantages of high positioning accuracy, low latency, cheap price, being light and portable, etc., which is very suitable for this kind of application scenario. This paper considers using two main technologies RTT ranging motion recognition and keyframe to achieve the aim.

keywords: ultra-wide band (UWB), RTT ranging motion recognition, keyframe.

1. Introduction

At present, VR games are popular, and there are many kinds of VR devices on the market, but after personal experiences, it is believed that human-computer interaction is relatively poor, and the player's willingness to move can only be expressed through hand movements, and the game experience is very single. Ultra-broadband technology is currently very popular in terms of positioning, and has the advantages of high speed, low latency, scalability, light weight, and low cost. Therefore, the author also envisions whether labels can be placed in human joints. Human body movements will change the position of the labels, and after the computer collects signals, it judges the human movements according to the software after deep learning and reflects them in the game. Human behavior recognition is a multidisciplinary and interdisciplinary research direction, involving image processing, computer vision, pattern recognition, machine learning, artificial intelligence and other disciplines. Human behavior recognition has a wide range of application prospects in the fields of intelligent video surveillance, natural human-computer interaction, smart homes and virtual reality [1].

Among them, the realization of keyframe extraction, automatic recognition and classification of three-dimensional human motion data is an important research content of human motion analysis, and an important premise and basis for effective management and reuse of motion capture data. As a time series data abstraction, keyframes are also one of the means of data compression. Keyframing technology represents the entire original motion sequence by using a few representative keyframes, which is also a data dimensionality reduction technique and feature extraction and representation

DOI: 10.54254/2753-8818/5/20230451

technique from another perspective. This paper envisages the construction of a motion posture recognition system that achieves computer interaction through the development and implementation of hardware and software modules. At the same time, the system also faces several problems, such as how to build coordinates, and how to collect data and process. This article presents preliminary ideas and solutions around several problems. First is about using the RTT ranging method. The RTT ranging method, also known as circular position line positioning, indirectly calculates the ranging value by measuring the round-trip time of the UWB pulse signal between the reference station and the moving label. It is represented in the axis.

For the processing of returned data, too much data is not conducive to analysis, and keyframe extraction is required to capture the data when the most easily recognized human body makes an action. The game does not require the character to completely coincide with the human body's movements, the only need is to complete the extraction of key actions such as running, walking, jumping, squatting, etc. to complete the interaction between the human body and the game character.

This paper takes into account image optical recognition and machine deep learning, for optical recognition there are camera-based close-range recognition [2], multi-perspective deep motion map human behavior recognition [3], and deep learning-based human behavior recognition [4], but this method requires a camera and is susceptible to interference and has delays. The basic principle of the action recognition device is to use the action recognition algorithm to extract a large amount of three-dimensional motion data to extract the motion trajectory, and then identify the corresponding action for different motion trajectories and then synchronize the action into the game by the computer.

Most of the current VR devices are based on two technologies for capturing motion postures, namely, optical capture and sensor capture. The first is susceptible to light interference. Some also need cameras to capture, which is expensive and inconvenient to carry, and is not suitable for civilian use. Second, if people want to capture the overall motion of the human body, they need to wear heavy equipment, which is not conducive to the game experience of players. Apple's latest airtag caught the author’s attention, it uses ultra-broadband technology, which is small and precise. Imagine that if the use of airtag is slightly changed to a wearable structure, which can be placed on each joint and combined with virtual reality devices, then a virtual reality game that recognizes human motion can be designed.

2. Difficulties

Many structures of human joints are complex, and various actions will cause different chain reactions of different joints, resulting in complex three-dimensional motion data, and the amount of calculation will increase accordingly, which will eventually lead to an increase in response time and reduce the game experience.

Keyframing of motion capture data provides the foundation for the compressed storage, retrieval, browsing, and further motion editing of motion capture data. When extracting keyframes, on the one hand, people need to consider which frames can best represent the original motion sequence, on the other hand, they must also consider the compression ratio effect of keyframes, and consider which factors will affect the decompression effect of the data and the effect of subsequent processing. Therefore, a simple and easy-to-control method is needed to complete the keyframe of extracting motion data.

There is no clear classification of human movement posture in the world, people need to define the required movements independently, classify the movement data, and use deep learning algorithms to accurately identify the movements.

3. Research design and methodology

Oversea countries are far ahead of the ultra-broadband technology, such as the United States company time domain, their products can be easily arranged compared with similar competitors, the accuracy can be less than 2cm and not easy to fluctuate. The system built in this paper is designed according to the actual needs, the software has positioning software and motion analysis software, and the hardware has wearable labels and reader synchronous distributors.

DOI: 10.54254/2753-8818/5/20230451

Considering that the three-dimensional data has high requirements for software and hardware water products, in order to simplify the data processing and operation, this paper considers splitting the three-dimensional human joint position information into a two-dimensional position analysis of a single joint, classifying the joints required for each action, and using the algorithm to deeply learn the different actions represented by different data. The following are a few basic movements of the game and their basic joints.

Table 1. Joints required for action

action				
Squat	hip joint	ankle	knee	
Jump	hip joint	ankle	knee	
Walking	ankle	shoulder joint	knee	elbow joint
Running	ankle	shoulder joint	knee	elbow joint
Dodge	head	shoulder joint	hip joint	ankle

At present, there are two international standardization organizations in the field of human modeling. One of them is the Moving Picture Expicture Group-MPEG, which publishes a standard for the description of virtual people known as the Face & Body Animation Specification (FBA). Another standardization organization is the Humanoid Animation Working Group, the virtual human modeling standard H-Animb (The Humanoid Animation Specification), which can achieve a unified description of the virtual human model in a network-wide 3D graphics and a multimedia environment,[5] allowing people to achieve control over the mannequin between different programs. It can achieve enhanced compatibility and flexibility for mannequin modeling. According to the standard, the virtual human skeleton model consists of the human center of gravity, human joints and human bone segments. The following figure shows a skeleton model commonly used in human motion analysis.

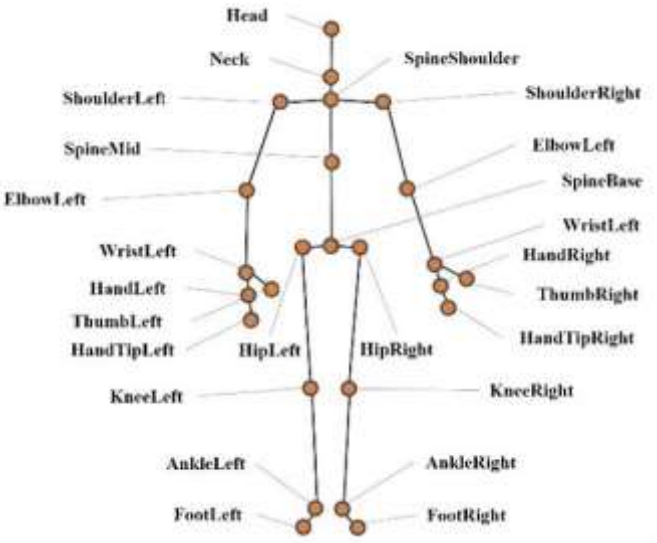

Figure 1. Main joints of human body.

The following figure is an example:

DOI: 10.54254/2753-8818/5/20230451

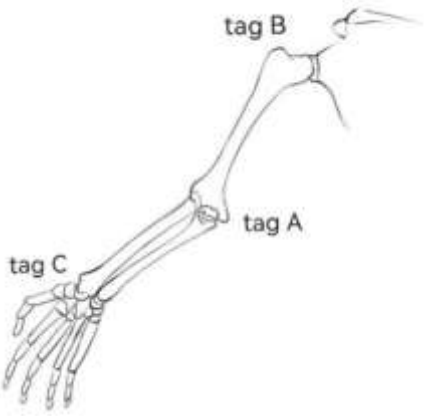

Figure 2. Joints of arms

Three labels A, B, C, are placed and the acceleration and direction of motion of the three labels are recorded respectively. Because individual differences in the human body exist objectively, it is necessary to avoid the selection of features with which they have a large correlation, so this paper also extracts the acceleration of the angle change between the bones of a feature. Combining the above three characteristic data can basically judge the human body movement.

For data measurements, RTT [6] ultra-wide band positioning technology based on round-trip time ranging is used to obtain position data with high accuracy, and because there is no large obstacle to the attachment on the human body, it does not cause large deviations. The schematic diagram is as follows.

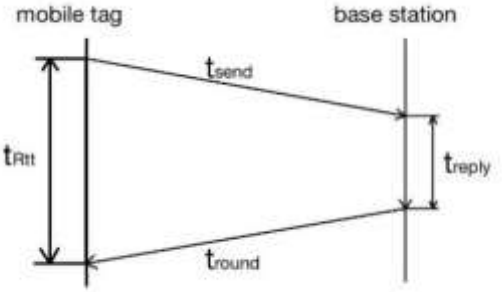

Figure 3. The principle of Rtt ranging.

The symbols and meanings are as follows: tRtt is the time information corrected by the standard time deviation; tsend represents the pulse signal transmission time; tround represents the pulse signal response time; treply represents the reflection time delay of the received signal, which is called the standard time deviation; di,k is the true distance between the k-time rover and the i base station; ri,k represents the pseudo-distance between the k-time rover station and the i base station; c is the speed of light; ξi,k is the error of the distance measurement at the k-time (mainly NLOS error).

Constructing a system of equations uses the least squares solution. [7]

As for data processing, since the computer will receive a large amount of motion data information, and there are many useless information, including slight shaking or unconscious shaking of the human body, data preprocessing can be a good way to eliminate useless actions and reduce errors. In this article, the amplitude of the acceleration vector AVA (AVA defined as the root number of the sum of the squares of the acceleration in the x and y axes) will be used to indicate the magnitude of the change in the action, and to determine whether it is constantly changing, and if the sum of the differences of multiple data is less than the threshold, the smaller amplitude is screened out [8].

DOI: 10.54254/2753-8818/5/20230451

After processing the redundant data, this paper chose to use a linear discriminant analysis classifier for classification identification, and data acquisition and machine learning as a sample set were required to reduce errors.

Finally, the key frame data is extracted, because the game only needs some simple basic actions, and it needs to select appropriate features and build models and classifiers. For this purpose, this article uses a cluster-based approach to cluster highly similar frames into one category.

The literature is clustered based on the curved convexity, and different candidate keyframes are obtained by clustering, and then the final keyframe is screened from each candidate key ton according to the set threshold. The literature proposes a key extraction method similar to the clustering method, which first maps the high-dimensional motion data to the low-dimensional implied state space using a linear time-invariant system, and then uses the error sum of squares criterion to divide the motion of the low-dimensional state variables of the motion sequence in the low-dimensional state space, believing that the segmentation point reflects the posture of the most violent change in motion, so the segmentation point between segments is used as a keyframe [9].

The process of the whole system is to install labels on the human body, and after the human body makes the action, the computer collects the information and preprocesses and then extracts the keyframe data to identify the human body action using a linear discriminative classifier.

4. Conclusion

This system is explained at every step of the way in this article, there are great possibilities for implementation, if successfully implemented, it will make VR technology more interactive with people, while small labels also achieve lightness and economy, the author also believes that this technology will make VR more popular among people.

This paper puts forward the idea of the practical application of ultra-broadband technology, the successful design can bring vitality and fresh blood to the market, but only for the ultra-broadband technology to propose an application method, in ranging, key frame extraction, action recognition will objectively have some errors. The author hopes that this idea can be practiced in the future and optimize the algorithm or model, with the calculation and hardware to do better, and finally achieve success, which has more promotional value. In the future, this technology may reform the way people play and entertainment habits.

References

[1] Zhu Honglei, Zhu Changsheng, Xu Zhigang. Research progress of human action recognition dataset [J]. Journal of Automation, 2018,44 (06) : 978-1004.)

[2] Xu Jian, Xu Kaikai, Zhai Yang, et al. Design of camera-based close-range gesture recognition system [J]. Computer measurement and control, 2016,24 (10) : 181-185.

[3] Liu Tingting, Li Yupeng, Zhang Liang. Human action recognition from multi-view depth motion images [J], Chinese Journal of Graphics, 2019, 24 (3) : 400-409.

[4] Shen Xiting, Yu Sheng, Dong Yao, et al. Human action recognition method based on deep learning [J]. Computer Engineering and Design, 2020,41 (4) : 1153-1157.

[5] Shen Junxing. Research on Motion Editing and Synthesis Technology: [D]. Hangzhou: Zhejiang University, 2004

[6] Zhang and Lei, a passive RTT measurement algorithm [J] Journal of Beijing University of Posts and Telecommunications, 2004 (5), 85-89

[7] Xu Aigong, Zhang Han, Song Jiapeng, et al. High-precision indoor positioning method using UWB / LiDAR odometers [J]. Mapping Science, 2022, 47 (2)

[8] Zhang Ping, Liu Zuoshi. Gesture recognition method based on inertial sensor MPU6050 [J]. Sensors and Microsystems, 2018,37 (1): 46 49.

[9] Zhu Dengming, Wang Zhaoqi Journal of Computer Aided Design and Graphics for Key Frame Extraction of Motion Capture Data Based on Motion Sequence Segmentation, 2008, 20 (6): 787-792