Capture of 3D Human Motion Pose in Virtual Reality Based on Video Recognition

Qiang Fu,1 Xingui Zhang1,2, Jinxiu Xu,3 and Haimin Zhang1

1Institute of Physical Culture, Henan Institute of Science and Technology, Xinxiang, Henan 453003, China
2Department of Sports, Tsinghua University, Beijing 100084, China
3Sanquan College of Xinxiang Medical University, Xinxiang, Henan 453003, China

Correspondence should be addressed to Xingui Zhang; zxg@mail.tsinghua.edu.cn

Received 30 September 2020; Revised 28 October 2020; Accepted 9 November 2020; Published 21 November 2020

Copyright © 2020 Qiang Fu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Motion capture technology can effectively solve the problem of difficulty in defining character motion in the process of 3D animation production and greatly reduce the workload of character motion control, thereby improving the efficiency of animation development and the fidelity of character motion. Motion gesture capture technology is widely used in virtual reality systems, virtual training grounds, and real-time tracking of the motion trajectories of general objects. This paper proposes an attitude estimation algorithm adapted to be embedded. The previous centralized Kalman filter is divided into two-step Kalman filtering. According to the different characteristics of the sensors, they are processed separately to isolate the cross-influence between sensors. An adaptive adjustment method based on fuzzy logic is proposed. The acceleration, angular velocity, and geomagnetic field strength of the environment are used as the input of fuzzy logic to judge the motion state of the carrier and then adjust the covariance matrix of the filter. The adaptive adjustment of the sensor is converted to the recognition of the motion state. For the study of human motion posture capture, this paper designs a verification experiment based on the existing robotic arm in the laboratory. The experiment shows that the studied motion posture capture method has better performance. The human body motion gesture is designed for capturing experiments, and the capture results show that the obtained pose angle information can better restore the human body motion. A visual model of human motion posture capture was established, and after comparing and analyzing with the real situation, it was found that the simulation approach reproduced the motion process of human motion well. For the research of human motion recognition, this paper designs a two-classification model and human daily behaviors for experiments. Experiments show that the accuracy of the two-category human motion gesture capture and recognition has achieved good results. The experimental effect of SVM on the recognition of two classifications is excellent. In the case of using all optimization algorithms, the accuracy rate is higher than 90%, and the final recognition accuracy rate is also higher than 90%. In terms of recognition time, the time required for human motion gesture capture and recognition is less than 2 s.

1. Introduction

Virtual reality (VR) is a virtual immersive interactive environment that uses modern high-tech with computer technology as the core to generate a specific range of realistic visual, auditory, and tactile virtual environments [1]. The research purpose of VR technology is to create such a simulated virtual environment so that users can interact with objects in the environment with necessary equipment to achieve an “immersive” effect, just like the feeling and experience of the real environment [2]. With the development of VR technology, natural interaction and multimodal interaction between humans and computers will become the main way of interaction between humans and computers. In this context, motion capture technology, as a key technology of virtual reality, has received more and more attention from researchers [3]. Motion capture technology refers to the use of optical, acoustic, or other motion data collection equipment to record the actions of the performer and subsequent processing to convert the recorded data into data that can drive the motion of the computer virtual character. The information collected by tracking can be used to simply
describe the position in the specific three-dimensional space of the captured object’s limbs, or it can be used to describe the subtle changes in the complex face or skin [4].

With the rapid development of computer technology, the New York Computer Graphics Technology Laboratory has designed an mercury mirror optical device to project the performance poses of real dancers on the computer screen as a reference for the key frames of digital dancer animation [5, 6]. It also promotes the development of motion gesture capture technology. Motion gesture capture technology has quickly attracted the attention of more scholars and even developers [7]. For example, the Biomechanics Laboratory in the United States and the Massachusetts Institute of Technology has begun to conduct research on motion gesture capture technology based on computer technology. American Pacific Data Graphics Corporation designed an input device with 8 degrees of freedom to control the position of animated characters and the movement of lips [8]. The appearance of this device shows that people’s research level of motion gesture capture technology has risen to a new level. In recent years, foreign sports gesture capture technology has moved from experimental research to practical application, and with the increasing demand for animation production efficiency and production quality, various types have appeared in some countries’ markets [9, 10]. There are also certain differences between the implementation schemes of the motion posture capture device, and the corresponding emphasis on different practical applications is also very different, and the price is basically more expensive [11]. At the same time, the application range of the motion posture capture device has gradually gone beyond the field of animation production, and many have been applied in many fields such as game development and ergonomics [12]. With the rapid development of the animation game industry and the field of film and television production, there are higher requirements for the fidelity of the effect of the work, as well as the production efficiency and quality [13, 14]. This also promotes the continuous development and development of motion gesture capture technology. After rapid progress, the 3D movie "Avatar" shot based on motion gesture capture technology has won unanimous praise from audiences all over the world, which has played a certain reference and promoted the domestic film and television creation field [15–17]. Especially, after discussion at international conferences in related fields, motion capture technology has attracted more and more attention from people [18]. A large part of the major universities across the country have opened animation production and other related professional disciplines [19, 20]. The production efficiency and the fidelity of animation promote the efficient and perfect integration of technology and art, and the demand for sports posture capture technology knowledge and sports posture capture system is huge [21–23]. It has received extensive attention from major scientific research institutions and companies [24]. A general acoustic motion capture device is mainly composed of a data processing module, an ultrasonic transmitter, and an ultrasonic receiver [25, 26]. The device fixes an ultrasonic generator at a specific position in the capture scene to transmit ultrasonic waves [27, 28]. A receiver composed of three ultrasonic probes (arranged in a triangle) receives the signal. It measures the time and phase difference when the ultrasonic waves are received by different receivers [29, 30]. Since ultrasound has good penetrability, this acoustic solution has certain advantages; that is, it can solve the problem of blind spots caused by occlusion [31]. At present, the ultrasonic motion capture devices developed by Logitech in Switzerland and SAC in Germany are relatively mature products [32]. Motion capture based on video sequences mainly imitates the principle of the human eye. It uses the comparison between two frames of images obtained by two spatial cameras at the same time to identify the motion object and complete the spatial positioning of the motion object [33]. This method uses intelligent algorithms to recognize actions, and there is no need to wear maker points on the action object, thus reducing the requirements for collection equipment [34, 35]. It can be considered that this is a special optical capture system.

The Kalman filter of the current inertial motion capture system has perfect modeling and mature algorithm, which realizes attitude estimation under various motion states, and has the advantages of high accuracy and good stability. However, the model algorithm is complex and the calculation is large, so high-speed data processing and complex attitude estimation algorithms can only be realized on a PC. Therefore, inertial motion capture system applications are mainly concentrated in offline processing applications. On the other hand, a large amount of data transmission puts forward corresponding requirements for data transmission and power supply, especially in high-speed motion capture and wireless data transmission networks. This article aims to research and organize related experiments on human motion tracking and recognition methods based on MEMS accelerometers and MEMS gyroscopes. We select the appropriate data acquisition device and determine the best fixed position, complete the analysis of the data components, and propose a preprocessing method. We design a wavelet threshold noise reduction method suitable for MEMS sensor data and organize human motion data noise reduction experiments to analyze and verify the designed noise reduction model. In the experiment of human motion posture capture, this article uses the stable and more accurate characteristics of the robot arm to verify the designed motion posture capture method and prove its correctness. Specifically, the technical contributions of this article can be summarized as follows:

Firstly, a two-step Kalman filter is proposed. This method can not only significantly reduce the calculation burden of the MCU but also effectively improve the data utilization. In order to further improve the estimation accuracy and system stability, the information distribution factor of the filter and the noise matrices R and Q are adjusted in real time, and an adaptive adjustment method based on fuzzy logic is adopted to enable it to be uniform in various motion states. It can ensure the stability of the output, further simplify the fault detection and isolation algorithm, and reduce the burden of fault-tolerant calculations.
Secondly, after collecting, denoising, and analyzing the human motion data, the initial and dynamic posture angles of the human body are obtained through the motion posture capture method, thus completing the human body capture work well. A visual model of human motion posture capture was established through the human VR model and MATLAB/VR toolbox, and the motion trajectory of the end of the human body in space was drawn through SimMechanics. These two approaches reproduce the motion process well. After comparison, it can be seen that the capture result basically conforms to the actual human movement.

Thirdly, grid search-support vector classifier (GS-SVC), genetic algorithm-support vector classifier (GA-SVC), and particle swarm optimization-support vector classifier (PSO-SVC) are three optimized SVCs for the two-classification mode recognition research. Experimental results show that no matter what optimization method is used, the recognition accuracy of the two classifications has reached more than 90%; the accuracy of the grid search algorithm GS-SVC in the multiclassification mode is not ideal, and the error rate is about 10%; the recognition accuracy of GA-SVC and PSO-SVC of the heuristic algorithm reached more than 92% in the two-classification mode.

The rest of this article is organized as follows. Section 2 discusses the key technologies of human motion gesture capture and recognition. Section 3 designs the pose estimation algorithm of the motion pose capture system. In Section 4, the human body motion gesture capture and recognition experiment is carried out. Section 5 summarizes the full text.

2. Key Technology of Human Motion Posture Capture and Recognition

2.1. Virtual Reality Scene Generation Based on Real Scenes. There are three aspects to be considered in the reality-based modeling in the virtual reality environment:

One is realism. The fidelity here is not exactly the same as what we mean by the fidelity in general 3D modeling. Generally, the fidelity of three-dimensional modeling is to make the user look like a certain scene in reality, and not specifically limited to which scene, as long as the user feels acceptable. And virtual reality scene modeling based on real scenes requires that all aspects of a scene be reflected as truthfully as possible, and it must withstand comparison.

The second is real-time. This includes the sense of continuity when walking through the scene and the smoothness of other scene animation processes. This is also the biggest difference from general three-dimensional modeling for the purpose of renderings. In order to ensure the real-time requirements of virtual reality, one cannot simply pursue beauty and excessive details. For complex real scenes, certain simplification strategies must be adopted.

The third is controllability. It means that the modeled objects must be suitable for control. This is the special requirement of human-scene interaction in virtual reality. The objects that should be merged must be merged, and the relationship between the objects must be determined according to the purpose to be achieved.

From the perspective of implementation, there are generally three schemes, which are directly drawn with graphics functions or modeled with third-party modeling software and then exported, and the other is to automatically generate with specific equipment and rely on noncontact vision technology. Since the complexity of the real scene is generally relatively high, including a large number of irregular surfaces, it is difficult to choose the parameters drawn directly by the function, and the method of reconstructing the geometric information of the object with the noncontact visual method requires special equipment and calibration, which is very complicated. Here, third-party modeling software is generally used to model and then export the object information. In addition, the reason for using third-party modeling software is to facilitate the realization of various functions for virtual objects in the virtual reality environment, such as collision detection. The composition of the virtual reality system is shown in Figure 1.

In order to more effectively improve the operating efficiency of the entire system, a multilevel detailed model mechanism is adopted to manually create a complex and simple model for each object. When the object is far away from the viewpoint, the relatively simple model is used. This can achieve a balance between realism and speed. Since the distance between indoor activities is not particularly obvious, we only set 2 to 3 levels. The scope of application of each level is dynamically allocated and there is intersection, which is different from the general system. The so-called dynamic allocation refers to the decision based on the total complexity of the current scene. If it is too high, a simple model of some objects will be used; otherwise, it will be rendered normally so that the choice of model can be determined by adaptive speed changes. The so-called intersection means that, under the premise of ensuring speed and certain fidelity, both simple models and complex models can be called within a certain distance range, that is, maintaining the inertia of "model calling" and reducing frequent switching of models in a fixed range boundary zone.

2.2. Sports Posture Acquisition Technology. Human body motion posture acquisition technology is a hot topic in human body motion analysis and intelligent information processing, and it is also the first step in human body posture recognition. According to whether the device needs to be worn or marked on the athlete, it is divided into contact recognition technology and noncontact recognition technology. The commonly used contact recognition technologies include mechanical, optical, and electromagnetic, and noncontact recognition is computer vision.

Mechanical motion posture capture is that the experimenter obtains real-time records of moving target tracking and measurement by wearing a mechanical device connected
to a sensor, and the driving of human motion is the only way for a mechanical device to obtain data. The advantages of this method are low cost, accuracy can be improved with the number of devices worn, and real-time measurements can be made. However, the wearing of mechanical devices is likely to cause movement restrictions on athletes.

Both electromagnetic and optical motion posture capture systems require the athlete to wear sensors or make marks on key parts of the body and move in a prescribed space with the athlete’s actions. The advantage of this kind of motion gesture capture is that it acquires multidimensional information, so the information content is rich. Moreover, the posture acquisition speed is fast, and the real-time performance is good. Compared with the mechanical motion posture capture wearable device, it is simpler and more portable and will not affect the motion posture of the athlete. The disadvantage is that it has strict environmental requirements. For example, the electromagnetic type is susceptible to the influence of metal electromagnetic field distribution, and the light in the optical type is affected by problems such as ground reflection.

Computer vision-based human motion gesture capture technology is to obtain video information through lens equipment and then recognize the human motion gesture in the video or image. The frame diagram of human motion posture capture is shown in Figure 2.

2.3. Human Body Motion Posture Feature Extraction Technology. Body movement can generally be described by body posture, movement trajectory, movement time, movement speed, movement strength, and movement rhythm. Among them, the most direct visual contact with people is the body posture. Therefore, the description of this type of information is of great significance in human body gesture recognition technology. Body posture refers to the state of the body and various parts of the body in different stages of motion. Describing actions based on the stretchability of the motion posture is a common way to describe body posture.

Describing human motion posture through feature information is a key step in human motion posture capture and recognition. Among them, the characterization ability of the extracted feature descriptors has a great impact on the accuracy and robustness of the recognition during the recognition of the motion gesture. Therefore, when the feature information accurately describes the motion posture of the same category, it also needs to reflect the degree of distinction between different categories. Moreover, similar features obtained by the same extraction method have different posture description capabilities for different types of motion. Therefore, it is necessary to select different types of features according to various considerations such as the quality of the acquired video, the actual type of exercise, the

![Figure 1: The composition of the virtual reality system.](image-url)
occurrence scene, and the number of people participating in the exercise. For example, in the long-range situation, it can be analyzed according to the trajectory changes in the moving target; in the close-range situation, the changes in the human body’s limbs and torso can be analyzed through the information extracted from the video image sequence.

The global feature describes the human body region of interest obtained by methods such as moving target tracking or foreground detection and describes it through silhouette contours, optical flow information, and moment features. Global features have the characteristics of good invariance, simple calculation process, and intuitive and rich expression, but they are more sensitive to noise, partial occlusion, and viewing angle changes and are very dependent on accurate underlying visual processing, and the amount of information is large.

Local feature extraction refers to the description of points or blocks of interest in the video, without the need to track moving targets or perform any contour trajectory modeling. Local features are insensitive to complex background, illumination changes, viewing angle changes, and partial occlusion issues, but they cannot describe the overall information of the moving target.

Just like when humans perform visual analysis, both global and local information are considered. Global features and local features in computer vision also have their own advantages and disadvantages. Therefore, the extraction of multiple feature information, which complements each other, is beneficial to the recognition and classification of motion gestures.

Feature-level fusion can be divided into two types: direct feature combination and feature selection combination according to analysis and processing methods. The direct feature combination method is to directly combine all feature vectors into a new feature vector group according to a certain simple rule, such as serial and parallel fusion methods. The feature selection and combination method is to put all the original feature vectors together and, according to a certain selection rule, select and retain some feature vectors from the original feature vectors as a new feature vector combination. The serial fusion method is the simplest and most effective feature fusion method. Its disadvantage is that the feature dimension increases after the fusion, all the feature information is retained, and there are a large number of redundant feature vectors. The parallel fusion method ensures that the feature dimension remains unchanged, but
the use of complex space vectors increases the complexity of operations and has a certain impact on the real-time performance of recognition.

2.4. Human Motion Posture Capture and Recognition Technology. The process of human motion posture capture and recognition is to complete the classifier training based on the feature information of the motion posture sample and calculate and compare the feature information of the test image sequence to achieve classification. This process is supervised learning.

2.4.1. Methods Based on Probability Statistics. The motion posture is represented by a continuous state sequence, in which the key frame static motion posture is set as a state, and the probabilistic relationship is used to represent the change between states, and the joint probability corresponding to the entire state sequence is calculated to perform action classification. It is usually divided into two categories: production classifiers and discriminative classifiers. The production classifier is to estimate the joint probability distribution, establish a posterior probability model, based on statistics and Bayesian theory, reflect the distribution of various types of data, and focus on the similarity of similar data, such as hidden Markov model. The discriminant classifier estimates the conditional probability density, looks for the optimal interface between different categories, and focuses on the differences between different types of data, such as support vector machines.

2.4.2. Method Based on Template Matching. In the template-based method, the known human motion posture video sequence and its extracted related features are used as a set of templates to represent the motion posture, and the similarity measure is used as the criterion to match the motion posture template to be tested with the known template. There are mainly template matching method and dynamic time warping. From the perspective of feature selection for template matching, it can be divided into colors, contours, and textures. The dynamic time warping algorithm is based on the principle of dynamic programming for time warping, which can reduce the time used for search and comparison in the matching process. Such template-based methods have the advantage of low computational complexity, but the recognition results are susceptible to changes in time intervals and noise.

2.4.3. Semantic-Based Approach. The grammar-based method is to add grammatical information based on the characteristics of the movement gesture in the process of human movement gesture recognition so that it has high-level meaning and forms an abstract description of the movement gesture. This is similar to how the human brain understands and recognizes motion gestures.

2.5. Human Body Posture Measurement Principle and Scheme Design. Human body posture measurement is the core algorithm for the realization of the motion capture system. The algorithm involves orientation detection, scale measurement, and spatial position coordinate positioning methods and finally synthesizes the data format of the human body motion capture information that the computer can recognize and process. The computer drives the virtual characters on the host computer to reproduce the captured skeletal motion trajectory by processing and analyzing the human body motion data and then realizes the extended application in other fields.

The human body posture measurement technology based on multiple sensors uses inertial devices such as accelerometers, gyroscopes, and magnetometers and magnetic sensitive devices to measure the rotation angle of human bone joints. Figure 3 shows the implementation process of the human motion posture measurement algorithm designed in this paper. First, we improve the posture information fusion algorithm according to the characteristics of the traditional posture calculation algorithm, fuse three different types of sensor data, and accurately obtain the posture angle of a single node in the geographic coordinate system and then establish a human skeleton model and define the skeleton parent-child relationship. Based on the principle of forward kinematics, we establish a human joint rotation angle model and a human bone motion position model and derive the rotation angle and relative displacement of each joint relative to the parent joint; then, the support leg is detected according to the gait characteristics of the human walking state; finally, the human body motion trajectory is captured.

3. Attitude Estimation Algorithm of Motion Attitude Capture System

3.1. Human Motion Model. In practical applications, people have introduced a threshold judgment method to detect the magnitude of motion acceleration, which is used as a criterion for judging. The maximum angle error $\theta_{\text{max}}$ judged by the threshold is

$$\theta_{\text{max}} = \arctan \left( \frac{x}{y} \right)$$

According to the calculation result of the threshold judgment, the angle estimation error still exists, which cannot meet the actual use requirements. Moreover, when there are different degrees of motion acceleration in the three axes, the amplitude may be close to $1g$. If only the acceleration amplitude is used to directly adjust the reliability of the $X$, $Y$, and $Z$ axes, it is easy to cause misjudgment.

Assuming that the rotational angular velocity of a joint is $\omega$, the angular acceleration is $\alpha$, and the rotation center node is $P$, the movement acceleration in the carrier coordinates of the point $P$ can be obtained:

$$l_p = \alpha \cdot o + w \cdot (o \cdot w) - o \cdot \|w\|^2.$$
The motion acceleration of the carrier coordinates is converted into geographic coordinates, and the chest is taken as the center to establish the motion acceleration estimation equation of the whole body:
\[ l_w^p = q_{jk} (\mathbf{a} \cdot \mathbf{w} + \mathbf{o} - o ||\mathbf{w}||^2). \]  
(3)

The estimated motion acceleration provides new parameters and methods for the error correction of the gravity field vector because the accuracy of the gravity acceleration of each axis decreases with the increase in the movement acceleration. When there is motion acceleration on the X axis, only the credibility of the X axis gravitational acceleration component is adjusted. At this time, the Y and Z axes gravitational acceleration components can still participate in the Kalman calculation. In this paper, the acceleration components of the three axes are processed separately according to the motion acceleration of each axis, which reduces the mutual interference between each axis, improves the utilization of data, and improves the accuracy of recognition of motion acceleration.

3.2. Two-Step Kalman Filter Design. Kalman filter is an optimal estimation theory with unbiased, linear, and minimal variance. Knowing the mathematical model of the state vector and the observation vector, the noise statistical characteristics of the state vector and the observation vector, and the initial value of the system state, the measurement data and state equation of the sensor can be used to obtain the relationship between the system state vector and the observation data. Optimal estimation is generally divided into two steps: estimation and correction, respectively, corresponding to the update equations of the state vector and the observation vector. The former is based on the optimal estimation of the state at time \( K - 1 \) and constructs the first state of the state at time \( K \). The latter is to combine the output of the current sensor to construct the posterior estimation of the state at time \( K \), and the two are alternately completed in turn to achieve the optimal estimation of the state.

The optimal estimation of the Kalman filter is achieved when the mathematical model of the state vector and the
observation vector is accurate, and the statistical characteristics of the noise and the initial value of the system state are known. Therefore, the accuracy of the statistical characteristics of the noise determines the Kalman to a large extent. The main reasons for the inaccuracy of the system model are, for example, lack of a complete understanding of actual physical problems or sufficient statistical data; in order to meet actual engineering requirements, the real physical model is simplified, or the nonlinear equation is linearized and approximated. Some minor external interferences are ignored, which leads to the fact that the mathematical model of the system or the statistical characteristics of noise can not accurately reflect the actual physical process of the system, cannot reflect the real system state, and even make the filter tend to diverge. In addition, it is difficult to accurately understand the statistical characteristics of time-varying noise during filter operation. The above factors make the theoretically optimal and stable Kalman filter, but it may be divergent in practical applications.

When the conventional centralized Kalman filter performs data fusion on the inertial motion attitude capture system, there are many serious problems, such as follows: (1) Kalman filter is an optimal estimation algorithm based on matrix operations, and its calculation will be based on the matrix dimension. It is difficult to guarantee real-time performance, especially in multisensor combined measurement. (2) The use of multiple sensor subsystems increases the probability of cross-influence between sensors, and a certain subsystem fails. If it is not detected and isolated in time, it will further affect the entire system and reduce reliability.

In response to these shortcomings, the decentralized Kalman filter achieves the optimal synthesis of multisensor information. It not only improves the measurement accuracy of the system but also has a certain fault tolerance, thus obtaining the best overall performance. For decentralized filters, sensors with different characteristics are isolated and processed in parallel, thereby avoiding cross-effects between sensors and reducing the amount of calculation.

Regardless of whether Euler angles or quaternion expressions are used, the two types use the updated carrier attitude of the gyroscope as the state vector of the Kalman filter, but the former uses the carrier attitude calculated by the accelerometer and magnetometer as the Kalman filter’s state vector. The observation vector is called the loose coupling mode; the latter uses the output of the accelerometer and magnetometer as the observation vector of the Kalman filter, which is called the tight coupling mode. The latter has a compact structure but involves 6-dimensional or even 9-dimensional matrix operations and is not suitable for embedded systems.

Taking into account the requirements of the calculation speed of the Kalman filter, this paper uses the carrier attitude calculated by the accelerometer and magnetometer as the observation vector of the filter, which avoids the nonlinear operation in the Kalman filter. However, in the process of calculating the attitude of the accelerometer and magnetometer, if the pitch and roll angles calculated by the accelerometer’s gravity vector are used as the input values for calculating the geomagnetic field vector, the yaw angle obtained will include the magnetic interference. Therefore, in the article, first we complete the angle estimation of the pitch and roll angles according to the gyroscope and accelerometer and use the estimation results as the input of the yaw angle solution to avoid transmission errors caused by motion acceleration.

In view of the sensor’s random noise, environmental interference, and measurement errors caused by motion, combined with the attitude calculation method of inertial tracking and decentralized Kalman filtering, this paper introduces a two-step Kalman, and the design filter is shown in Figure 4. Because during the movement of the carrier, the accelerometer is mainly affected by the movement acceleration, and the magnetometer is mainly affected by the surrounding electromagnetic interference, so the error of the sensor output is different, and the accelerometer and the magnetometer are treated separately. Because the reasons for the errors are different, in the two-step Kalman, the adaptive adjustment basis is also different. The two-step Kalman filter consists of two sub-Kalman filters, which are as follows:

The first step is to perform Kalman filtering of pitch and roll angles. The Kalman filter for pitch and roll angles is composed of a gyroscope and an accelerometer. The gyroscope updates the state vector, and the corrected gravity vector updates the observation vector to form a 2-dimensional Kalman filter, as shown in the following formula. $\Delta T$ is the sampling interval.

$$X_k^r = w_k + A_{k-1} \cdot X_{k-1} + B_{k-1} \cdot u_k,$$

$$Z_k = X_k^r \cdot H_k - v_k. \quad (4)$$

The second step is to perform the Kalman filter of the yaw angle. According to the estimation result of the first step, it is substituted into the geomagnetic field vector equation. We convert it from the carrier coordinates to the geographic coordinates in the X-Y plane to obtain the observation vector of the yaw angle, and the gyroscope updates the state vector of the yaw angle to form a 1-dimensional Kalman filter.

The two-step Kalman filter in this paper performs the state update and angle estimation of the pitch, roll, and yaw angles in sequence, which simplifies the complex error analysis and matrix calculations in the centralized Kalman filter, significantly reduces the calculation burden of the MCU, and improves the speed of processing. On the contrary, substituting the pitch and roll angles into the solution equation of the yaw angle provides more accurate observations for the Kalman filter equation of the yaw angle. The two-step Kalman filter decouples the traditional Kalman filter step by step. Two local filters are performed alternately, and different sensor characteristics are processed separately, which can greatly reduce the MCU calculation, avoid cross-effects between sensors, and improve the ability of system error isolation and reconstruction.
3.3. Two-Step Kalman Filter Algorithm Based on Fuzzy Logic. Kalman filter is to optimally allocate the information weight between the state vector and the observation vector according to the error size of the state vector and the observation vector, so as to realize the optimal estimation of the state. The size of the error between the state vector and the observation vector is reflected by the covariance matrices $Q$ and $R$. The error of the state vector and the observation vector together form the innovation sequence of Kalman filter, so the adaptive adjustment of Kalman filter is mainly carried out around innovation. In order to improve the estimation accuracy of the filter, this paper proposes an adaptive adjustment based on the motion state. According to the judgment result of fuzzy logic, the covariance matrices $Q$ and $R$ are adjusted online in real time.

Compared with conventional logic, fuzzy logic establishes control rules based on actual operating experience, so it can more directly reflect human thinking and judgment results and can realize reasonable judgments and operations without complete information. Fuzzy logic reduces the complexity of system modeling and analysis, greatly reduces the amount of calculations, and can realize control calculations more quickly and accurately. In addition, fuzzy logic requires less storage space, thus reducing the technical requirements for hardware.

The actual variance matrix $C_r$ is approximated by the average of the window estimates:

$$C_r = \frac{1}{M} \sum r_i \cdot T \cdot r_i.$$  \hspace{1cm} (5)

We define the difference between the theoretical and actual variance values as $\text{DoM}$ (degree of matching). If the $\text{DoM}$ value is close to zero, it means that the theoretical and actual values of the matrix $R$ are well estimated, and only a small adjustment to $R_k$ is required or not. If $\text{DoM}$ is positive, we decrease $R_k$; otherwise, we increase $R_k$. If $\text{DoM}$ is used as the input value of fuzzy logic, the corresponding output value is $AR$.

The adaptive multimode Kalman filter replaces the previous amplitude adjustment function with the new and relevant time function of the Kalman filter and then assigns the probability distribution of each hypothesis test model. The improved algorithm is called RCKFB (residual correlation Kalman filter bank). This method detects and judges system failures based on new information within a certain period of time, so there will be a certain time delay in system failures. However, the requirement for input information is reduced, and the accuracy of fault identification is improved.

Although the implementation methods of various adaptive filtering are different, they are all based on the statistical innovation of the innovation sequence. The theoretical calculation values tend to be consistent, thereby reducing the accuracy requirements for the statistical characteristics of the noise of the observation vector, especially when the noise characteristics are changed, and then, the Kalman gain matrix $K$ is changed to improve its estimation accuracy. This method realizes the adaptive adjustment of $Q$ or $R$ and avoids filtering divergence. However, the variance matching uses the average value of a certain amount of innovation as its actual variance value. It relies more on statistical characteristics and cannot accurately distinguish it.

According to the two-step Kalman filter described above, the output of the accelerometer and the magnetometer are processed separately in two subfilters, and the motion acceleration of each axis is estimated, so the error of the observation value of each angle in the Euler angle can be respectively estimated; that is, the motion acceleration of the $X$-axis, the motion acceleration of the $Y$-axis and the $Z$-axis, and the intensity of the geomagnetic field, respectively, correspond to the error size of the pitch, roll, and yaw angle, combined with the angular velocity of the carrier to adjust the matrix $R$. According to the angular velocity, motion acceleration, and geomagnetic field strength of each axis, this paper proposes an adaptive adjustment method based on the motion state. The output of the sensor and the estimated motion acceleration of each axis are used as the judgment basis of fuzzy logic to identify the current carrier motion. According to the different motion states of the carrier, respectively, we estimate the data reliability and angle estimation error of each sensor in the current state and adjust the variance matrix $Q$ of the state vector and the variance matrix $R$ of the observation vector. For example, when the carrier rotates rapidly, the accuracy of the state vector is improved; that is, $Q$ is reduced, and $R$ is increased; when the carrier is rotating at a slow speed or in a stationary state, $Q$ is increased and $R$ is decreased. According to the motion state, Kalman’s adaptive adjustment is performed, which avoids relying solely on the innovation sequence, and can more accurately estimate the error size of each sensor in the current state.

The measurement error of the gyroscope mainly includes zero offset and quantization error. The zero offset will continue to accumulate over time, but it will be a fixed value.
within a fixed sampling interval. The quantization error will increase linearly with the increase in angular velocity. The angular velocity in geographic coordinates is related to the three axial angular velocities in the carrier coordinates. For ease of operation, this article is simplified to have a linear relationship with the amplitude of the angular velocity. Therefore, the carrier attitude error updated by the gyroscope includes three parts.

When the carrier is in different motion states, the reliability of the accelerometer to measure the pitch and roll angle changes. According to the foregoing, the pitch and roll angles in the observation vector are directly related to the motion acceleration of each axis, and the estimation accuracy of the motion acceleration is affected by both the IMU installation error and the angular velocity measurement error.

When the magnetometer is subjected to external electromagnetic interference, the reliability of the yaw angle changes; on the other hand, the calculation of the geomagnetic field vector is based on the estimated values of the pitch and roll angles, and the angular velocity of each axis reflects the rotation of the carrier state, so the matrix $R$ is the equation between magnetic field strength, angular velocity, and innovation.

In the previous analysis, an adaptive adjustment method based on fuzzy logic is proposed. According to the motion state of the carrier, a Kalman filter adaptive adjustment parameter table is established. Each motion state has a corresponding adjustment parameter. The adaptive adjustment of the Kalman filter is the fuzzy recognition of the motion state. The basis for judgment mainly includes the acceleration of each axial movement, the angular velocity, and the strength of the geomagnetic field. According to the principle of maximum membership, the state of motion of the carrier is judged separately.

Assuming that the corrected $Y$-axis gravity field component is $g_y$, the estimated error of motion acceleration is $a_y$, the $Z$-axis gravity field component is $g_z$, and the estimated error of motion acceleration is $a_z$. The estimated error of motion acceleration is $K$, and then, the estimated error of pitch angle is as follows:

$$
\Delta \psi = 0.2 \cdot \left[ (g_y - g_z + K)^5 - (g_y)^5 + (g_y - g_z + K)^3 - (g_y)^3 \right].
$$

In a certain attitude, the angle estimation error and the motion acceleration estimation error are approximately linear changes.

In the two-step Kalman filter, sensors with different characteristics are treated differently to reduce the cross-effects caused by filter failure or reduced accuracy. According to the estimation of different motion states of the carrier, the measurement accuracy of each sensor is judged, and the variance matrices $Q$ and $R$ are adjusted adaptively, which better reflects the change in the estimation accuracy of each filter. This avoids the dependence on the statistical data of innovation, reduces the judgment time of system failure, and has better dynamic characteristics. Moreover, the introduction of fuzzy logic and the establishment of the motion state transform the adaptive adjustment of the filter into the judgment and recognition of the motion state. This not only can effectively avoid the divergence of the filter but also avoid the mathematical modeling between sensor output and filter adjustment. It can rely more on a large amount of adjustment experience and system error analysis to establish fuzzy logic judgment rules, which simplifies fault detection. And the isolated algorithm greatly reduces the computational burden of fault tolerance estimation. According to the incomplete carrier state data, the adjustment factor of the filter can be directly determined, which realizes a more accurate judgment of the sensor error distribution and improves the estimation accuracy of the filter.

### 4. Human Motion Pose Capture and Recognition Experiment

#### 4.1. Experimental Program Design
First, the human body motion posture capture experiment is based on the characteristics of the robot arm running smoothly and can more accurately control the motion speed and position; then, we design the human motion motion for experimental research and adopt the motion posture capture method to carry out attitude angle capture; finally, we use VR technology and Sim Mechanics simulation module to realize the visual verification of human motion posture capture, compare the results with the real situation, and draw corresponding conclusions.

The wireless motion capture system unit is installed at each joint position of the human body to realize the posture measurement of the limbs. The Zigbee wireless network is responsible for data transmission between the motion capture unit and the PC. According to the real-time posture of the limbs, the joint movement of the character model is driven. The designed three-dimensional motion capture method is used to calculate the initial attitude angle and dynamic attitude angle of the manipulator. At the same time, the initial angle of the robotic arm can be obtained through the control algorithm, and the true rotation angle and angular velocity can be measured through the circular magnetic grid, and then, the values of the angular velocity and the angular velocity can be compared.

Based on a series of research and design of human motion recognition, the experimental plan is shown in Figure 5.

In the experimental research of human motion recognition, different types of motion programs are designed and experiments are performed separately. This paper uses these experimental schemes to verify the recognition performance of SVC in different situations. The research of human movement recognition in this paper is oriented to the daily movement of the human body.

#### 4.2. Experimental Verification of Motion Pose Capture Method
Due to the large uncertainty of human motion, the motion pose capture method designed in the previous article cannot be directly verified. However, the robotic arm can
perform more precise control of its own motion from the control algorithm and can record the operating angle and angular velocity of the robotic arm through the circular magnetic grid that measures the angle of the mechanism.

4.2.1. Verification of the Initial Attitude Angle Capture Method. We collect the acceleration data of the initial position of the parallelogram mechanism. The initial position acceleration acquisition process needs at least 0.8 s to collect enough data. The calculated angle between the Z axis and the X axis and the horizontal position is shown in Figure 6.

It can be seen from Figure 6 that the angle between the X axis and the horizontal position is between 32° and 36°, and the average value is 33.8°; the angle between the Z axis and the horizontal position is between 0.9° and 0.98°, and the average value is 0.94°. This is not much different from the ideal angle. Considering that there is a certain error in the fixation of the data acquisition device, it can meet the requirement of capturing the initial attitude angle.

4.2.2. Verification of Dynamic Attitude Angle Capture Method. During the rotation process, the angle and angular velocity data are measured by the circular magnetic grid, and the data obtained are shown in Figure 7. In addition, the angular velocity data collected by the gyroscope and processed by noise reduction are shown in Figure 8.

It can be seen from Figure 8 that the Z-axis angular velocity collected by the gyroscope varies from −3 to 2, which indicates that the angular velocity data after noise reduction processing reflects the real motion of the robotic arm. The angular velocity data of the remaining two axes remain basically stable, with only small fluctuations not exceeding ±0.1°, and the interference to the Z axis of the gyroscope can be ignored. The dynamic attitude angle calculated using the data in Figure 8 and calculated according to the motion attitude capture method is shown in Figure 9. For comparison with the actual motion attitude angle data, the initial attitude angle is not considered.

It can be seen from Figure 9 that the dynamic attitude angle data of the Z axis of the gyroscope is basically the same as the angle change value of the parallelogram joint of the robotic arm. The root mean square error of the dynamic attitude angle and zero position of the X axis and Y-axis are calculated to be 0.036 and 0.032, respectively. For the nonworking X axis and Y-axis of the gyroscope, the calculated angle fluctuation value is extremely small.

After the experimental verification based on the robotic arm, from the capture results of the initial attitude angle and the dynamic attitude angle, the motion attitude capture method designed in this paper shows good performance in motion attitude capture with small errors, so it can be used.

4.3. Experiment and Analysis of Human Body Motion Posture Capture. The movement of the human body has strong uncertainty and instability. In the process of human movement, the speed, trajectory, and start and stop positions cannot be known in advance and can only move roughly along the expected trajectory to collect motion data. Based on the abovementioned human movement characteristics, this article designs human movement movements and conducts experiments and analyses by capturing the movement process of these two movements of the same volunteer.
Figure 6: Initial attitude angle data: (a) the angle between the $X$ axis and the horizontal position; (b) the angle between the $Z$ axis and the horizontal position.

Figure 7: Measurement data of circular magnetic grid: (a) angle data; (b) angular velocity data.

Figure 8: Continued.
Figure 8: Sensor tag data acquisition device collects data: (a) Z-axis angular velocity; (b) Y-axis angular velocity; (c) X-axis angular velocity.

Figure 9: Calculation results of the motion pose capture method: dynamic attitude angle of (a) Z axis, (b) Y axis, and (c) X axis.

Figure 10: Action angular velocity data of (a) X-axis, (b) Y-axis, and (c) Z-axis.
In order to study the results of the experiment, the process of the forward swing movement is stipulated as follows: the torso of the volunteer can stand in any direction. The human body swings in front and repeats several times and finally returns to the initial position and remains still.

The acceleration and angular velocity data of the action are collected, and the angular velocity data obtained after applying the wavelet threshold to reduce noise are shown in Figure 10.

It can be seen from Figure 10 that the volunteer’s upper limbs perform forward swing motion for 0.8 seconds. Figure 10(a) shows the angular velocity data of the X axis, which fluctuate greatly. Figures 10(b) and 10(c) are the angular velocity data of the Y axis and Z axis. It can be seen that the angular velocity data of the two axes have a small range of fluctuations after the start of the action. These movements that deviate from the expected movement are all human movements.

In this experiment, the first 20 sets of acceleration data after applying wavelet threshold denoising in the static state are used to capture the initial attitude angle. The calculated initial roll angle $\gamma$ and pitch angle $\theta$ of the human body are shown in Figure 11.

The static state of the human body is not absolute and has a certain amount of jitter. Since the acceleration data used to capture the initial attitude angle has a very small fluctuation range, we take the average value of the initial attitude angle as the capture result. The average values of the initial roll angle $\gamma$ and the initial pitch angle $\theta$ are $-3.12^\circ$ and $1.14^\circ$, respectively; that is, the angles between the Y axis and X axis of the data acquisition device and the plane in the geographic coordinate system are $-3.12^\circ$ and $1.14^\circ$, respectively. The capture result of the initial posture angle shows that the human body deviates from the plumb line to the rear of the human body by a certain angle before the movement starts.

In Figure 12, the human body dynamic posture angle data and angular velocity data reflect the same movement trend. In Figure 12(a), the roll angle $\gamma$ reflects the forward swing motion of the human body, and several small swings occur in the intermediate stage. In the beginning stage, the roll angle $\gamma$ keeps the initial attitude angle unchanged, and in the end stage, the human body roll angle $\gamma$ basically returns to the zero position, which means that the human body basically maintains the plumb weight state after the exercise. The periodic changes in the upper limb pitch angle $\theta$ and yaw angle $\psi$ in Figures 12(b) and 12(c) are caused by the uncertainty and instability of human motion.

Using the MATLAB/VR toolbox, combined with the capture method of the initial posture angle and dynamic posture angle of human motion, we apply the collected acceleration and angular velocity data of the action to perform VR-based visual simulation and verification. The visual model in the VR environment still cannot show all the details of the human body movement. The human body is modeled through the Sim Mechanics simulation module, and according to the human body posture angle, the trajectory of the end of the human body in space can be obtained.

4.4. Human Motion Recognition Experiment. We define the category labels of the standing and exercise states as 1 and 2. Because the two-classification mode is relatively simple, this experiment collected the motion data of 2 volunteers and performed noise reduction and segmentation operations and obtained 68 sets of feature values. After all eigenvalues are processed by dimensionality reduction, the first 10 principal components reach an interpretation level of more than 90%.

Three optimized SVCs, GS-SVC, GA-SVC, and PSO-SVC, are used for training and recognition. Under the K-CV method, the SVC is optimal for the training set. Figure 13 shows the partial results of human motion gesture capture based on video recognition.

The training and recognition results of the two-classification mode using three optimized SVCs are shown in Figure 14. It can be seen from Figure 14 that, consistent with our previous analysis of SVC, SVC has excellent
Not only does the accuracy rate of all optimization algorithms reach more than 90% but also the final recognition accuracy rate has reached more than 90%. In addition, the time required to capture and recognize human motion gestures is less than 2 s.

5. Conclusion

In view of the different characteristics of sensors in the inertial motion attitude capture system, this paper proposes a two-step Kalman filter, which processes the accelerometer and magnetometer separately and simplifies the filtering
algorithm from 9-dimensional to 3-dimensional matrix operation. The algorithm improves the calculation speed and enables the angle estimation algorithm to realize high-speed sampling estimation in the embedded system. In order to improve the estimation accuracy of the filter, an adaptive adjustment method based on fuzzy logic is established. The corresponding adjustment factor is determined for different motion states, and the adjustment of the filter is converted into the recognition of the motion state, which reduces the computational burden. After completing the research on the calculation method of human body posture angle, the research work of human body capture is completed by capturing the initial and dynamic posture angle of the human body. For this process, this article uses VR technology to establish a visualization model and simulate the human body capture process. This paper studies the recognition of human motion based on SVM. First, the SVC for human motion recognition is designed based on SVM theory, and then, a variety of optimization methods including grid search algorithm and heuristic algorithm are used to optimize the design of SVC to obtain better performance. Combining the research of human motion posture capture and recognition, we formulate the corresponding experimental plan. For human body capture, the motion posture capture method is verified, and the human body motion is designed for the capture experiment. The capture process was verified through a visual model. For the capture and recognition of three-dimensional human motion gestures, a two-classification model is designed for experiments, and the experimental results are analyzed. The sensor static correction method is divided into two steps: online data collection and offline processing. However, when in different experimental environments, it is necessary to reprocess the error correction of the sensor and modify the corresponding error model, which brings great inconvenience to actual use. Therefore, it is necessary to introduce online correction of the sensor.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported in part by the 2020 General Program of the Ministry of Education on Humanities and Social Sciences (An Experimental Study on the Physical and Mental Problems of Autistic Children and the Intervention, Subject number 20YJC890033).

References


