

Research Article

Design of Precise Estimation Algorithm for Sports Pose Based on Multifeature Fusion

Gao Songai, Zhang Ningning, and Zheng Xiang 

Chengde Medical College, Chengde 067000, Hebei, China

Correspondence should be addressed to Zheng Xiang; gsiii@cdmc.edu.cn

Received 23 August 2022; Accepted 17 September 2022; Published 11 October 2022

Academic Editor: Imran Khan

Copyright © 2022 Gao Songai 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.

The estimation time and estimation precision of motion pose samples are problematic for the pose estimation algorithm of sports movements. This paper proposes a multifeature fusion-based algorithm for accurate posture estimation. The human rod model is constructed after analyzing the human pose estimation technology. Using the Kalman filter method, the degree of freedom and range of motion of the major joints of the human body were determined. The eight-star model was used to extract the sports posture features, and the weighted average method was used to process the grayscale images of sports. Using the multifeature fusion method, the extracted multisource feature vector information is thoroughly analyzed and processed, and a new group of fusion feature vectors is created. Using a mixture Gaussian distribution model, the posture estimation of an athlete's body is accomplished. Experimental results indicate that when the amount of sports pose sample data is 900 GB, the accurate estimation time of the proposed method is 5.3 s, and its accuracy is 100 percent. Improve the estimation accuracy of samples of sports posture.

1. Introduction

With the advancement of society and the improvement of living standards, people are beginning to pay more attention to physical health and other aspects and are constantly improving their physical quality through sports. In the process of movement, standard movement posture can not only determine the effect of movement to a certain extent but also provide the greatest possible protection against injury [1]. In badminton, for instance, the use of standard badminton movements for daily exercise can not only effectively exercise the body but also improve a player's level of competition [2]. However, the definition of standard motor posture during movement is primarily based on images or verbal instruction, resulting in the absence of quantitative evaluation criteria for standard motor posture. Nevertheless, the estimation of human pose and recognition of human motion from the collected image or video sequence provide a theoretical foundation for measuring standard motion pose. Early human action recognition requires external equipment to perceive the change in human posture in order to

recognize human action. With the development of machine learning and deep learning, there are numerous research directions in the academic field, including image processing, SVM classifier, and deep neural network [3]. These advancements have made it possible for computers to detect human motion using only devices like cameras, thereby drastically reducing the number of external perceptrons.

Human pose estimation is the most critical technology in the process of sports pose recognition. In order to guide people's posture during the movement process, a series of features derived from human pose estimation are compared with some standard actions. Consequently, the human motion assessment system has evolved significantly. Simultaneously, people's focus is gradually shifting from the professional type to the portable type: a portable human motion assessment system that is unaffected by location and environment can enhance the quality of people's daily exercise. Human pose recognition has garnered the interest of numerous domestic and international research institutes and university-level laboratories [4] due to its high theoretical value and vast application potential. A number of

algorithms for human pose recognition and related specific problems have been proposed domestically and internationally. In addition, advances have been made in the areas of human object recognition, keyframe extraction, human pose estimation, and other specialized issues. The traditional human gesture recognition technology has low recognition rate, high difficulty, high cost, slow speed, and large storage space for human behavior recognition, which is not the requirement for the application of human-computer interaction technology in the future. Combining the keyframe selection technique with the human pose estimation technique drastically reduces computation and saves storage space. In recent years, the research status of human pose estimation at home and abroad is summarized as follows [5].

The authors of [6] proposed a 3D human pose estimation method based on semisupervised learning convolutional neural network, collected video data containing motion pose, built a semisupervised learning convolutional neural network model, and estimated the corresponding 3D pose skeleton changing pose. This method can accurately estimate the 3D pose of the corresponding frame in a video, but sports pose estimation efficiency is low. The method proposed by the authors of [7] for acquiring specific motion frames in motion videos is based on human pose estimation and clustering. Select the HRNet pose estimation model as the basis, establish the model lightweight process and combine with the DARK data encoding, construct the Small-HRNet network model, extract from the video human body joint points, and use the human skeleton features of each frame of video as sample points. The standard skeleton feature of the motion frame is the cluster center, which clusters the entire video. The video's specific motion frame is obtained through clustering, and the experiment is conducted on the motion dataset. This method can reduce the number of body pose estimation model parameters while maintaining accuracy, but the classification effect of the proposed method is inadequate. A human pose estimation optimization algorithm for deformable models is proposed with reference to [8]. The model parameters are extracted from each frame of a color image by a neural network, and the parameters are optimized and solved using human key points and contour as constraints. According to the interframe coherence of the video sequence, the error of the pose estimation results of all video frames is then corrected, making the motion sequence more fluid and smooth. The point cloud obtained from the depth map and the corresponding color map were used as the joint input, followed by the use of the distance constraint between the point cloud and the corresponding point of the model to optimize and solve the problem. Eventually, a result resembling the actual human pose was achieved. Using point cloud data optimization, the algorithm can effectively correct the single-frame pose estimation results and greatly improve its accuracy. Nonetheless, the accuracy of posture estimation in sports is poor.

This paper proposes an accurate estimation algorithm for sports pose based on the fusion of multiple features.

2. Acquisition of Sports Posture Features

2.1. Human Pose Estimation Techniques. According to the Oxford Dictionary, a posture is a particular posture of the body, and the way a person maintains his or her body. The human pose estimation method is to extract, classify, distinguish, and describe the pose features of the human body, which has been widely paid attention to in recent years. The application of human physiology, digital image processing, pattern recognition, and other disciplines is an interdisciplinary research field [9]. From the perspective of the collection method of human pose information, human pose estimation technology can be divided into two categories: contact recognition and noncontact recognition [10]. According to different classification criteria, it can be divided into a variety of human pose recognition algorithms. From the perspective of implementation methods, it is usually divided into three types: 3D model reconstruction method, which extracts 3D features from effective samples to construct 3D models, human appearance model method, by obtaining the shape features of the human body to establish a two-dimensional model and use the model matching method to complete the recognition, motion model method, according to the motion characteristics of the classification. From the perspective of pattern recognition, attitude recognition is a problem of classifying time-varying features, that is, matching the test sequence with the preset sequence according to the obtained feature information.

2.2. Construction of Human Rod Model. According to the knowledge of human anatomy, there are more than 200 bones in the human body, and the bones are connected by joints, which makes the whole system extremely complex, and the actual data acquisition is very complicated. Due to the particularity of the acquisition unit used in this system, we could not collect all the data for each bone [11].

At present, most of the studies on the human body are simplified modeling and analysis with the rod model; that is, the main joint of the human body is abstracted into a point, and the limbs between the joints are abstracted into a link, mainly considering elbow, shoulder, hip, hip, knee, ankle, and other joints. In this way, the whole human body is simplified into a rod-like model, which is also the premise for the analysis of human posture, as shown in Figure 1.

After the simplified rod human body model has been established, the range of motion of each joint must be constrained in order to bring the model closer to the normal human posture. In rehabilitation medicine, the initial posture of human movement is defined as follows: the body is upright and facing forward, the eyes are flat, the feet are standing together, the toes are pointing forward, the upper limbs are hanging at the side of the body, and the palm is attached to the outer thigh of the side of the body. The joint has three components: the articular surface, the articular capsule, and the articular cavity [12]. The shape and structure of the articular surface determine the axis of joint activity, and the freedom of joint movement is closely associated with the movement around the axis. There are

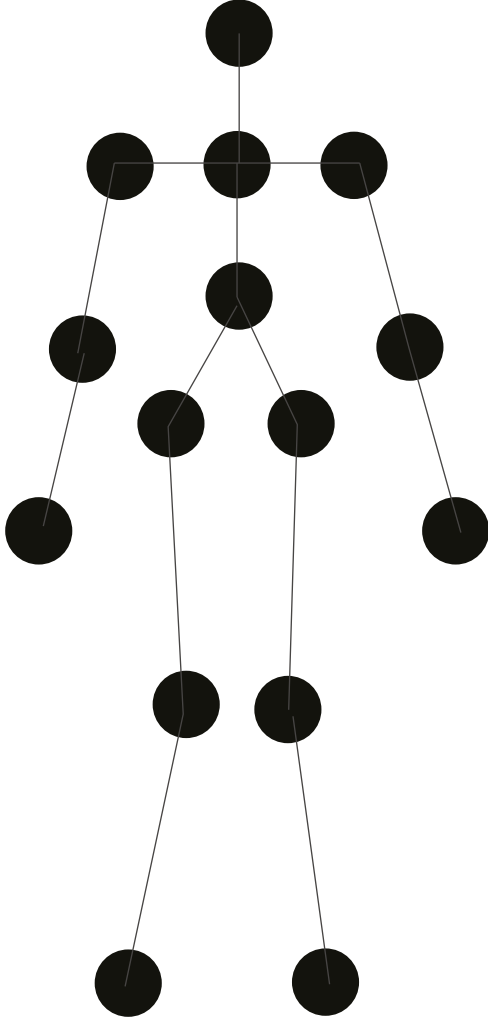


FIGURE 1: Schematic diagram of rod mannequin.

several degrees of freedom if there are multiple directions of joint activity. All joints with more than two degrees of freedom are capable of producing circumferential motion, and the limb typically rotates about the joint axis. The degrees of freedom and motion range of each major human joint are as follows:

- (1) Shoulder joint: there are three degrees of freedom, and the upper arm sags in a neutral position. Its range of motion is forward flexion: 70~90°, posterior extension: 40~45°, forward flexion: 150~170°, upward lift: 160~180°, abduction: 80~90°, adduction: 20~40°, internal rotation: 70~90°, external rotation: 40~50°.
- (2) Elbow joint: there is one degree of freedom. The neutral position of the elbow joint is forearm extension. Flexion: 135~150°, overextension: 10°, pronation: 80~90°, supination: 80~90°.
- (3) Wrist joint: there are two degrees of freedom. The neutral position is the hand in line with the forearm, and the palm is downward. Dorsal extension: 80~90°, palmar flexion: 50~60°, radial tilt: 25~30°, ulnar tilt: 30~40°.
- (4) Hip joint: there are three degrees of freedom. The neutral position is hip joint extension and patella forward. Flexion can reach 130~140°, posterior extension can reach 10~15°, abduction can reach 30~45°, adduction can reach 20~30°, knee flexion 90°, normal external rotation 30~40°, internal rotation 40~50°.
- (5) Knee joint: there are two degrees of freedom. The neutral position is knee joint extension, flexion: 120~150°, hyperextension: 5~10°, internal rotation about 10°, and external rotation: 20° when bending the knee.
- (6) Ankle joint: there are two degrees of freedom. The neutral position is a 90 degree Angle between the foot and the calf, without varus or valgus. Dorsiflexion is about 20~30°, plantar flexion is about 40~50°, varus is 30°, and valgus is 30~35°.

In this paper, attitude solution refers to the use of existing nine-axis sensor data and relevant theories and algorithms to get attitude Angle information from these data. Attitude solving algorithm belongs to the category of data fusion, that is, attitude solving by fusing multiple sensor data [13]. The commonly used algorithms for attitude resolution include Kalman filter, extended Kalman filter, complementary filter, and gradient descent algorithm.

The model-based linear minimum variance estimation of Kalman filter is a very widely used attitude calculation algorithm. As an important optimal estimation theory, it is widely used in various fields such as automatic control and aerospace [14]. Since discrete Kalman filter algorithm can be directly implemented on computer to estimate multidimensional stochastic processes, it will be introduced in the following.

Let the system noise sequence W_{k-1} act on the estimated state X_k at time t_k and can be described as follows:

$$X_k = \Phi_{k,k-1}X_{k-1} + \Gamma_{k-1}W_{k-1}. \quad (1)$$

The measurement equation is a linear function, as shown in the following equation:

$$Z_k = H_kX_k + V_k, \quad (2)$$

where $\Phi_{k,k-1}$ is the one-step state transition matrix from time t_{k-1} to time t_k , Γ_{k-1} is the system noise control matrix, H_k is the observation matrix, V_k is the measurement noise sequence, and w_k is the disturbance noise vector of the system [15].

Assuming that the system noise variance matrix Q_k is nonnegative definite and the observation noise variance matrix R_k is positive definite, the estimator can be solved in the following way:

Single-step state prediction equation is as follows:

$$x'_{k,k-1} = \Phi_{k,k-1}x'_{k-1}. \quad (3)$$

State estimation equation is as follows:

$$x'_k = x'_{k,k-1} + K_k [Z_k - H_k X'_{k,k-1}]. \quad (4)$$

Mean square error estimation matrix is as follows:

$$P_k = [I - K_k H_k] P_{k,k-1}, \quad (5)$$

where the initial values P_0 and x'_0 need to be given in advance, so that the state estimate x'_k at time k can be obtained from the measurement equation Z_k at time k recursively.

2.3. Attitude Feature Extraction Based on Eight-Star Model.

The image of the human object is a two-dimensional array. The extracted features of the human object are transformed into a one-dimensional feature vector, which greatly reduces the feature vector's dimension and computational complexity. This article focuses on walking, running, and jumping in three postures for feature extraction and recognition, through the analysis of three kinds of attitude, and the legs move in arm swinging amplitude, arising in the process of adopting the star model to the human body modeling, namely through the calculation of center of mass of human movement target image points and local contour pole and its mutual relations, in order to obtain the attitude. In this paper, an eight-star model is utilized to extract a set of feature vectors containing seventeen feature values to characterize the human pose in each frame [16]. Compared to the six-star model, it can describe the human body's contours, including the head and limbs, more precisely, and it is not redundant.

Eight-star model is an improvement of the traditional star model, which is a feature description model formed according to the eight local contour poles and centroid points of the human object. The human pose model is established as shown in the following equation:

$$Fet = \{D, A, e\}, \quad (6)$$

where D is the distance between the eight local contour poles of the eight-star model and the centroid of the human target. A is the minimum Angle formed by the straight line and horizontal line from the eight local contour poles to the centroid point of the eight-star model; e is the eccentricity of the human target. The specific feature extraction process is as follows:

Calculate the centroid of human object (x_c, y_c) :

$$x_c = \frac{1}{N_b} \sum_{i=1}^{N_b} x_i, y_c = \frac{1}{N_b} \sum_{i=1}^{N_b} y_i, \quad (7)$$

where N_b is the number of target pixels.

The target is divided into four parts by horizontal lines and vertical lines passing through the center of mass. Based on the overall contour of the target, eight local contour pole coordinates (x_i, y_i) of the rightmost, uppermost, leftmost, and bottommost contour in each part are extracted, respectively.

The Euclidean distances between the eight local contour poles (x_i, y_i) and the centroid point (x_c, y_c) are calculated,

which are, respectively, expressed as d_i , and the calculation formula is shown in the following equation:

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}, i = 1, 2, \dots, 8. \quad (8)$$

Then, $D = [d_1, d_2, \dots, d_8]$. The extraction process is shown in Figure 2.

After feature extraction and fusion, the feature vector of the human pose in the video at frame t is shown in the following equation:

$$Fet_t = \{d_1, d_2, \dots, d_8, a_8, e\}. \quad (9)$$

The feature vector is a 17-dimensional feature vector constructed from the local contour pole distance feature, local contour pole Angle feature, and eccentricity feature of the human eight-star model, where d_i is the pole distance feature of the eight-star local contour, a_i is the Angle feature of the human local contour pole, and e is the eccentricity of the human body. The computation complexity and memory of the gesture description operator are relatively low, so the system can achieve a good real-time recognition effect.

In practice, the video image will contain some noise and affect the effect of detection and recognition. Therefore, the primary task of video image processing is image pre-processing, and its effect directly affects the subsequent detection and recognition.

2.4. Sports Color Image Grayscale. The color value of the true color image pixel is composed of three components: red R, green G, and blue B. The value of each component has 256 levels of brightness, which is represented by numbers from 0 to 255. Therefore, there are 224 colors, which is relatively complicated to process directly. The gray image is a special color image, whose R, G, and B components are all equal, so there are only 28 colors left, which is convenient for image processing [17]. Therefore, in order to realize the real-time performance of the algorithm, the original image is converted into gray image before the video image processing, so as to reduce the computation amount of the algorithm. In general, the gray-level methods used include component method, maximum method, average method, and weighted average method.

In the component method, one of the three components of RGB is used as the gray value of the pixel, as shown in the following equation:

$$I = R \quad \text{or} \quad I = G \quad \text{or} \quad I = B, \quad (10)$$

where I represents the pixel brightness of gray image, and R, G, and B, respectively, represent the brightness values of RGB components. A gray image can be selected according to the application needs. Taking the 7th frame of LENA_WALK1 video, the 10th frame of PETS2006 video set, and the 21st frame of test video set in Weizmann database as examples, the grayscale is realized by component method.

The above method is simple to implement but lacks consideration of the importance of different components. According to the importance of components and other needs of images, the weighted average of RGB three

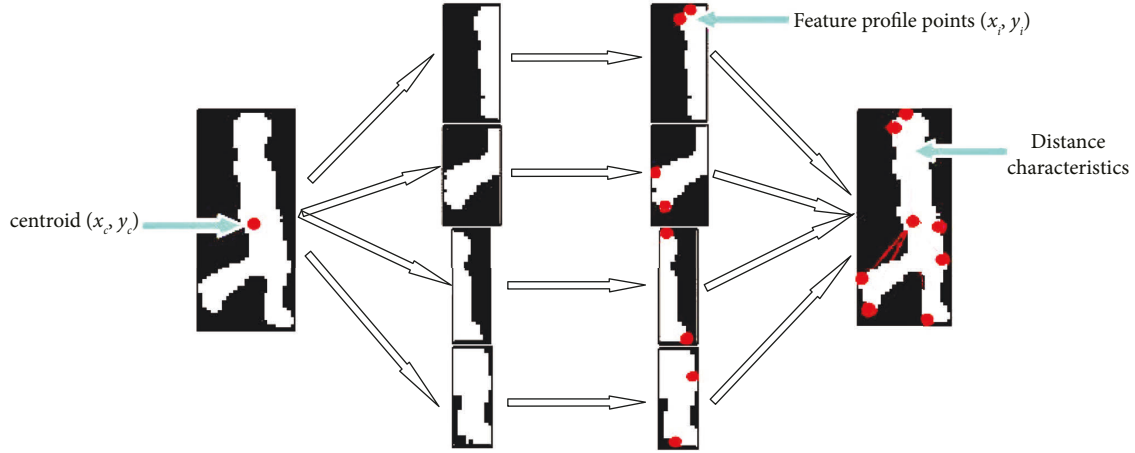


FIGURE 2: Feature extraction process of pole distance of local contour of human eight-star model.

components can be taken as the gray value of pixels [18]. From the perspective of human physiology, the human eye is sensitive to green and insensitive to blue. Therefore, according to the visual model of human eyes, the weight of G component is set to be large, while the weight of B component is small, and its commonly used formula is shown in the following equation:

$$I = 0.3R + 0.5G + 0.11B. \quad (11)$$

The gray weights adopted by OpenCV open source computer vision library also follow the visual model of human eyes, and the weighted average formula is shown in the following equation:

$$I = 0.212671R + 0.715160G + 0.72169B. \quad (12)$$

From the perspective of theory and practice, the weighted average method is more reasonable and more in line with the needs of practical application, so the weighted average method is the most commonly used method of image graying. In this paper, the weighted average method is used to gray-scale the image [19], and the `cvCvtColor` function in the OpenCV computer vision open source library is used to complete the gray-scale of the image by setting its third parameter as `CV_BGR2GRAY`. Or using `cvLoadImage` function, by setting its second parameter to 0 can realize the color image gray.

3. An Accurate Estimation Algorithm for Sports Posture

3.1. Median Filtering Method for Posture Image Processing. In this method, the arithmetic mean or median of the pixels of the same coordinate in the continuous image frame is taken as the pixel value of the point in the background model, so as to realize the background modeling. The background modeling process is shown in (13) and (14), respectively.

$$B_t(x, y) = \frac{1}{k} \sum_{i=1}^k I_{t=i}(x, y). \quad (13)$$

$$B_t(x, y) = \text{med}(I_i(x, y)_{i=1,2,\dots,k}). \quad (14)$$

Among them, `med` means taking the median operation, $I_t(x, y)$ means the video frame at time t , and $B_t(x, y)$ means the updated background model at time t . However, it can be seen from formula (14) that the memory requirement increases because k frames of images need to be cached for updating the background model. Therefore, in order to reduce memory and computation, the background model parameters are usually updated according to the following equation:

$$B_{t+1}(x, y) = \lambda I_t(x, y) + (1 - \lambda)B_t(x, y), \quad (15)$$

where λ is called the update rate, which controls the update rate of the background model. Through experimental verification, in order to reduce the influence of foreground change on the model, the value of λ is usually small, generally set at 0.05.

3.2. Multifeature Fusion of Human Motion Posture. Multifeature fusion refers to the comprehensive analysis and processing of multisource feature vector information extracted from different feature extraction methods through feature fusion algorithm to form a new fusion feature vector group. In general, it can be divided into two categories: direct feature combination and feature selection combination. Direct feature combination method is to directly synthesize new feature vector groups by all feature vectors according to some simple rules, for example, serial and parallel fusion method. The method of feature selection and combination is to put all the original feature vectors together, and according to some selection rules, select and retain some feature vectors from the original feature vectors as a new combination of feature vectors.

Serial feature fusion method: the feature vectors in the sample space are directly merged into new feature vectors in turn. Assume that A and B are standardized features in sample space Ω , any sample $\xi \in \Omega$, feature vector is expressed as $\alpha \in \Omega$, and $\beta \in \Omega$ fused feature is expressed as

$$\gamma = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \quad (16)$$

Parallel feature fusion method [20]: two groups of feature vectors in the sample space are merged into a new feature vector by a complex vector. Assuming that A and B are standardized features in sample space Ω , and any sample $\xi \in \Omega$, feature vectors are expressed as $\alpha \in \Omega$ and $\beta \in \Omega$, and then, the fused features are expressed as

$$\gamma = \alpha + i\beta. \quad (17)$$

In (17), i is an imaginary unit. If the dimensions of features α and β are not equal, feature fusion will be carried out after zero-complementing for features of low dimension.

The serial fusion method is the simplest and most effective method for fusing features. The dimension of the fused feature is increased, all feature information is retained, and there are numerous redundant feature vectors. The parallel fusion method guarantees the invariance of feature dimension, but the use of complex space vectors increases the operation's complexity, which has an effect on the recognition's real-time performance. In order to improve accuracy and real-time performance, the feature selection method generates new features that can ensure the effectiveness within and between feature classes without generating an excessive amount of redundancy. Feature level fusion method is roughly divided into five categories: probability statistical method (Bayesian estimation, Kalman filter, etc.), logic reasoning method (D-S evidence theory, fuzzy logic, etc.), neural network methods, fusion method based on feature extraction (PCA, LDA, etc.), and fusion method based on search (genetic algorithm, particle swarm optimization (psa), etc.).

3.3. Human Pose Estimation in Sports Based on Mixed Gaussian Distribution Model. The method of mixed Gaussian background modeling is to update the parameters dynamically, so that the background model established by it has better adaptability. Its basic principle is to assume that every pixel in the image is independent and can be mixed by K Gaussian distributions, where K is generally 3~5.

At time t , the pixel value at position (i, j) is x_t , and the probability of each pixel value can be represented by the weighted sum of K Gaussian probability density functions, as shown in the following equation:

$$p(x_t) = \sum_{i=1}^K \omega_{i,t} \times \eta \left(x_t, \mu_{i,t}, \sum_{i,t} \right), \quad (18)$$

where $\omega_{i,t}$ is the weighting coefficient of the i th Gaussian distribution at time t , the mean and covariance matrix of the i th Gaussian distribution at time t are $\mu_{i,t}$, $\sum_{i,t}$, and

$\sum_{i=1}^K \omega_{i,t} = 1$, and the probability density function of its Gaussian distribution is $\eta(x_t, \mu_{i,t}, \sum_{i,t})$. The specific calculation is shown in the following equation:

$$\eta \left(x_t, \mu_{i,t}, \sum_{i,t} \right) = \frac{1}{(2\pi)^{n/2} |\sum_{i,t}|^{1/2}} e^{-1/2 (x_t - \mu_{i,t})^T \sum_{i,t}^{-1} (x_t - \mu_{i,t})}, \quad (19)$$

when a new video frame is input, according to (19), the pixel value x_t of each newly input image is successively compared with the K models already established at the current pixel point of this position. Where D is the confidence parameter, usually set at 2.5.

$$|x_t - \mu_{i,t-1}| \leq D \times \sigma_{i,t-1}. \quad (20)$$

If the new input pixel value x_t satisfies (20), then x_t is judged as the background point after successful matching; otherwise, it is the foreground point. After the new video frame input pixel matching is completed, the model parameters of this point are updated and $M_{k,t} = 1$ is taken, and the parameter update method is shown in equations (18) to (20), where α is the learning rate is a fixed value, and the value is $0 < \alpha < 1$, and ρ can be updated through the following equation:

$$\omega_{k,t} = (1 - \alpha) \omega_{k,t-1} + \alpha \times M_{k,t}, \quad (21)$$

$$\mu_t = (1 - \rho) \times \mu_{t-1} + \rho \times x_t, \quad (22)$$

$$\sigma_t^2 = (1 - \rho) \times \sigma_{t-1}^2 + \rho \times (x_t - \mu_t)^T (x_t - \mu_t),$$

$$\rho = \alpha \times \eta(x_t | \mu_k, \sigma_k). \quad (23)$$

If no matching is achieved, $M_{k,t} = 0$ is taken, which is only newer (23), and the rest is unchanged. If none of the models is matched successfully, then the model with the lowest weight in the mixed Gaussian model is replaced, and then, the new parameter mean is x_t , the standard deviation is the initial maximum value σ_0 , and the weight is the small value ω_0 .

In order to obtain the model with good correlation between background and Gaussian model, that is, the model with heavy weight and small standard deviation, all Gaussian models are arranged from large to small according to ω/α^2 value to obtain the sports pose estimation function B , which is expressed as

$$B = \arg \min_b \left(\sum_{k=1}^b \omega_k > T \right). \quad (24)$$

In the formula, the parameter T represents the proportion of background. If x_t in the input video frame successfully matches any of the best description model, x_t is judged to be the background point. Otherwise, it is judged as the foreground point. At this time, the obtained sports pose estimation function value is the result of sports pose estimation, so as to achieve accurate sports pose estimation.

4. Experiment

4.1. Experimental Design. The experiment was conducted on two benchmark datasets, JPL First Person Interaction Dataset (tJPL) and DogCentric Activity Dataset (Dogcentric), containing 84 and 209 videos, respectively. The former is the human-computer interaction data set, and the latter is the animal-human interaction data set from the first perspective. The video in both data sets has a resolution of 320 240 and a frame rate of 30 FPS.

In this paper, multiple evaluation indicators are used to evaluate the model, including sample accuracy, sample level category accuracy, sample level average category accuracy, and video accuracy. Sample accuracy, also known as overall accuracy, refers to the proportion of correctly classified samples in the total number of samples and is the simplest and intuitive evaluation index in classification problems. Specifically, assuming that N sample segments are obtained from the video according to the sampling strategy, and the number of segments correctly classified by the model is N^{cor} , and then, the formula for calculating the sample accuracy is as follows:

$$A_{cc} = \frac{N^{cor}}{N}. \quad (25)$$

However, when the proportion of categories is seriously unbalanced, it is not reliable to rely only on the sample accuracy as the evaluation standard. Suppose an extreme case: when the proportion of negative samples is 99% of the total samples, the classifier can also obtain a high accuracy by predicting all samples as negative samples. When the data are skewed, the categories with a large proportion become the main factors affecting the assessment criteria. Therefore, we also use the sample level category accuracy and sample level average category accuracy. Among them, the sample-level category accuracy is to group all sample segments according to categories and obtain the probability of correct prediction in each category. The average category accuracy at sample level is averaged according to the weighted summation of each category accuracy. Specifically, assuming that the number of classes is $num_classes$ ($num_classes > 0$), each group is grouped according to the category, and the number of samples of $num_classes$ ($num_classes > 0$) in each group is N_n^{cor} ($0 \leq N_n^{cor} \leq N_n$), and then,

$$class(A_{cc}) = \frac{N_n^{cor}}{N_n}, \quad (26)$$

$$Mclass(A_{cc}) = \frac{\sum_{n=0}^{num_class-1} N_n^{cor} / N_n}{num_classes}$$

In order to obtain the video level prediction from the prediction metrics of the sample segments, the method described in Section 3.2 is used for voting fusion, that is, the fusion of the maximum number of category predictions, the maximum category scores, and the maximum weighted scores, and the estimation accuracy of sports posture is calculated according to the fusion results.

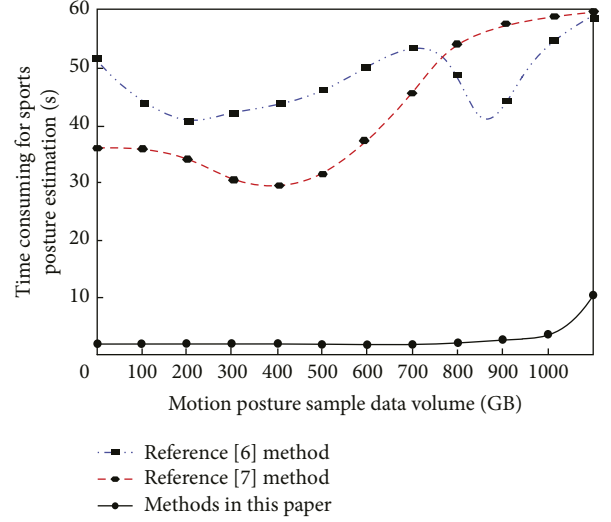


FIGURE 3: Time consumption for accurate estimation of sports posture.

4.2. Experimental Result

4.2.1. Accurate Estimation of Athletic Posture Takes Time.

In order to verify the precision estimation efficiency of the proposed method, the method in literature [6], the method in literature [7], and the method in this paper are used to estimate the precise estimation time of sports posture, and the results are shown in Figure 3.

Analyzing Figure 3 reveals that when the sample data amount of sports posture is 100 GB, the estimation time of the method in reference [6] is 43.5 s, the estimation time of the method in reference [7] is 35.2 s, and the estimation time of this method is 1.5 s. When the sample size of sports posture data is 300 GB, the estimation time of the method in reference [6] is 42.1 s, reference [7] method is 31.6 s, and the proposed method takes 2.3 s. When the sample data amount of sports posture is 900 GB, the accurate estimation time of the method in reference [6] is 46.5 s, the method in reference [7] is 58.3 s, and this method's accurate estimation time is 5.3 s. The posture accuracy of the proposed method is significantly higher than that of other methods, indicating that it can effectively improve the posture accuracy of sports.

4.2.2. Sample Accuracy of Posture Estimation in Sports.

In order to verify the sample validity of the proposed motion pose estimation method, literature [6], literature [7] and the methods proposed in the paper are used for comparison, and the results are shown in Figure 4.

According to Figure 4, when the amount of sports pose sample data is 200 GB, the estimation accuracy of the method in reference [6] is 72.5 percent, the method in reference [7] is 48.6 percent, and the proposed method is 95 percent. When the amount of data for the sports pose sample is 600 GB, the estimation accuracy of the method in reference [6] is 69.9%, and that of the method in reference [7] is 71.5%. The proposed method has a 99.9% estimation

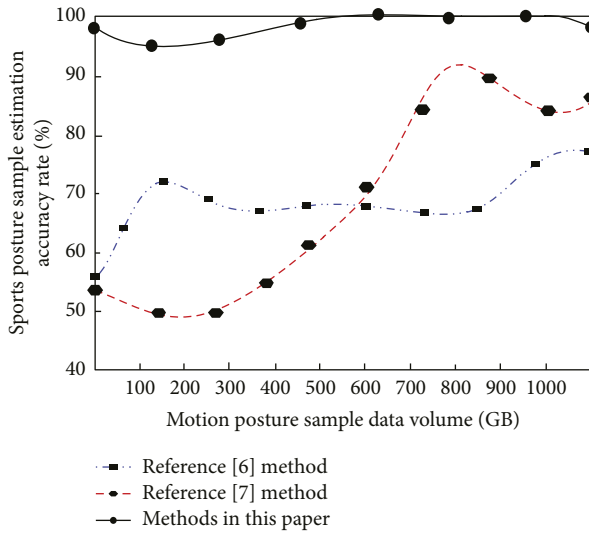


FIGURE 4: Estimation accuracy of sports posture samples.

precision. The estimation accuracy of the sports pose sample is 69.2% for the method in reference [6] and 89.6% for the method in reference [7] when the sample data size is 900 GB. The estimation precision of the proposed method is one hundred percent. This indicates that the proposed method can effectively improve the estimation accuracy of sports pose samples.

5. Conclusion

In this paper, a human rod model is constructed using a precise estimation algorithm for sports posture based on the fusion of multiple features. Based on the Kalman filter method to estimate the motion posture, determine the degree of freedom and range of motion of the main body joints, use the eight-star model to extract the motion posture features, use the multifeature fusion method to extract the multisource vector information features, and use the mixed Gaussian distribution model to realize the motion body posture. Experimental results indicate that when the amount of sports pose sample data is 900 GB, the accurate estimation time of the proposed method is 5.3 s, and its accuracy is 100 percent. Improve the estimation accuracy of samples of sports posture.

In the future, we will introduce a multihead attention mechanism and more advanced 6D recognition technology to further improve the accuracy of motion gesture recognition.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The thesis was supported by the Hebei Province Sports Science and Technology Research Project; Project Name: Research on Sports Health of the Urban Elderly in Hebei under the Background of “Integration of Sports and Medicine”; Project No.: 20202006.

References

- [1] N. Zhou, X. Jiang, and X. Wang, “Research progress of sports protective clothing based on biomechanics [J],” *Journal of clothing*, vol. 5, no. 03, pp. 210–215, 2020.
- [2] Y. Wang, H. Zhao, and W. Zhi, “SEMA4D under the post-transcriptional regulation of HuR and miR-4319 boosts cancer progression in esophageal squamous cell carcinoma,” *Cancer Biology & Therapy*, vol. 21, no. 2, pp. 122–129, 2020.
- [3] Z. Lv, F. pan, and X. Xu, “A new depth neural network compression model for airport target detection,” *Transactions of Nan Jing University of Aeronautics & Astronautics*, vol. 37, no. 04, pp. 562–573, 2020.
- [4] H. Song, Yi Zheng, and K. Wang, “Research progress of wearable human posture recognition system,” *Laser and infrared*, vol. 51, no. 09, pp. 1123–1128, 2021.
- [5] J. Wang and X. Lu, “Video key frame extraction algorithm based on semantic correlation,” *Computer engineering and application*, vol. 57, no. 04, pp. 192–198, 2021.
- [6] H. Lin, W. Li, Y. Wu, and Y. Tan, “Three dimensional human posture estimation based on semi supervised learning convolutional neural network,” *Modern computer*, vol. 27, no. 31, pp. 56–60, 2021.
- [7] M. Cai, J. Huang, L. Xiao, and X. Zhou, “A method for acquiring specific motion frames based on human posture estimation and clustering,” *Journal of graphics*, vol. 43, no. 01, pp. 44–52, 2022.
- [8] Li Jian, H. Zhang, and B. He, “Optimization algorithm of human posture estimation based on deformation model,” *Journal of Xi’an University of Engineering Science and Technology*, vol. 47, no. 02, pp. 23–31, 2020.
- [9] J. Gu, “Qualitative research on human physiology online teaching experience based on Bloom’s teaching goal classification theory,” *Nursing and rehabilitation*, vol. 21, no. 01, pp. 44–46, 2022.
- [10] B. Gao, Ke Ma, Bi Hongbo, and L. Wang, “Lightweight design of human posture estimation network,” *Laboratory research and exploration*, vol. 39, no. 01, pp. 79–82, 2020.
- [11] Y. Li, L. Fang, L. Liao, and Y. Yang, “Digital watermarking algorithm using 3D animation bone data,” *Computer application and software*, vol. 37, no. 05, pp. 123–126, 2020.
- [12] H. Cheng, Z. Shi, Bo Yu, Y. Zhu, and H. Ding, “Experimental research on the measurement of return difference of small joints of service robots,” *Journal of Instrumentation*, vol. 41, no. 05, pp. 48–57, 2020.
- [13] L. Chun, T. Liu, H. Zhang, W. Jixiang, and Z. Wang, “Application of wavelet transform in attitude calculation,” *Journal of Electronic Measurement and Instrument*, vol. 35, no. 01, pp. 183–190, 2021.
- [14] J. Zhou, X. Sun, K. Liu, M. Wang, X. Yang, and G. Liu, “Research on SOC estimation algorithm of sliding mode observer combined with extended Kalman filter,” *Chinese Journal of electrical engineering*, vol. 41, no. 02, pp. 692–703, 2021.
- [15] S. Li, Z. Xia, X. Cheng et al., “Continuous estimation method of grid inertia normalization based on noise like disturbance,”

- Chinese Journal of electrical engineering*, vol. 40, no. 14, pp. 4430–4439, 2020.
- [16] J. Liu, Y. Liu, X. Jia, and S. Guo, “Research on human pose recognition algorithm based on model constraint,” *Journal of Instrumentation*, vol. 41, no. 04, pp. 208–217, 2020.
- [17] L. Xu, H. Zhang, S. Yan, D. Gao, J. ye, and Y. Shao, “Analysis of factors affecting PSNR in image processing,” *Journal of Tonghua Normal University*, vol. 42, no. 04, pp. 44–48, 2021.
- [18] J. Xing, “An analysis of the gray value and exposure control of the digital conversion model (black and white) of ancient books,” *Digital and microfilm*, vol. 2, no. 02, pp. 15–18, 2020.
- [19] X Han, B. Guo, C. Li, and S. Xiao, “Research on parallel algorithm of image graying based on OpenCL,” *Journal of Jiangxi Normal University (NATURAL SCIENCE EDITION)*, vol. 44, no. 05, pp. 462–471, 2020.
- [20] D. Ma, S. He, C. Yang, and chunman Yan, “Semantic segmentation of feature fusion convolutional neural network,” *Computer engineering and application*, vol. 56, no. 10, pp. 193–198, 2020.