Research Article

3D Human Motion Posture Tracking Method Using Multilabel Transfer Learning

Libin Zhu and Lihui Liu

Physical Education Department, Suihua University, Suihua 152061, China

Correspondence should be addressed to Lihui Liu; 568025518@qq.com

Received 21 June 2022; Accepted 20 July 2022; Published 8 August 2022

Abstract

To overcome the high position and posture angle tracking error, long tracking loss time and posture tracking update response time, and low fitness problem of traditional human motion posture tracking methods, in this paper, a three-dimensional (3D) human motion posture tracking method using multilabel transfer learning is proposed. According to the human structure composition and degree of freedom constraints, the 3D human joint skeleton model is constructed to generate the 3D human pose image and perform the noise reduction operation. The background difference is used to detect the 3D human moving target. Using multilabel transfer learning, human motion posture features are extracted from joint position and joint angle, and the estimation results of 3D human motion posture are obtained. The tracking error of human motion posture is corrected by three-step search, and the visual 3D human motion posture tracking results are output. The results show that, compared with the traditional human motion posture tracking method, the position and posture angle tracking errors of the proposed method are 2.18 mm and 0.178 deg, respectively. The tracking loss time and posture tracking update response time are shorter, which proves that the proposed method has more advantages in tracking accuracy and higher adaptability.

1. Introduction

Human motion posture tracking is a technology that estimates the posture of the target human in each frame image through the analysis of continuous images and accurately captures the posture changes of the human in the continuous image sequence. The results of human motion posture tracking can be used as a reference for motion evaluation and objectively and accurately reflect the motion flexibility of human, and they can also be used as a diagnostic reference for clinical medicine [1]. With the rapid development of artificial intelligence technology, human motion posture tracking technology is applied more widely, and three-dimensional (3D) human motion tracking technology appears. Human motion posture tracking in 3D space refers to estimating the 3D coordinates of human joint points and maintaining the tracking of human motion posture. The 3D human motion posture tracking is a hot issue in machine vision. Its subsequent applications include motion recognition, human-computer interaction, and visual understanding. Therefore, it puts forward higher requirements for the accuracy of human motion posture tracking, and it is of great significance to study a new 3D human motion posture tracking.

To break the limitation of spatial dimension and show the tracking results of human motion posture more comprehensively, a 3D human motion posture tracking method is proposed in this paper. From the current research situation, there are many mature human motion posture tracking methods at present. For example, Li et al. [2] proposed a 3D human motion posture tracking method based on dual Kinect sensors. In this method, the human motion posture is expressed by the human joint degree of freedom vector, the human motion posture is tracked by the traceless Kalman filter method, and a human motion posture tracking system based on dual Kinect sensors is built. However, this method has the problem of high tracking loss time, and the practical application effect is not good. Ma
et al. [3] proposed a multifeature fusion human motion posture tracking method in video. In the process of human motion posture tracking, the motion information of the target area of human was calculated by using the video time information, and the motion information is used to transfer the posture model of the human part between frames. This method used the deep learning network to generate additional candidate samples of human arm parts and the probability map of arm feature consistency. Combined with the video spatial information, the optimal position posture is calculated, and each part is reorganized into a complete human motion posture tracking result. However, the position tracking error of this method is high. Xiao et al. [4] presented a 3D human motion posture estimation method based on CNN. This method improves a sequential convolution neural network to extract human spatial information and texture information and finds out the accurate position of human head and limb joint points by estimating the two-dimensional posture of human in video. Through projecting joint points into 3D space, the 3D posture of each person is estimated and tracked. However, the posture angle tracking error of this method is large, and the practical application effect is not good. Huang et al. [5] proposed a human motion posture tracking recognition method based on indoor location technology. Firstly, a wearable receiving tag is installed at the key nodes of the human, and the UWB ranging method is used to locate and track the key parts of the human. In the posture estimation algorithm, the least square algorithm and the improved extended Kalman filter algorithm are used to suppress the noise and improve the tracking accuracy, but this method has the problem of large position tracking error. Tani et al. [6] designed a walking tracking method based on human motion posture estimation in Omnidirectional mobile walker. RGB-D camera is used to measure the abdominal surface of gait students, and the measured point cloud is ellipse fitted on the horizontal plane. Based on the center position and direction of the approximate ellipse, the position coordinates and direction of the gait trainer are estimated, and the posture of the trainees is tracked by proportional control. However, this method has the problem of taking much tracking loss time.

The traditional human motion posture tracking method mainly takes two-dimensional video images as the research aim. If it is applied to the tracking of 3D human motion posture, there will be some problems, such as large tracking error, tracking loss, and low tracking timeliness. Therefore, based on the above existing tracking methods, a new 3D human motion posture tracking method is designed by introducing multilabel transfer learning. The contributions of this paper are as follows: (1) Through multilabel transfer learning, multiple targets to be processed are tagged separately, and high-precision classification results are obtained through the iteration of transfer learning. (2) The tracking error of human motion posture is corrected by three-step search to improve the tracking accuracy. (3) Different datasets are used to verify the application effect of this method. Through experimental verification, this paper can realize fast and accurate tracking of 3D human motion posture.

2. Methodology

The optimized 3D human pose tracking flow is shown in Figure 1. It can be divided into three stages: initialization stage, feature extraction stage, and tracking stage. In the initialization stage, initialize the 3D human skeleton model and training sequence, respectively, in this result. In the feature extraction stage, determine the tracking detection target, obtain the feature information of 3D human motion posture from the video image sequence through the multilabel transfer learning, segment the background and foreground with appropriate methods, and determine the tracking mark points of human motion posture. Output the posture feature extraction results. In the tracking stage, according to the current 3D human motion posture feature extraction results, the human motion posture parameters at the next moment are predicted. Verify whether the sample is reasonable by combining a priori model and a posteriori model. If it is unreasonable, resampling is required. The predicted results are used as the tracking search range, and the final posture tracking results are obtained.

2.1. Dataset. This experiment needs a large amount of 3D human motion posture data as support. Therefore, the public dataset Human3.6M and HumanEva dataset are selected as the training and testing data sources. The public dataset Human3.6M contains 65000 target characters. For each sampling point on the human, it is corresponding to the corresponding surface points by using the method of manual annotation, so as to realize the dense correspondence between the original image and the 3D surface model. In data annotation, six different visual effects are provided for each human joint, so that the appearance of each joint can be seen from any angle, which greatly improves the annotation accuracy of the overall data and provides support for the application of tag transfer learning. The data in HumanEva public dataset is stored in the form of dynamic image or video, and the amount of stored data is about 35.69 GB. The synchronization of 3D human motion posture is realized by using seven calibrated video sequences and action capture system. The public dataset includes three parts: training, verification, and test set. In this experiment, the data with supporting code is selected as the research data. According to the application requirements of the optimization design method, the selected experimental data are transformed into a form that can be directly identified and processed by the 3D human motion posture tracking method. Figure 2 represents the preparation results for a portion of the experimental dataset.

From this, we can get the data preparation results of other human motion posture types. The total number of data samples set in the experiment is 3500. The 3D human motion posture types involved include lunge jump, jumping jack, walking, running, squatting, and waving. In order to avoid the influence of accidental events on the experimental results, the experimental data prepared above are divided into seven groups on average, and the posture types contained in each group are different.
The optimized 3D human motion posture tracking method is realized by using the multilabel migration learning framework. During data sample training, four NVIDIA GeForce GTX 63 graphics cards are used unless otherwise specified. In the training stage, the size of each batch of multilabel transfer learning is 1024. The initial
weight of each layer and the weighted attenuation parameters of 0.01 and 0.005 are obtained through Gaussian distribution. The whole dataset is iterated for 40 times. After each iteration, the learning rate decreases by 0.90. Using the regularization probability of 0.20, some characteristic parameters are ignored, overcoming the problems of over-fitting and time consumption of artificial neural network. In order to optimize the generator and discriminator of the network, the generation and discrimination will be updated in each iteration. When the accuracy of the classifier exceeds 0.95, it will be updated. If the accuracy of the classifier is lower than 0.1, it will not be updated. After completing the experimental environment and parameter configuration, the prepared experimental data are substituted into the 3D human motion posture tracking method of optimal design piece by piece to obtain the tracking results.

2.2. Construction of 3D Human Joint Skeleton Model. According to the human motion structure and the relationship between joints, the human joint skeleton model is constructed, as shown in Figure 3.

In order to find a balance between the computational complexity and the expression accuracy of the model and meet the actual needs of various human bone models, when parameterizing, we should avoid making the parameters in the parameterization greater than the actual degrees of freedom, so as to reduce the possibility of false posture. Therefore, the degree of freedom constraints is set for each joint in the human joint skeleton model as shown in Table 1.

The degree of freedom constraints in Table 1 are added to the 3D human joint skeleton model in the form of direction vector. Taking the shoulder joint as an example, the quantitative constraints can be expressed as

\[ w = [w_x, w_y, w_z]. \]

To make the generated 3D human motion posture meet the constraints of human motion posture in 3D space, the angles between direction vectors need to meet the following relationship:

\[
\begin{align*}
\sin \beta &= \frac{w_y w_z - w_z w_y}{\|w_y\| \|w_z\|} \geq 0, \\
\sin \delta &= \frac{w_y w_z - w_z w_y}{\|w_y\| \|w_z\|} \leq 1,
\end{align*}
\]

where \( \beta \) and \( \delta \) refer to the posture angles of human shoulder joint in \( XZ \) dimension and \( YZ \) dimension, respectively. Similarly, the 3D degree of freedom constraints of other human joints can be obtained to complete the construction of 3D human joint skeleton.

2.3. Generation and Preprocessing of 3D Human Motion Posture Image. The acquisition of 3D human motion posture information mainly comes from digital image acquisition. Kinect is used to obtain the depth image of 3D human. It first uses the infrared transmitter on the left to radiate the infrared to the surrounding environment. In any space, different light spots will be generated, resulting in a 3D effect of “light coding.” The depth image data in the field of view is obtained by using the initial parameters of infrared image and Kinect. On this basis, the depth of each part of the image is analyzed from the front, side, and top view, and the

![Figure 3: 3D human joint skeleton model.](image)

| Joint name | Rotation axis | Range of degrees of freedom |
|------------|---------------|-----------------------------|
| Left shoulder | X | \([-\pi, 5\pi/6]\) |
| Right shoulder | X | \([-7\pi/10, -\pi/4]\) |
| Left elbow | Z | \([-\pi/6, \pi]\) |
| Right elbow | Z | \([-\pi/4, 7\pi/10]\) |
| Left hip | Z | \([-3\pi/4, 3\pi/4]\) |
| Right hip | Z | \([-3\pi/4, \pi/4]\) |
| Left knee | Z | \([-\pi/2, \pi/3]\) |
| Right knee | Z | \([-\pi/3, 2\pi/3]\) |
visual output results of 3D human motion posture data are obtained. When generating 3D human motion posture images, there are often a series of data acquisition, transmission, storage, and recording. In these processes, there will be noise, resulting in the pollution of digital images. In order to reduce the negative impact of noise on image quality, Gaussian filter is used to denoise the initial 3D image before target posture detection and tracking [7]. The expression of the function of Gaussian filter is as follows:

\[
g(x, y, z) = e^{-\frac{(x^2 + y^2 + z^2)}{2\sigma^2}},
\]

where \(\sigma\) refers to the filter bandwidth and \(x, y,\) and \(z\) are the pixel value of the human motion posture image to be processed at three dimensions human motion posture [8]. In order to ensure that the initial 3D human motion posture image can effectively retain the edge information after Gaussian filtering, the filter bandwidth is set to 0.8.

2.4. Detection of 3D Human Motion Objects. The detection of 3D human motion target is mainly realized in two steps. The background image is differentially processed by using the parameter value of the background image, so as to obtain the target motion area. On this basis, threshold segmentation is used to realize the segmentation between 3D human motion foreground and background [9], as shown in Figure 4. \(I_f\) and \(B_f\) are the initial image of 3D human motion posture and its background image, and \(F_f\) denotes the extraction results of human motion targets. It mainly uses the background difference to segment the initial image into target image and background image and extracts the moving target by constantly updating the background image, so as to improve the target extraction accuracy. The specific background differential processing process can be expressed as

\[
F(a, b) = 1 - 2\sqrt{(a + 1)(b + 1) - (a + 1 + b + 1)} \cdot \frac{2\sqrt{(255 - a)(255 - b)}}{(255 - a + 255 - b)},
\]

where \(a(x, y, z)\) and \(b(x, y, z)\) are the grey value of the current frame and the background image at \((x, y, z)\). After the background model is obtained, the pixels higher than the threshold in the difference result are taken as the foreground, and vice versa [10]. In addition, because the background of human motion posture of motion will change with the influence of space object movement, illumination, target shadow, and other elements, the background image model needs to be updated in real time in the process of background difference. The update process can be expressed as follows:

\[
B_k(x, y, z) = yB_{k-1}(x, y, z) + (1 - y)I_k(x, y, z),
\]

where \(y\) refers to the background updated coefficient [11]. The segmentation threshold between human motion foreground and environmental background is obtained through background difference, and the image segmentation is realized by

\[
G(a, b) = \begin{cases} 
0, & F(a, b) < T, \\
1, & F(a, b) \geq T.
\end{cases}
\]

The background difference results of (4) are substituted into (6) to derive the detection results of 3D human motion targets.

2.5. Extracting Human Motion Posture Features. In order to ensure the integrity of human motion posture feature extraction, each joint point is marked according to the human skeleton model, and then the multilabel transfer learning is used to obtain the feature extraction results of human motion posture from the aspects of edge, feature point, shape, and so on [12]. Figure 5 represents the feature extraction process of the multilabel transfer learning.

According to Figure 5, the feature extraction process of multilabel transfer learning is as follows: the training samples are divided into source domain samples and target domain samples. The two groups are projected into the same feature subspace through the mapping matrix to generate a multilabel classifier. The multilabel classifier is used to label the target domain samples in the shared feature subspace. Combined with the marking results, the features are extracted from the position of each joint point and the contour of human motion posture.

According to the process in Figure 4, the labeled human joint source domain sample set and the unlabeled target domain sample set are recorded as \(D^{(0)}\) and \(D^{(f)}\). The labeled matrix of the training sample is regarded as a hypergraph, and its structure is expressed by the Laplace matrix of the hypergraph, which is mapped into the eigenvector of the matrix.

\[
\tau = \min_{W \in R^{m \times n}, \forall (a,b) \in H} \left\{ W^T X^2 + \| S - HH^T \|^2 \right\},
\]

where \(W\) refers to the mapping matrix of subspace of sample features and \(X, H, S\) are the sample characteristic matrix of the source domain, the sample marking matrix of the target domain, and the sample potential characteristic matrix of the target domain [13]. Then, the training samples of the source domain and the target domain are projected into the same feature subspace through the mapping matrix \(W\), and the feature extraction results are obtained. In the actual 3D human motion posture feature extraction process, according to the operation process of the above multilabel transfer learning, feature extraction is carried out from the position of each joint point and the contour of human motion posture. Taking contour feature extraction as an example, all surface points are set in the contour, and the contour line is obtained by Gaussian mixture method, and the contour line of the model along the projection direction can be easily obtained by this method [14]. The extraction process of 3D human motion posture profile features can be expressed as
where \( x_{i}^{\text{img}} \) is the projection point of a point on the surface and \( I_{\text{dist}} \) is the distance between any pixel and the contour. The objective function of contour features makes the optimal solution meet the model projection in the contour to the greatest extent [15]. Similarly, we can get the extraction results of other features and finally output the final feature extraction results through the fusion of multiple types of 3D human motion posture features.

2.6. Estimated 3D Human Motion Posture. According to the extraction results of human motion posture features, the two-dimensional coordinate value \((u_t, v_t)\) of each joint point in the human joint skeleton model at any time can be obtained. According to the coordinate sequence of the two-dimensional joint points, the 3D coordinate value \((x_t, y_t, z_t)\) of each joint point is obtained by using the proportional orthogonal projection method. The specific solution formula is as follows:

\[
\begin{align*}
{x_t} &= \frac{u_t}{s}, \\
{y_t} &= \frac{v_t}{s}, \\
{z_t} &= l_{\text{limbs}} \cdot \sin \theta,
\end{align*}
\]

where \(s\) is the proportional factor and \(l_{\text{limbs}}\) and \(\theta\) are limb length and posture angle, respectively. Then, in the estimation of 3D human motion posture, it is only necessary to predict the two-dimensional position coordinates and...
posture angle of each joint point [16], in which the estimate value of the posture angle is

\[ \psi_t = (t - t_0) \phi \theta, \] (10)

where \( t - t_0 \) refers to the time interval of human motion posture motion and \( \phi \theta \) is posture angle change feature extracted by multilabel transfer learning [17]. Similarly, the two-dimensional position coordinate estimation results of joint points can be obtained, and the 3D coordinate estimation results of human joints can be obtained by substituting them into equation (9).

### 2.7 3D Human Motion Posture Real-Time Tracking Method.

**Input:** 3D human motion posture training samples

**Output:** 3D human motion posture real-time tracking results

In order to meet the speed and accuracy requirements of human motion posture tracking, a three-step search method is adopted to search 8 pixels around the 3D human motion posture estimation coordinates in turn, and the purpose of effective tracking estimation is achieved through error compensation [18]. The three-step search is based on the block matching criterion, and the joint tracking position is corrected by

\[ \text{MAD}(d_x, d_y, d_z) = \frac{1}{MN} \sum |f_k(x_i + d_x, y_i + d_y, z_i + d_z)|, \] (11)

where \((d_x, d_y, d_z)\) is the motion vector of tracking target in three dimensions and \(f_k\) is the grey value of the obtained and processed image. If a certain point is at \((x, y, z)\), the calculation result of (11) reaches the minimum, and this point is the optimal matching point to be found, that is, the 3D posture tracking point of human joints [19,20]. Finally, all the joint posture parameters are substituted into the 3D human skeleton model, and the visual output results of real-time posture tracking are obtained.

By constructing the 3D human joint skeleton model, the 3D human motion posture image is generated and denoised, and the 3D human moving target is detected by background difference. The multilabel transfer learning is used to extract the features of human motion posture, and the estimation results of 3D human motion posture are obtained. The tracking error of human motion posture is corrected through three-step search, and the visual 3D human motion posture tracking results are output. The process of the 3D human motion posture tracking method using multilabel transfer learning is shown in Figure 6.

### 2.8 Experimental Indicators.

The method of literature [2] (Method1), the method of literature [3] (Method2), the method of literature [4] (Method3), the method of literature [5] (Method4), the method of literature [6] (Method5), and the proposed method were used as experimental methods to verify the effectiveness of the application of different methods by comparing different experimental indicators.

Position tracking error refers to the error in longitude and latitude position in the process of tracking 3D human motion posture. The equation for calculating this indicator is as follows:

\[ \varepsilon_p = \sum_{i=1}^{n} \left| (q_{t-\text{act}} - q_{t-\text{track}}) \right|, q = x, y, z, \] (12)

where \(q_{t-\text{act}}\) and \(q_{t-\text{track}}\) refer to the coordinate of the actual position of joint \(i\) and the position coordinate tracking result. Parameter \(n\) refers to the number of human joints to be tracked.

Posture angle tracking error refers to the error in posture angle in the process of tracking human motion posture.

\[ \varepsilon_{\theta} = |\theta_{t-\text{act}} - \theta_{t-\text{track}}|, \] (13)

where \(\theta_{t-\text{act}}\) and \(\theta_{t-\text{track}}\) are the actual value and the tracking value, respectively, of the posture angle of joint \(i\).

Time of posture tracking loss refers to the sum of all the lost time of target tracking.

\[ \Delta t_{\text{miss}} = T_s - T_e, \] (14)

where \(T_s\) and \(T_e\) are the start time and end time, respectively, of posture tracking loss. Generally, it is required that the number of tracking joint points shall not be less than 8. If it is detected that there are less than 8 joint points in the output.

---

**Figure 6: 3D human motion posture tracking method process.**
interface of tracking results, this time shall be recorded as the start time of posture tracking loss. When the joint points displayed on the output interface are restored to 8, this indicates that the posture tracking loss is over. The smaller the tracking error, the shorter the time of posture tracking loss, which shows that the accuracy and performance of the corresponding tracking method are higher.

The calculation equation of posture tracking update response time index is as follows:

$$\Delta t_{\text{response}} = T_k - T_{k-1},$$

(15)

where $T_k$ and $T_{k-1}$ are the output time of the tracking data of $k$ and $k - 1$, respectively. The shorter the update response time, the better the response performance of the corresponding tracking method.

The numerical results of the fitness index of 3D human motion posture tracking are as follows:

$$\eta_{\text{adaptation}} = \frac{N_{\text{type}}}{N_{\text{all}}} \times 100\%,$$

(16)

where variables $N_{\text{type}}$ and $N_{\text{all}}$ are the number of samples supported for tracking and total number of samples prepared, respectively. The larger the value of $\eta_{\text{adaptation}}$ is, the better the adaptation performance of the corresponding tracking method is.

### 3. Results and Discussion

The actual posture data of human and the output results of tracking method are extracted, and the test results of position tracking error are obtained through the calculation of (12), as shown in Table 2.

According to the data in Table 2, the average position tracking error of method1 is 14.14 mm, the average position tracking error of method2 is 11.30 mm, the average position tracking error of method3 is 11.52 mm, the average position tracking error of method4 is 7.26 mm, and the average position tracking error of method5 is 5.96 mm. The average position tracking error of the proposed method is 2.18 mm, and the tracking error is the lowest among the six experimental methods, which shows that the position tracking error accuracy of this method is high.

The results of the posture angle tracking error tests for the six methods are shown in Table 3.

The data in Table 3 were averaged to derive the average of the posture angle tracking errors for the six tracking methods. The average posture angle tracking error of method1 is 3.438 deg, the average posture angle tracking error of method2 is 2.302 deg, and the average posture angle tracking error of method3 is 2.252 deg, the average posture angle tracking error of method4 is 2.454 deg, and the average posture angle tracking error of method5 is 1.326 deg. The average posture angle tracking error of the proposed method is 0.178 deg, which is much lower than the traditional posture tracking method, indicating that the posture angle tracking error of this method is low and the accuracy is high.

The test results for the length of posture tracking loss are shown in Figure 7.

It can be seen in Figure 7 that the time of posture tracking loss of method1 is up to 190 s, the time of posture tracking loss of method2 is up to 205 s, the time of posture tracking loss of method3 is up to 210 s, the time of posture tracking loss of method4 is up to 210 s, and time of posture tracking loss of method5 is up to 180 s. The time of posture tracking loss in the proposed method is up to 16 s, which can prove that the lost time of posture tracking in this method is shorter and the integrity of tracking results is higher.

In terms of the timeliness performance of the tracking method, the output time of two consecutive tracking results is recorded and substituted into (15), and the test and comparison results of the update response time index of 3D human motion posture tracking are obtained, as shown in Figure 8.

According to Figure 8, the 3D human motion posture tracking update response time of method1 is between 1300 ms and 7500 ms, the 3D human motion posture tracking update response time of method2 is between

| Method   | Frame number | 1   | 2   | 3   | 4   | 5   |
|----------|--------------|-----|-----|-----|-----|-----|
| Method1  | $\varepsilon_p$/mm | 9.9 | 11.7 | 14.4 | 16.2 | 18.5 |
| Method2  | $\varepsilon_p$/mm | 15.4 | 8.7 | 11.9 | 10.4 | 10.1 |
| Method3  | $\varepsilon_p$/mm | 11.4 | 10.7 | 12.5 | 13.2 | 9.8 |
| Method4  | $\varepsilon_p$/mm | 6.2 | 7.7 | 4.9 | 9.2 | 8.3 |
| Method5  | $\varepsilon_p$/mm | 5.2 | 7.6 | 4.3 | 5.9 | 6.8 |
| Proposed | $\varepsilon_p$/mm | 1.6 | 3.3 | 2.8 | 1.5 | 1.7 |

| Method   | $\varepsilon_\theta$/deg | 2.17 | 3.53 | 4.18 | 4.46 | 2.85 |
|----------|--------------------------|-----|-----|-----|-----|-----|
| Method2  | $\varepsilon_\theta$/deg | 0.97 | 1.39 | 3.42 | 3.56 | 2.17 |
| Method3  | $\varepsilon_\theta$/deg | 2.38 | 3.04 | 1.92 | 1.37 | 2.55 |
| Method4  | $\varepsilon_\theta$/deg | 3.08 | 3.21 | 2.66 | 1.73 | 1.59 |
| Method5  | $\varepsilon_\theta$/deg | 1.35 | 1.48 | 1.52 | 1.21 | 1.07 |
| Proposed | $\varepsilon_\theta$/deg | 0.21 | 0.13 | 0.25 | 0.14 | 0.16 |

**Figure 7: Posture tracking lost time test comparison results.**

**Figure 8: 3D human motion posture tracking response time.
1300 ms and 7200 ms, the 3D human motion posture tracking update response time of method3 is between 1300 ms and 6900 ms, the update response time of 3D human motion posture tracking of method4 is between 1300 ms and 5000 ms, and the update response time of 3D human motion posture tracking of method5 is between 1300 ms and 4300 s. The update response time of the 3D human motion posture tracking of the proposed method is between 1300 ms and 1600 ms. Compared with the five traditional methods, the update time of the optimized 3D human motion posture tracking method is shorter, which proves that the optimization method has more advantages in time performance.

After the calculation of (16), the results of the adaptation performance test of the posture tracking method are obtained, as shown in Table 4.

From the analysis of the data in Table 4, it can be concluded that the average fitness indexes of the five comparison methods are 96.77%, 95.39%, 94.79%, 95.34%, and 98.46%, respectively, while the average fitness of the optimized design tracking method is 99.89%, which is higher than all traditional tracking methods; that is, the designed 3D human motion posture tracking method based on multilabel transfer learning has higher fitness performance.

4. Conclusions

In order to improve the performance of 3D human motion posture tracking, a 3D human motion posture tracking method using multilabel transfer learning is proposed. The results show that the position and posture angle tracking error of this method is lower, the time of posture tracking loss and posture tracking update response time are shorter, and the adaptability is higher. From the experimental results, it can be seen that the application of multilabel transfer learning effectively improves the application performance of 3D human posture tracking method and plays a positive role in the development of sports, medical treatment, virtual reality, and other fields. However, due to the limitation of time and space, the number of public datasets used in the

| Experimental group | Method1 | Method2 | Method3 | Method4 | Method5 | Proposed |
|--------------------|--------|--------|--------|--------|--------|----------|
| 1                  | 95.2   | 94.4   | 92.0   | 96.4   | 98.8   | 99.8     |
| 2                  | 98.3   | 97.6   | 95.8   | 98.2   | 97.6   | 99.9     |
| 3                  | 97.5   | 98.5   | 99.4   | 93.3   | 98.5   | 99.7     |
| 4                  | 96.1   | 93.2   | 90.9   | 91.9   | 99.2   | 100      |
| 5                  | 99.8   | 91.5   | 96.9   | 98.4   | 99.0   | 100      |
| 6                  | 93.4   | 97.7   | 93.7   | 95.5   | 98.4   | 99.9     |
| 7                  | 97.1   | 94.8   | 94.8   | 93.7   | 97.7   | 99.9     |

Table 4: Posture tracking method fitness (%).
experiment is small, and the experimental results have some limitations. Therefore, in future work, the application data needs to be further supplemented.

**Data Availability**

The data used to support the findings of this study are included within the article. Readers can access the data supporting the conclusions of the study from Human3.6M dataset and HumanEva dataset.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**References**

[1] S. Jiang, Y. Xu, T. Wang et al., “Multi-label metric transfer learning jointly considering instance space and label space distribution divergence,” *IEEE Access*, vol. 7, pp. 10362–10373, 2019.

[2] Q. Li, X. D. Wang, and H. Li, “3D human posture tracking method based on dual Kinect sensors,” *Journal of System Simulation*, vol. 32, no. 8, pp. 1446–1454, 2020.

[3] M. Ma, Y. B. Li, and X. Q. Wu, “Human posture tracking based on multi feature fusion in video,” *Chinese Journal of Image and Graphics*, vol. 25, no. 7, pp. 1459–1472, 2020.

[4] A. W. Xiao, J. Liu, and S. P. Zhang, “3D human pose estimation method based on CNN,” *Journal of Wuhan University of Engineering*, vol. 41, no. 2, pp. 5–15, 2019.

[5] X. P. Huang, J. Zhang, and Z. L. Hu, “Human posture recognition method based on indoor positioning technology,” *Journal of China University of Science and Technology*, vol. 49, no. 10, pp. 805–811, 2019.

[6] Y. Tani, R. Tasaki, and K. Terashima, “Walking tracking system based on estimation of human posture in omni-directional mobile walker,” *Journal of the Robotics Society of Japan*, vol. 37, no. 2, pp. 161–167, 2019.

[7] D. Groos, L. Adde, and E. Ihlen, “Approaching human precision on automatic markerless tracking of human movements,” *Gait & Posture*, vol. 81, no. 11, pp. 117-118, 2020.

[8] M. Airaksinen, O. Räsänen, E. Ilén et al., “Automatic posture and movement tracking of infants with wearable movement sensors,” *Scientific Reports*, vol. 10, no. 1, p. 169, 2020.

[9] A. Sengupta, F. Jin, R. Zhang, and S. Cao, “Mm-pose: real-time human skeletal posture estimation using mmWave radars and CNNs,” *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10032–10044, 2020.

[10] Q. Fu, X. Zhang, J. Xu, and H. Zhang, “Capture of 3D human motion pose in virtual reality based on video recognition,” *Complexity*, no. 3, pp. 1–17, 2020.

[11] K. Ishii, Y. Ohno, M. Oikawa, and N. Onishi, “Relationship between human exhalation diffusion and posture in face-to-face scenario with utterance,” *Physics of Fluids*, vol. 33, no. 2, p. 027101, 2021.

[12] G. Li, Z. Liu, L. Cai, and J. Yan, “Standing-posture recognition in human–robot collaboration based on deep learning and the dempster–shafer evidence theory,” *Sensors*, vol. 20, no. 4, p. 1158, 2020.

[13] E. M. Taylor and M. E. Cinelli, “The effects of human following behaviours on decision making during aperture crossing scenarios,” *Gait & Posture*, vol. 83, pp. 232–236, 2021.