Research Article

Research on Three-dimensional Motion History Image Model and Extreme Learning Machine for Human Body Movement Trajectory Recognition

Zheng Chang, Xiaojuan Ban, Qing Shen, and Jing Guo

School of Computer and Communication Engineering, University of Science and Technology Beijing, Haidian District, Beijing 100083, China

Correspondence should be addressed to Xiaojuan Ban; banxj@ustb.edu.cn

Received 15 August 2014; Revised 3 November 2014; Accepted 5 November 2014

Academic Editor: Zhan-li Sun

Copyright © 2015 Zheng Chang 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.

Based on the traditional machine vision recognition technology and traditional artificial neural networks about body movement trajectory, this paper finds out the shortcomings of the traditional recognition technology. By combining the invariant moments of the three-dimensional motion history image (computed as the eigenvector of body movements) and the extreme learning machine (constructed as the classification artificial neural network of body movements), the paper applies the method to the machine vision of the body movement trajectory. In detail, the paper gives a detailed introduction about the algorithm and realization scheme of the body movement trajectory recognition based on the three-dimensional motion history image and the extreme learning machine. Finally, by comparing with the results of the recognition experiments, it attempts to verify that the method of body movement trajectory recognition technology based on the three-dimensional motion history image and extreme learning machine has a more accurate recognition rate and better robustness.

1. Introduction

With the rapid development of the natural human-computer interaction technology, the body movement trajectory tracking and recognition technology has become an important and indispensable research direction in the natural human-computer interaction technology. As we all know, since the body movement is a natural and intuitive communication mode [1], therefore, beyond all doubts, the body movement recognition technology has become a useful technology to the new generation of natural human-computer interaction interface [2–4], especially for the disabled and patients who can only use their body movements to give orders to the auxiliary equipment (such as the wheelchair, the smart television, and disabled scooter), which will bring them more convenience.

The prior body movement trajectory recognition researches on the human-computer interaction mainly focus on the modelling of human skin colour and the extraction of dynamic body movements based on image attributes

of the robust feature [5] and the artificial neural network; however, due to the diversity, ambiguity, and disparity in time and space of the body movements, the traditional body movement trajectory recognition researches have great limitations. The paper attempts to introduce the invariant moments of the three-dimensional motion history image and the extreme learning machine into the body movement trajectory recognition, which will make the machine vision recognition of the body movement trajectory more accurate, efficient, and robust [6].

The paper is organized as follows. Section 1 describes the background of the human body movement trajectory recognition and the importance of the human body movement trajectory recognition. Section 2 describes the main problems in the human body movement trajectory recognition, followed by a summary of the normal algorithm. Section 3 describes the principal ideas of our new method and the main steps including the feature extraction of the motion history image, calculation of the invariant moments, and the movement recognition based on extreme learning machine. Section 4

FIGURE 1: Two-dimensional body movement image.

makes a comparison of the recognition performance between our new method and the other algorithms. Section 5 summarizes the conclusions of this study.

2. Problem Description

The essence of the body movement trajectory recognition is to extract motion features and to classify the sample data accurately. It is a contrast between the body movement trajectory captured by the sensor and the predefined sample movement trajectory. Therefore, there are two steps in the body movement trajectory recognition. The first step is the features extraction of dynamic body movements. And the second step is the classification of the body movement trajectory captured by the sensor.

In the first extraction step, the traditional method extracts the dynamic body movements by applying the hidden Markov model [7, 8] as is shown in Figure 1.

In the hidden Markov model method, human movement trajectory data $\mathcal{M}$ can be regarded as the state series $\mathcal{M} = (\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \dots, \mathcal{F}_T)$ which is obtained from sampling frames. In one single frame, the state of the human movement is represented by $\mathcal{F} = (r_1, r_2, r_3, \dots, r_n)$, wherein r_n represents the value of the n th characteristic values in the current frame. The movement trajectory recognition is the process of comparing the real-time trajectory captured by sensor with the state series $\mathcal{M}$ and then matching $\mathcal{M}$ with the state series of the predefined sample trajectory.

Based on the hidden Markov model (HMM), the whole process of the comparison between the real trajectory from the sensor and the predefined sample is shown in Figure 2 [9]. But based on the hidden Markov model (HMM), the recognition process still has some limitations; for example, consider the following.

- (i) Light: when the light condition is changed, the luminance information of the body will change since the images captured by the sensor are easily affected by natural light and artificial light. Different skin colors can also easily affect the luminance information under the same light condition.
- (ii) Obstruction: during the whole process, the body movement trajectory may be blocked by some objects in the environment or the other parts of the body since the obstruction can lead to the loss of the identification information of body, which will affect the reliability of the body movement recognition greatly.
- (iii) Background: in the real-time body movement recognition process, if the factors (color, texture, shape,

etc.) of the body movement area and the background area are similar, it will also affect the performance of the recognition [10].

The three-dimensional hidden Markov model (3DHMM) is a better performance, but this method has been restricted to very few application fields because of the huge amount of calculation, the inefficiency of the training, and the easy accessibility to the local optimal value, and so forth.

In the second classification step, the traditional methods involve many machine learning algorithms (such as K nearest neighbor method and gradient descent-based feed-forward network learning methods) [11–13]. But these methods still have some limitations as follows.

- (i) K nearest neighbor method: the K nearest neighbor method (KNN) needs to compute the distance (such as Euclidean distance, Mahalanobis distance, or Pearson correlation) between the real body movement trajectory captured by the sensor and every predefined sample body movement trajectory. Therefore, because of the heavy computation in this method, it has been restricted to very few application fields.
- (ii) Gradient descent-based feed-forward network learning methods: the BP artificial network is one of the typical gradient descent-based feed-forward network learning methods. All the parameters of the feed-forward networks need to be turned and thus there exists the dependency between different layers of parameters (weight and biases) in the BP artificial network. Therefore, the training process is very slow. And this gradient descent-based learning method may easily converge to local minima and overfitting problem [14–16].

To solve these problems, this paper attempts to combine the three-dimensional motion history image and the extreme learning machine in order to overcome those shortcomings.

In the comparative experiment section (Section 4.3), a detailed and specific experiment about the comparison of the different classification methods will be conducted.

The three-dimensional motion history image can easily present the feature of the human body movement trajectory, such as the space feature and the time feature. And the seven invariant moments of the three-dimensional motion history image have translation invariance, scaling invariance, and rotation invariance. These invariances could tactfully overcome the observation position sensitivity of the human body movement trajectory.

The comparison between the motion history image and the traditional MHH method will be described in the comparative experiment at the end of Section 4.2.

The extreme learning machine has a faster and better generalization performance than another traditional feed-forward artificial network (such as BP network) [17–19]. What is more, the extreme learning machine could obtain the global optimal value easily. Therefore, it adopts a new method to achieve the body movement trajectory recognition goal in this paper (Figure 3).

FIGURE 2: The HMM based on two-dimensional image.

FIGURE 3: The ELM based on 3D motion history image.

In the first step, this paper attempts to combine the motion history image (MHI) with the three-dimensional depth data of the body movements in order to get the three-dimensional motion history image (3DMHI) of body movements (Figure 4) because a single human body gesture cannot show the meaning of the human action. Using this method we could get the human body movement trajectory presentation from a few of previous video frames or snippets. The snippets which we set in our experiments usually take three minutes or more. In order to improve the accuracy of the human body movement trajectory recognition process, it adopts a high sampling rate, thirty frames per second in the paper.

At the end of Section 3.1, the grey scale of the motion history image (MHI) will be given a detailed and popular description.

FIGURE 4: The MHI based on three-dimensional depth data.

And then it calculates seven invariant moments of the three-dimensional motion history image working as the eigenvector of the body movements. To calculate the invariant moments, after getting the three-dimensional motion history image, the image was projected in the XY plane, YZ plane, and XZ plane. In each projection plane, a set of invariant moments is calculated. Each three sets of invariant moments work as the eigenvector of the human body movement trajectory.

In the second step, this paper attempts to use the extreme learning machine instead of the traditional methods (like KNN, BP, or SVM). On the one hand, not only is the new process free from the effects of the illumination, obstruction,

background, and other environmental factors but it also improves the efficiency, accuracy, and robustness of body movement recognition. The extreme learning machine has a faster and better generalization performance than another traditional feed-forward artificial network (such as BP network) [20–22]. According to the previous splendid work, the extreme learning machine tends to have a better scalability and achieve a similar (for regression and binary class cases) and much better generalization performance (for multiclass cases) at a much faster learning speed (up to thousands of times) than traditional machine learning methods (such as KNN, BP, and SVM) [23–25].

3. Problem Solving

3.1. Body Movements Characterized by 3D Motion History Image. To characterize the 3D motion information, the paper proposes a new method named three-dimensional motion history images approach. This method improves the application of the traditional motion history images approach based on two-dimensional image in order to combine with the three-dimensional depth data. By combining the invariant moments of the three-dimensional motion history image (computed as the eigenvector of body movements) and the extreme learning machine (constructed as the classification artificial neural network of body movements), the paper applies the method to the machine vision of the body movement trajectory.

The motion history image approach is a kind of special finite-difference time-domain method; it is a branch of the Finite Difference Time Domain (FDTD) method [26, 27]. The mechanism of the FDTD method is to get different images from continuous image sequences by comparing with two or three adjacent pixels in the corresponding frames and then to extract the human body moving regions in the image by setting the threshold. By introducing the 3D data, the paper presents the improved FDTD method named three-dimensional motion history image approach as follows:

$$\mathcal{D}(x, y, z, n) = \mathcal{F}(x, y, z, n-1) - 2\mathcal{F}(x, y, z, n) + \mathcal{F}(x, y, z, n+1). \quad (1)$$

Among them, N means the number of the current frame and N mainly shows the relation between the adjacent frames. $\mathcal{F}(x, y, z, n)$ represents the pixel grey value in the position (x, y, z) in three-dimensional space; $\mathcal{D}(x, y, z, n)$ is the result of the three consecutive frames' difference and also represents body movement changed area. The threshold $\mathcal{D}(x, y, z, n)$ is as follows:

$$\mathcal{B}(x, y, z, n) = \begin{cases} 1 & \mathcal{D}(x, y, z, n) > \Gamma \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where Γ is the specially selected threshold. If the value is too low, it cannot effectively remove noises in images, but if the value is too high, it will impede the valuable variation of the image. So the value of the threshold should be adjustable for different experiment conditions. The experiment should be

FIGURE 5: Three-dimensional motion history image of body movements (MHI).

repeated many times to determine the value of the threshold. Under my experiment condition, the value of the threshold is $(10, 10, 15, n)$ in my follow-up work.

Three-dimensional motion history image approach of body movements is as follows:

$$\mathcal{H}_\tau(x, y, z, t) = \begin{cases} \tau & \mathcal{B}(x, y, z, t) = 1 \\ \max(0, \mathcal{H}_\tau(x, y, z, t-1) - 1) & \text{otherwise.} \end{cases} \quad (3)$$

Among them, $\mathcal{H}_\tau(x, y, z, t)$ represents the pixel gray value in the position (x, y, z) and t in three-dimensional motion history image. The motion history image MHI not only reflects the external shape of the body movements (space feature) but also reflects the direction and state of them (time feature).

Generally speaking, a sequence of the human body movement trajectory image is compressed into a single and special image by the previous algorithm. The single and special image works as the motion history image (MHI).

In the motion history image, the grey value of each pixel is in proportion with the duration of the body movement in the position. The recent body gestures have the maximum grey value. Grey value changes reflect the direction of the body movements (Figure 5). With time elapsing, the grey value of the human body movement trajectory would decrease.

In other words, a sequence of the human body movement trajectory image is compressed into a single and special image by the previous algorithm. But each image's gray scale is changed. The longer the motion lasts in the sequence of the human body movement trajectory image, the lighter the gray scale presents.

3.2. The Calculation of the Invariant Moments of the Motion History Image. Although the three-dimensional motion history image approach based on the MHI is simple and efficient,

FIGURE 6: XY surface projection of the MHI.

FIGURE 8: XZ surface projection of the MHI.

FIGURE 7: YZ surface projection of the MHI.

it is too sensitive to the observation position. Seen from different observation positions, the three-dimensional motion history image will have different results. These differences will affect the accuracy of human body movement recognition process greatly. In order to overcome this shortcoming, this paper selects the invariant moments as eigenvector of the motion history image. The method of invariant moments is a classical method to extract image feature. Its translation invariance, scaling invariance, and rotation invariance properties rule out the impact on the position, distance, and angle.

To calculate the invariant moments, after getting the three-dimensional motion history image, our method attempts to project it in the XY plane (Figure 6), YZ plane (Figure 7), and XZ plane (Figure 8). This projection could simplify the invariant moment's calculation process. The

direct calculation of the three-dimensional motion history image could bring huge amounts of three-dimensional invariant moment's calculation instead of a much better performance of the human body movement trajectory recognition process. This method can get three views of three-dimensional motion history image with one motion. Then the calculation of invariant moments for the three main views is obtained.

For a size of $\mathcal{M} \times \mathcal{N}$ digital image $f(x, y)$, the $p + q$ order moment m_{pq} is defined as follows:

$$m_{pq} = \sum_{x=1}^N \sum_{y=1}^M f(x, y) x^p y^q. \quad (4)$$

Among them, $p, q = 0, 1, 2, \dots$

$p + q$ order central moment μ_{pq} is defined as follows:

$$\mu_{pq} = \sum_{x=1}^N \sum_{y=1}^M f(x, y) (x - \bar{x})^p (y - \bar{y})^q, \quad (5)$$

where (x, y) represents the object image point and $(\bar{x}, \bar{y})$ is the object centroid:

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}}. \quad (6)$$

Among them,

$$\begin{aligned} m_{00} &= \sum_{x=1}^N \sum_{y=1}^M f(x, y), \\ m_{10} &= \sum_{x=1}^N \sum_{y=1}^M f(x, y) x^1, \\ m_{01} &= \sum_{x=1}^N \sum_{y=1}^M f(x, y) y^1. \end{aligned} \quad (7)$$

Then through the normalizing of the central moment by the zero-order central moments μ_{00} , the normalized center moment of the motion history image could be got:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^r}, \quad r = \frac{p+q+2}{2}, \quad p+q = 2, 3, 4, \dots \quad (8)$$

Ming-kuei Hu got seven invariant moments based on the linear combination of two-order and three-order normalized central moment. And the seven invariant moments of the three-dimensional motion history image have translation invariance, scaling invariance, and rotation invariance. These invariances could skillfully overcome the observation position sensitivity of the human body movement trajectory. The image translation, rotation, and scaling are unchanged and the invariant moments are as follows [28]:

$$\begin{aligned} M_1 &= \eta_{20} + \eta_{02}, \\ M_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \\ M_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2, \\ M_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2, \\ M_5 &= (\eta_{30} - 3\eta_{12}) \times (\eta_{30} + \eta_{12}) \\ &\quad \times [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad + (\eta_{03} - 3\eta_{21})(\eta_{03} + \eta_{21}) \\ &\quad \times [(\eta_{03} + \eta_{21})^2 - 3(\eta_{12} + \eta_{30})^2], \\ M_6 &= (\eta_{20} - \eta_{02}) [(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ &\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}), \\ M_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{21}) \\ &\quad \times [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad - (3\eta_{12} - \eta_{30})(\eta_{03} + \eta_{21}) \\ &\quad \times [(\eta_{03} + \eta_{21})^2 - 3(\eta_{12} + \eta_{30})^2]. \end{aligned} \quad (9)$$

Because the values of the invariant moments are too small, it is compressed by the absolute value of the logarithm and so the actual values need to be amended in accordance with the following formula:

$$M_k = \log |M_k|, \quad k = 1, 2, 3, 4, 5, 6, 7. \quad (10)$$

Because formula (10) does not change the feature of the invariant moments, the seven new values of the invariant moments of the three-dimensional motion history image have translation invariance, scaling invariance, and rotation invariance. The invariant moments still have translation, rotation, and scaling invariance after amendment.

Through the calculation of the projection images in three directions, we will get a 3×7 eigenvalue matrix. This eigenvalue matrix is the eigenvector for motion history image

because the three-dimensional motion history image could easily present the human body movement trajectory including the space feature and the time feature. And the seven invariant moments of the three-dimensional motion history image have translation invariance, scaling invariance, and rotation invariance. These 3×7 eigenvalue matrixes could tactfully overcome the observation position sensitivity of the human body movement trajectory and present the human body movement trajectory easily.

3.3. The Body Movements Recognition Based on Extreme Learning Machine. In the process of recognition, first the samples of body movement are collected and then a training sample set was built to obtain a better recognition performance. In order to get a better human body movement trajectory recognition performance, the samples of the human body movement trajectory must be definite, clear, and slow; in other words, all of the training samples must comply with the criterion under my experiment conditions.

For the same body movement, different people involved should repeat the action several times. And then multiple groups of three-dimensional motion history images are collected for each human body movement trajectory; after that, the training sample set for each human body movement trajectory will be established.

According to many precedent outstanding researches, we can draw the conclusion that the input weights and hidden layer biases of a single-hidden layer feed-forward neural network (SLFN) can be randomly assigned if the activation functions in the hidden layer are infinitely differentiable [29].

Based on the conclusion, this paper proposes the extreme learning machine (ELM) for SLFN. The extreme learning machine need not adjust the most parameters of the artificial neural network repeatedly, such as the input weights and the hidden layer biases. Contrary to the traditional artificial neural network, this randomly assigned approach could reduce huge amounts of the calculation and increase the artificial neural network's training efficiency. What is more, the precedent researcher has proved that this simplified approach does not reduce the test accuracy of the artificial neural network. Because of these advantages, the ELM could transform the complex machine learning problems into a simple linear problem which could be determined through a generalized inverse calculation of the hidden layer weight matrices.

In some applications, the extreme learning machine tends to have a better scalability and achieve a similar (for regression and binary class cases) and much better generalization performance (for multiclass cases) at a much faster learning speed (up to thousands of times) than traditional machine learning methods (such as KNN, BP, and SVM).

Because the ELM can have much better efficiency and accuracy; moreover, it can overcome the traditional methods' shortcomings. For our body movement trajectory recognition process, the paper attempts to construct a SLFN (Figure 9) as follows.

There are $\mathcal{N}$ hidden nodes and $\mathcal{M}$ input lay nodes in our SLFN. Our training sample set is (X_j, T_j) , where

FIGURE 9: ELM artificial neural network.

$X_j = [X_{j1}, X_{j2}, \dots, X_{jM}]$ is the invariant moments of training samples and $T_j = [T_{j1}, T_{j2}, \dots, T_{jM}]^T$ is the kind of training samples. There are K output layer nodes; it means that there are K types of human body movement trajectory. Each node in the output layer presents a type of human body movement trajectory. And activation function $\ell(x)$ is mathematically modelled as

$$\sum_{i=1}^N \beta_i \ell(W_i \cdot X_j + b_i) = R_j, \quad (11)$$

where $R_j = [R_{j1}, R_{j2}, \dots, R_{jM}]$ is the output vector from our SLFN, $W_i = [W_{i1}, W_{i2}, \dots, W_{iM}]^T$ is the weight vector which connects the i th hidden layer node and the input nodes, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{iN}]^T$ is the weight vector which connects the i th hidden layer node and the output nodes, and b_i is the threshold of the i th hidden layer node.

Each element in the two vectors $T_j = [T_{j1}, T_{j2}, \dots, T_{jM}]^T$ and $R_j = [R_{j1}, R_{j2}, \dots, R_{jM}]$ is a K dimensional vector. In other words, T_{ji} and R_{ji} serve as the real type and computational type of the invariant moments of training sample X_{ji} , respectively.

Our goal is to make T_j close to R_j , which is equivalent to minimizing the cost function:

$$C = \sum_{i=1}^N \beta_i \ell(W_i \cdot X_j + b_i) - T_j = R_j - T_j. \quad (12)$$

Previous excellent papers have strict calculation that SLFN with N hidden nodes can distinguish N samples exactly for any infinitely differentiable activation function and SLFN may require less than N hidden nodes if learning error is allowed [30]. Of course, the number of the hidden layer nodes in the artificial neural network could have some impacts on the human body movement trajectory recognition performance. In the experiment section, this paper will show these impacts under different numbers of the hidden layer nodes.

The equation is presented as follows:

$$\sum_{i=1}^N \beta_i \ell(W_i \cdot X_j + b_i) = T_j. \quad (13)$$

The above equations could be abbreviated as

$$BF = T, \quad (14)$$

where

$$F = \begin{bmatrix} \ell(W_1 \cdot X_1 + b_1) & \cdots & \ell(W_1 \cdot X_M + b_1) \\ \vdots & \cdots & \vdots \\ \ell(W_N \cdot X_1 + b_N) & \cdots & \ell(W_N \cdot X_M + b_N) \end{bmatrix}_{N \times M},$$

$$B = [\beta_1, \beta_2, \dots, \beta_N],$$

$$T = [T_1, T_2, \dots, T_M]. \quad (15)$$

On both sides of the equation, the transpose operation could be applied in order to obtain the standard ELM model:

$$\mathcal{H}\beta = \mathcal{T}, \quad (16)$$

where

$$\mathcal{H} = F^T,$$

$$\beta = B^T, \quad (17)$$

$$\mathcal{T} = T^T.$$

So we could obtain the result like this:

$$\beta = \mathcal{H}^\dagger \mathcal{T}, \quad (18)$$

where $\mathcal{H}^\dagger$ is the Moore-Penrose generalized inverse of matrix $\mathcal{H}$.

If $N = M$ and $\mathcal{H}$ has an inverse, the solution is existing and unique. And the answer is, evidently, $\mathcal{H}^\dagger = \mathcal{H}^{-1}$. The result is shown as follows:

$$\beta = \mathcal{H}^{-1} \mathcal{T}. \quad (19)$$

There are several methods which could be used to calculate the Moore-Penrose generalized inverse of matrix $\mathcal{H}$. This paper attempts to apply the spectral theorem and Tikhonov's regularization method to calculate $\mathcal{H}^\dagger$ Moore-Penrose generalized inverse of matrix $\mathcal{H}$ [31]:

$$\mathcal{H}^\dagger = \sum_{\substack{b=1 \\ \alpha_b \neq 0}}^s \frac{1}{\alpha_b} \left[\prod_{\substack{l=1 \\ l \neq b}}^s (\alpha_b - \alpha_l)^{-1} \right]$$

$$\times \left[\prod_{\substack{l=1 \\ l \neq b}}^s (\mathcal{H}^* \mathcal{H} - \alpha_l E_n) \right] \mathcal{H}^*, \quad (20)$$

where $\mathcal{H}^*$ represents the adjoint matrix of $\mathcal{H}$ and $\alpha_b, b = 1, 2, \dots, s$, are the eigenvalues of $\mathcal{H}^* \mathcal{H}$ (or the singular values of $\mathcal{H}$).

According to the training sample set, we could obtain the weight vector β and then complete the training of ELM. As presented in the previous outstanding paper, ELM has many important properties, as follows:

- (i) the smallest norm of weights;
- (ii) minimum approximation error;
- (iii) the minimum norm least-squares solution of $\mathcal{H}\beta = \mathcal{T}$ which is unique.

After the training process, the invariant moments of real-time human body movement trajectory captured by the structured light sensor can be put into the ELM. In the process of the pretrained ELM, the type of the real-time human body movement trajectory could be obtained. Although the training process of the ELM would cost some time, the recognition process of the real-time human body movement trajectory is very fast. In the following experiment analysis section, this paper will present the accuracy and efficiency of this method.

4. Results of the Experiment

Supported by the laboratory's National Science Funding, the researches on the human body movement trajectory recognition are required to be conducted under the office environment, and the experimental data in this paper therefore are required to be specific and unique. The public data sets are more universal and unsuitable for our specific research field. In consequence, the human body movement trajectory data set needs to be collected by ourselves.

In the experiment of different lighting conditions, four people are asked to do four kinds of body movements, as is shown in Figures 10, 11, 12, and 13. Each kind of body movements is repeated 10 times and generates 40 samples for each body movement. Every movement lasts five to fifteen seconds with the image size of 1200×900 .

In the other experiments, if we adopt same raw data in the first experiment, the testing process will be too quick to record the testing time. So the other five people whose heights are from 155 cm to 180 cm are tested. The five people are asked to do five kinds of body movements. These movements and their projection images are shown in Figures 17, 18, 19, 20, and 21.

All the data preprocessing and the comparative experiments (including Sections 4.2, 4.3, 4.4, 4.5, and 4.6) are carried out in MATLAB 2010b environment running in Intel Core 4 Quad 2.9 GHZ CPU with 6 GB RAM. Microsoft's structured light sensor Kinect serves as the three-dimension human body movement trajectory data capture sensor to capture the human body movement trajectory.

4.1. Experiment Condition and Data Preprocessing. This human body movement trajectory recognition experiments are done in normal laboratory environment. In the experiment, people should keep the body forward, perpendicular

FIGURE 10: Motion history image for movement A.

FIGURE 11: Motion history image for movement B.

FIGURE 12: Motion history image for movement C.

FIGURE 13: Motion history image for movement D.

to the horizontal plane, and be about 1.2 to 2 meters to the structured light sensor.

As is mentioned previously, the human body movement trajectory of a single motion has been collected. The motion history image is constructed with whole image sequences as in Figure 1. In consequence, there is no need to find the key frame and detect human motion in some motion image sequences. Each frame in the motion image sequences is equal and important for the three-dimension motion history image. The three-dimension motion history image from the whole motion image sequences is calculated. Besides the three-dimension motion history image represents the feature of human motion.

In this paper, the physical movements monitored are debounced and the centre position data of the prior frame are recorded to compare with the centre position data of the current frame. If the deviation is within the threshold range, the position data of the prior frame is chosen to neglect the jitter of the current frame [27, 32].

During the comparative experiments, repeated experiments are conducted many times to determine the value of the threshold. In the end, the paper sets the value of the threshold as 100 pixels in the follow-up work.

When using the real-time human body movement trajectory, invalid frames will appear at the beginning and the end of the movement. By eliminating the jitter of the physical movements, we could remove the useless part of the physical movements, and all the frames left are presented clearly.

As is mentioned previously, MATLAB program is used to preprocess all raw human body movement trajectory data. Then the motion history image with the preprocessed human body movement trajectory data is calculated.

4.2. The Experiments under Different Feature Extracting Algorithms. In order to verify the robustness of the MHI method

FIGURE 14: Recognition rate in different lighting conditions.

FIGURE 15: Recognition rate in different obstruction conditions.

in the extracting feature step, here some comparing experiments are carried out under different conditions, such as different light, different obstruction, and different background factor. The recognition rate between the HMM method and the MHI method is under different conditions checked.

In the experiment, four people are asked to do four kinds of body movements, as is shown in Figures 10, 11, 12, and 13. Each kind of body movements is repeated 10 times in the different conditions. It generates 40 samples for each body movement in each experiment condition. Every movement lasts five to fifteen seconds with the image size of 1200×900 .

In order to keep the same experiment condition, the same human body movement trajectory raw data is adopted. The first 20 samples of each kind of movements are chosen as training data to get the standard movement templates using the traditional hidden Markov model (HMM) and three-dimensional motion history image approach (MHI) in each experimental condition. Then the other 20 samples left are used as testing data in each experimental condition.

Figure 14 is the recognition accuracy in normal light and the weak light for every human body movement trajectory. The recognition accuracy rate declines sharply by the traditional method in low light conditions. The three-dimensional motion history image approach with 3D depth data could capture the trajectory of human body movement very well even under weak light environment. The experiment shows that the new method has better robustness under the weak light conditions than the traditional method.

By adopting the hat, glass, or mask as obstruction, Figure 15 is the recognition accuracy in different obstruction experimental condition for each kind of human body

FIGURE 16: Recognition rate in different background conditions.

movement trajectory. The recognition accuracy rate declines sharply by the traditional method in strange obstruction experimental condition, such as hat, glass, or mask. The three-dimensional motion history image approach with 3D depth data will capture the trajectory of human body movement very well even under obstruction experimental condition. The experiment shows that the new method has a better robustness under the obstruction experimental conditions than the traditional method.

By adopting some texture and pattern wall as comparing background, Figure 16 is the recognition accuracy in the different background conditions for each human body movement trajectory. As we can see, the recognition accuracy rate declines sharply by the traditional method in same texture and pattern background. The three-dimensional motion history image approach with three-dimensional depth data will capture the trajectory of human body movement very well even under the same texture and pattern background condition. The experiment shows that the new method has better robustness under the same texture and pattern background conditions than the traditional method.

In the favourable conditions, the recognition rate of the MHI method and the HMM method is almost the same, but in the unfavourable conditions (weak light, obstruction, and same background) the MHI method has a higher recognition rate than the traditional HMM method.

According to the result of the comparing experiments, the conclusion could be safely drawn that the MHI method has a better performance than HMM method.

4.3. The Experiments under Different Recognition Algorithms. In order to verify the efficiency and accuracy of the body movement recognition process based on the extreme learning machine, here some experiments under different recognition algorithms are carried out (ELM, KNN, BP, and SVM).

We could compare with the recognition rate and training time of different recognition algorithms (ELM, KNN, BP, and SVM) in the same raw data and experimental conditions.

In these contrast experiments, if we adopt the same raw data in the first experiment, the testing process will be too quick to record the testing time. Therefore more human body movement trajectories are collected and other five people whose heights are from 155 cm to 180 cm are tested. Five people are asked to do five kinds of body movements. These movements and their projection images are as shown in Figures 17, 18, 19, 20, and 21.

Each kind of body movements is repeated for 60 times and generates 1500 samples for each body movement. 200 samples out of 1500 samples are randomly chosen as the test data set and the rest as the training data set.

To keep the same experiment condition, the same numbers of the hidden nodes (50 hidden nodes) are applied and the same activation function is adopted (standard sigmoid function):

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (21)$$

Figures 22 and 23 are the recognition accuracy and elapsed time in several different recognition algorithms for each action. As we can see, there is no significant recognition rate difference between the two algorithms, but the extreme learning machine costs less time than the other recognition algorithms in the training process.

From two aspects (see Figures 22 and 23), we could come to the conclusion that the new method in this paper has a better accuracy, efficiency, and robustness.

4.4. The Experiment under Different Hidden Layer Node Number Conditions. In order to verify the impact of the number of the hidden layer nodes in the extreme learning machine artificial neural network, here the experiment is carried out under different hidden layer node number conditions.

In this comparative experiment, the training time, recognition process time, and recognition accuracy in different numbers of hidden layer node condition are recorded and compared. By comparison, it is quite obvious that the extreme learning machine has a very good performance and a very fast processing speed in our application.

To keep the same experiment condition, the same experimental raw data are chosen in the second experiment and the bright normal laboratory environment. And the same activation function is selected (standard trigonometric function “Sin”).

Table 1 shows that the training time is less than 1 second when the number of the hidden layer nodes is less than 300. The number of the hidden layer nodes is increased to more than 3000, the human body movement trajectory recognition process is still very fast, and the training time is less than 12 seconds, and the recognition accuracy increased. And the testing time is less than 100 milliseconds. So from this experiment, it is quite obvious that the extreme learning machine has a very good performance and a very fast processing speed.

4.5. The Experiment under Different Training Sample Number Conditions. In order to verify the impact of the number of the training sample on the experimental result, here some experiments are conducted under different numbers of training sample conditions.

In this comparative experiment, the training time, recognition process time, and recognition accuracy in different numbers of training sample condition are recorded and compared. From this record, it is easy to see that the extreme learning machine has a very good performance and a very

(a)

(b)

(c)

(d)

FIGURE 17: Motion history image for movement E.

(a)

(b)

(c)

(d)

FIGURE 18: Motion history image for movement F.

FIGURE 19: Motion history image for movement G.

FIGURE 20: Motion history image for movement H.

fast processing speed in a large quantity of training sample application.

To keep the same experiment condition, the same experimental raw data is chosen in the second experiment, the bright normal laboratory environment and 1000 hidden layer

FIGURE 21: Motion history image for movement I.

FIGURE 22: Recognition rate under different recognition algorithms.

FIGURE 23: Training time under different recognition algorithms.

nodes. And the same activation function is adopted (standard trigonometric function “Sin”).

Figure 24 shows that the number of the training sample has some impact on the recognition accuracy. The human body movement trajectory accuracy improves when the number of the training sample increases. But this influence is limited; even the number of the training samples is 500; 96.5%

accuracy of the human body movement trajectory recognition is quite a good result.

4.6. The Experiment under Different Activation Function Conditions. In order to verify the impact of the different activation function on the experimental result, here some

TABLE 1: Different hidden layer node number experiment.

Hidden layer node number	Training time (second)	Testing time (second)	Accuracy
50	0.0312	0.00006	0.8550
100	0.0624	0.00007	0.8000
150	0.0936	0.00005	0.8300
200	0.1248	0.00003	0.9400
250	0.2028	0.00006	0.7850
300	0.2184	0.00004	0.9000
1000	3.3540	0.04690	0.9650
1500	6.3960	0.04680	0.9800
2000	8.3939	0.07800	0.9750
2500	9.5785	0.06800	0.9600
3000	10.3741	0.07800	0.9500
3500	11.2789	0.07800	0.9600

FIGURE 24: Different training sample number experiment.

experiments are carried out under different activation function conditions.

In this comparative experiment, the training time, recognition process time, and recognition accuracy in different activation function condition are recorded and compared. By comparison, it is obviously presented that activation function has a very good performance and a very fast processing speed in our application.

To keep the same experiment condition, the same experimental raw data are chosen in the second experiment, the bright normal laboratory environment and 1000 hidden layer nodes. And the 1350 training sample data and 200 testing data are adopted.

The experiments are repeated many times under the different activation functions condition, and we finally get a series of results. Figure 25 shows that different activation functions have a great impact on the recognition accuracy. The human body movement trajectory accuracy applying the standard trigonometric function "Sin" is much better than other activation functions. But the recognition efficiency using the "sigmoid" function is much better than other activation functions. So in my experiments, the "sigmoid" function is much more suitable in my application.

FIGURE 25: Different activation function experiment.

5. Conclusion

Combined with the three-dimensional motion history image and the extreme learning machine, this paper overcomes the shortcomings of the traditional body movement trajectory recognition method, such as the light, the obstruction, the background, the influence of some specific data, the heavy calculation, and the slow recognition process. The new process in this paper realizes the recognition of body movement trajectory robustly, efficiently, and accurately. Finally after some complicated experiments, it is quite obvious that the new process in the paper has better robustness, efficiency, and accuracy.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by National Nature Science Foundation of China (nos. 61272357 and 61300074) and the new century personnel plan for the Ministry of Education (NCET-10-0221).

References

- [1] D. Weinland, R. Ronfard, and E. Boyer, "Automatic discovery of action taxonomies from multiple views," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 1639–1645, New York, NY, USA, June 2006.
- [2] L. Cheng, Q. Sun, H. Su, Y. Cong, and S. Zhao, "Design and implementation of human-robot interactive demonstration system based on Kinect," in *Proceedings of the 24th Chinese Control and Decision Conference (CCDC '12)*, pp. 971–975, Taiyuan, China, May 2012.
- [3] D. Weinland, R. Ronfard, and E. Boyer, "Free viewpoint action recognition using motion history volumes," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 249–257, 2006.
- [4] Y. Qi, K. Suzuki, H. Wu, and Q. Chen, "EK-means tracker: a pixel-wise tracking algorithm using kinect," in *Proceedings of the 3rd Chinese Conference on Intelligent Visual Surveillance (IVS '11)*, pp. 77–80, Beijing, China, December 2011.

- [5] D. Weinland, R. Ronfard, and E. Boyer, "Motion history volumes for free viewpoint action recognition," in *Proceedings of the IEEE International Workshop on Modeling People and Human Interaction*, 2005.
- [6] S. B. Lee, "Real-time stereo view generation using kinect depth camera," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1153–1156, October 2011.
- [7] G. Huang and X. Cheng, "An automatic recognition approach of human gestures," *Journal of Southwest China Normal University*, vol. 35, no. 4, pp. 136–140, 2010.
- [8] O. Rashid, A. Al-Hamadi, and B. Michaelis, "Robust hand posture recognition with micro and macro level features using kinect," in *Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems*, pp. 631–635, 2011.
- [9] S. Xu and Q. Peng, "Three-dimensional object recognition based combined moment invariants and neural network," *Computer Engineering and Applications*, vol. 44, no. 31, pp. 78–80, 2008.
- [10] J. Liu and Z. Qu, "Real-time detecting and tracking of multiple moving object based on improved motion history image," *Computer Applications*, vol. 28, no. 6, pp. 198–201, 2008.
- [11] J. Liu, M. Shah, B. Kuipers, and S. Savarese, "Cross-view action recognition via view knowledge transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 3209–3216, June 2011.
- [12] Y. M. Lui, J. R. Beveridge, and M. Kirby, "Action classification on product manifolds," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 833–839, San Francisco, Calif, USA, June 2010.
- [13] F. Lv and R. Nevatia, "Single view human action recognition using key pose matching and viterbi path searching," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, Minneapolis, Minn, USA, June 2007.
- [14] K. Guo, P. Ishwar, and J. Konard, "Action recognition in videos by covariance matching of Sillhouette tunnels," in *Proceedings of the 22nd Brazilian Symposium on Computer Graphics and Image Processing*, pp. 299–306, 2009.
- [15] K. Guo, P. Ishwar, and J. Konrad, "Action recognition using sparse representation on covariance manifolds of optical flow," in *Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS '10)*, pp. 188–195, IEEE, Boston, Mass, USA, August-September 2010.
- [16] A. Kovashka and K. Grauman, "Learning a hierarchy of discriminative space-time neighborhood features for human action recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2046–2053, San Francisco, Calif, USA, June 2010.
- [17] J. Liu and M. Shah, "Learning human actions via information maximization," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.
- [18] R. Minhas, A. Baradarani, S. Seifzadeh, and Q. J. Wu, "Human action recognition using non-separable oriented 3d dual-tree complex wavelet," in *Proceedings of the 9th Asian Conference on Computer Vision (ACCV '09)*, pp. 226–235, 2009.
- [19] R. Minhas, A. Baradarani, S. Seifzadeh, and Q. M. J. Wu, "Human action recognition using extreme learning machine based on visual vocabularies," *Neurocomputing*, vol. 73, no. 10–12, pp. 1906–1917, 2010.
- [20] G.-B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107–122, 2011.
- [21] Y. Gu, J. Liu, Y. Chen, and X. Jiang, "Constraint online sequential extreme learning machine for lifelong indoor localization system," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '14)*, pp. 732–738, July 2014.
- [22] R. Minhas, A. A. Mohammed, and Q. M. J. Wu, "Incremental learning in human action recognition based on Snippets," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 11, pp. 1529–1541, 2012.
- [23] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, 2001.
- [24] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2247–2253, 2007.
- [25] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Transactions on Neural Networks*, vol. 17, no. 6, pp. 1411–1423, 2006.
- [26] K. Schindler and L. van Gool, "Action Snippets: how many frames does human action recognition require?" in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.
- [27] L. Yeffet and L. Wolf, "Local trinary patterns for human action recognition," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 492–497, Kyoto, Japan, October 2009.
- [28] X. Liu and K. Yuan, "Weight moment method based on hu invariant moments and applications," *Journal of Dalian Nationalities University*, vol. 12, no. 5, pp. 470–472, 2010.
- [29] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN '04)*, pp. 985–990, Budapest, Hungary, July 2004.
- [30] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [31] J. C. A. Barata and M. S. Hussein, "The Moore-Penrose pseudoinverse: a tutorial review of the theory," *Brazilian Journal of Physics*, vol. 42, no. 1-2, pp. 146–165, 2012.
- [32] L. Shao and R. Mattivi, "Feature detector and descriptor evaluation in human action recognition," in *Proceedings of the ACM International Conference on Image and Video Retrieval (CIVR '10)*, pp. 477–484, July 2010.