Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Recognition of Human Body Movements Trajectory Based on the Three-dimensional Depth Data

Zheng Chang Xiaojuan Ban
Qing Shen Jing Guo

School of Computer and Communication Engineering
University of Science and Technology Beijing.
E-mail:
elephant.cz@gmail.com
banxj@ustb.edu.cn
shenqingcc222333@gmail.com
guojing7192356@163.com

Abstract: based on the traditional machine vision recognition technology about human body movement trajectory, the paper finds out the shortcomings of the traditional recognition technology. By combining the three - dimensional depth data, the three-dimensional motion history image and the invariant moments of the three - dimensional motion history image computed as the eigenvector of body movements, the paper applies the method to the machine vision of the human body movements trajectory. In detail, the paper describes the algorithm and realization scheme of the human body movements trajectory recognition based on the three-dimensional depth data. Finally, comparing with the results of the recognition experiment, we verify that the method of human body movement trajectory recognition technology based on the three-dimensional data has a more accurate recognition rate and a better robustness.

Keywords: Artificial intelligence, Human-machine interface, Image recognition, Motion estimation.

1. INTRODUCTION

With the rapid development of the machine vision technology, the machine vision recognition and natural human-computer interaction technology have also become an important research direction in the machine vision technology. As the human body movement is a natural and intuitive communication mode (Daniel Weinland, 2006), beyond all doubts, the human body movement recognition technology has become an indispensable technology to the new generation of natural human-computer interaction interface (Remi Ronfard, 2006), especially for the disabled who can only use their body movements to give the wheelchair and the auxiliary equipment orders, which will bring them more convenience.

The prior human body movements 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 (Edmond Boyer 2003), however, due to the diversity, ambiguity and disparity in time and space of the human body movements, the traditional body movements trajectory recognition researches have great limitations. The paper attempts to introduce three-dimensional depth data into the human body movement trajectory recognition, which will make the machine vision recognition of the human body movement trajectory more accurate and possess more robust feature.

2. PROBLEM DESCRIPTION

Human body movement trajectory recognition is the matching between the human body movements trajectory captured by the sensor and the pre-defined sample movement trajectory. The traditional method applies the Hidden Markov Model to realize the match process (Guofan Huang et al., 2010).

Figure 1. Human Body Movements Two-dimensional Image

As is shown in Fig1, based on the two-dimensional images and the Hidden Markov Model, the whole process of comparing between the real-time trajectory and the pre-defined sample trajectory is shown in Fig2 (Sheng Xu et al., 2008). But the recognition based on two-dimensional images still has several limitations, as follows:

- Light: when the light condition is changed, the luminance information of the human head will change since the images captured by the sensor are easily affected by natural light and artificial light.
- Obstruction: In the recognition process, the human body movement trajectory may be obscured by some objects in the environment. While the obstruction can cause the loss of the identification information, it will affect the reliability of the recognition greatly.

- Background: In the actual recognition process, if the factors (color, texture, shape etc.) of the body movement area and the background area are similar, it will also increase the difficulty of the recognition (Junxue Liu et al., 2008).

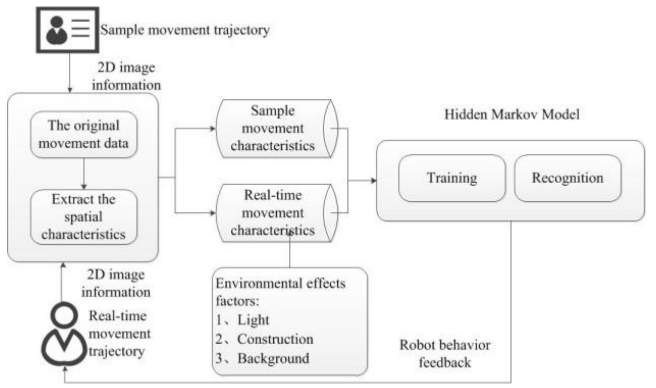

Figure 2. The HMM Based on Two-dimensional Image

Based on the three-dimensional depth data, the three-dimensional Hidden Markov Model (3DHMM) is free from the effects of light, obstruction and background, but this method has been restricted to the application field of real-time monitoring because of the huge amount of calculation, the inefficiency of the training and the easy accessibility to the local optimal value etc.

To solve these problems, this paper attempts to compare the motion history image (MHI) with the three-dimensional depth data of the human body movements in order to get the three-dimensional motion history image (3DMHI) of body movements (Fig3). And then it calculates seven invariant moments of the three-dimensional motion history image working as the eigenvector of the human body movements. Finally we build the template set of the human body movements and calculate the mean vector of the template set and the covariance matrix. In the recognition process, we use the Mahalanobis distance to measure similarity between the new input body movements and the movement template. As long as the calculated Mahalanobis distance is within the range of the pre-determined threshold value, the body movement trajectory is successful. The new method not only is free from the effects of the illumination, obstruction, background and other environmental factors, but also improves the efficiency and accuracy of human body movement recognition.

Figure 3. The MHI Based on Three-dimensional Depth Data

3. PROBLEM SOLVING

3.1. Body Movements Characterized by Three-dimensional Motion History Image

To characterize the 3D motion information, the paper improves the application of traditional motion history images approach based on two-dimensional image so as to combine with the three-dimensional depth data.

The motion history image approach is a branch of The Finite Difference Time Domain Method (FDTD). The mechanism of the FDTD method is to get different images from a continuous image sequences by comparing with two or three adjacent pixels in the corresponding frames, then to extract the moving regions in the image by setting the threshold. By introducing the 3D data, the paper presents the improved FDTD method as follows:

$$\mathcal{D}(x,y,z,n) = \mathcal{I}(x,y,z,n-1) - 2\mathcal{I}(x,y,z,n) + \mathcal{I}(x,y,z,n+1) \tag{1}$$

Among them, $\mathcal{I}(x,y,z,n)$ represents the pixel gray 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 human body movement changed area. The threshold $\mathcal{D}(x,y,z,n)$ is as follows:

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

Γ is the specially selected threshold. If the value is too low, it can't 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 repeated many times to determine the value of the threshold. In my experiment environment, 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:

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

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, but also reflects the direction and state of them. In the motion history image, the gray value of each pixel is in proportion with the duration of the body movement in the position. The recent body gestures have the maximum gray value. Gray value changes reflect the direction of the body movements.

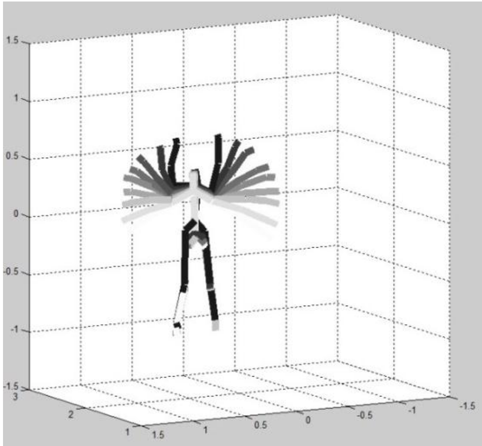

Figure 4. Three-dimensional Motion History Image of Body Movements (MHI)

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, it is too sensitive to the observation position. 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 and angle.

To calculate the invariant moments, after getting the three-dimensional motion history image, we project it in the XY plane (Fig 5), YZ plane (Fig 6) and XZ plane (Fig 7). This method can get three views of three-dimensional motion history image with one gesture. Then we have the calculation of invariant moments for the three main views.

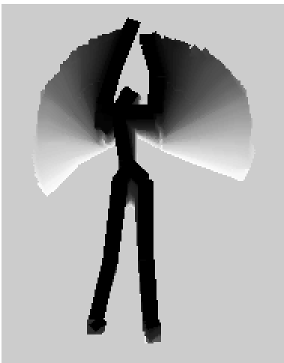

Figure 5. XY Surface Projection of the MHI

Figure 6. YZ Surface Projection of the MHI

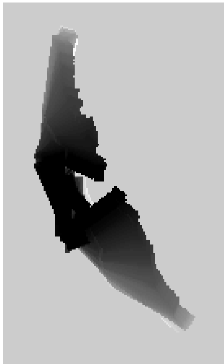

Figure 7. XZ Surface Projection of the MHI

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:

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

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

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

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

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

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}}。 \tag{6}$$

Among them:

$$m_{00} = \sum_{x=1}^{N} \sum_{y=1}^{M} f(x, y)$$

$$m_{10} = \sum_{x=1}^{N} \sum_{y=1}^{M} f(x, y) x^{1}$$

$$m_{01} = \sum_{x=1}^{N} \sum_{y=1}^{M} f(x, y) y^{1}.$$

Then through the normalizing of the central moment by the zero order central moments μ_{00} ,we can get the normalized center moment of the motion history image.

$$\eta_{pq} = \frac{\mu_{pq}}{{\mu_{00}}^{r}}, r = \frac{p+q+2}{2}, p + q = 2,3,4, \ldots \tag{7}$$

Hu·M·K get seven invariant moments based on the linear combination of two order and three order normalized central moment. The image translation, rotation and scaling are unchanged and the invariant moments are as follows:

$$M_1 = \eta_{20} + \eta_{02} \quad (8)$$

$$M_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}2 \quad (9)$$

$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (10)$$

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (11)$$

$$M_5 = (\eta_{30} - 3\eta_{12}) \times (\eta_{30} + \eta_{12}) \times [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (\eta_{03} - 3\eta_{21})(\eta_{03} + \eta_{21})[(\eta_{03} + \eta_{21})^2 - 3(\eta_{12} + \eta_{30})^2] \quad (12)$$

$$M_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (13)$$

$$M_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{21})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (3\eta_{12} - \eta_{30})(\eta_{03} + \eta_{21})[(\eta_{03} + \eta_{21})^2 - 3(\eta_{12} + \eta_{30})^2] \quad (14)$$

Because 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.

$$M_k = \log|M_k|, k = 1,2,3,4,5,6,7 \quad (15)$$

The invariant moments still have a 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 volume (Xiaoniu Liu et al., 2010).

3.3. The Motion History Image Recognition of Body Movements

In the process of recognition, we collect the sample of human body movement first and build a training template library so that we can get the standard eigenvector. In order to get better recognition results, the sample of the human body movement must be definite and slow therefore, all of the training template must comply with the criterion.

For the same body movement, different people involved shall repeat the action several times. We get multiple groups of three-dimensional motion history images for each movement and then get the mean of these eigenvectors and the covariance matrix. After doing this, the template for each gesture is established.

For the Mahalanobis distance between new movement calculation and standard action template, it is defined as follows:

$$\gamma^2 = (f - \mu_r)^T c^{-1} (f - \mu_r) \quad (16)$$

γ is Mahalanobis distance, f is the eigenvector of motion history image, μ_r is the mean vector of the eigenvectors trained. c is the covariance matrix of the eigenvectors trained.

In the recognition process, an optimal threshold is determined according to the order of each invariant moment using the classical AdaBoost algorithm. Then we use Mahalanobis distance to measure the similarity between the new input gestures and body movements which has been trained by template. If the Mahalanobis distance is within the scope of the provisions of the threshold, it can be considered as a successful match. If we get more than one template matching, we choose the minimum distance as the template.

4. RESULTS OF THE EXPERIMENT

4.1. Data Pre-processing

This trajectory recognition experiments are done in normal laboratory environment. In the experiment, people should keep the body forward, perpendicular to the horizontal plane and be about 1.2 to 2 meters to the Kinect. In this paper, we debounce the physical movements monitored and record the centre position data of the prior frame to compare with the centre position data of the current frame. If the deviation is within the threshold range, we can show the position data of the prior frame to ignore the jitter of the current frame.

When using the real time trajectory, invalid frames will appear at the beginning and the end of the movement. In order to remove the invalid part and get the middle part, we debounce the physical movements to make sure that the motion part displacement will decrease and all the frames can be used.

In the experiment, we let four people to do four kinds of body movements, as is shown in Figure 8, Figure 9, Figure 10 and Figure 11. Each action is repeated 10 times and generates 40 samples for each body movement. Every movement lasts five to fifteen seconds with the image size of1200×900.

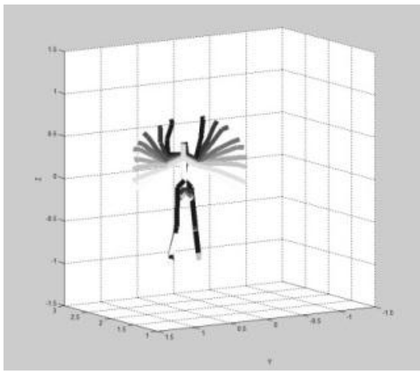

Figure 8. Motion History Image for Movement A

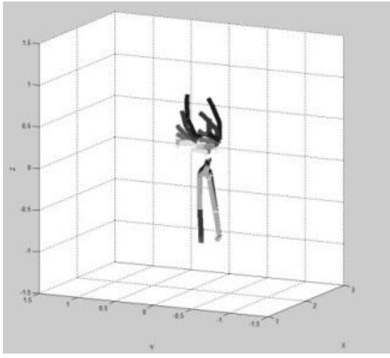

Figure 9. Motion History Image for Movement B

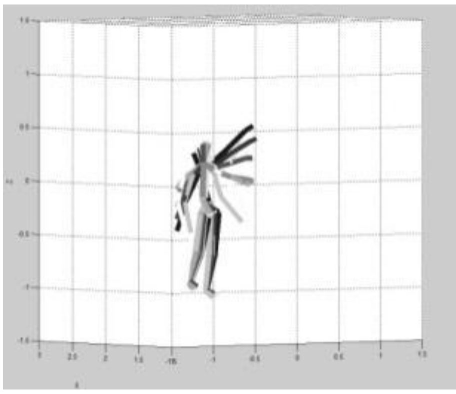

Figure 10. Motion History Image for Movement C

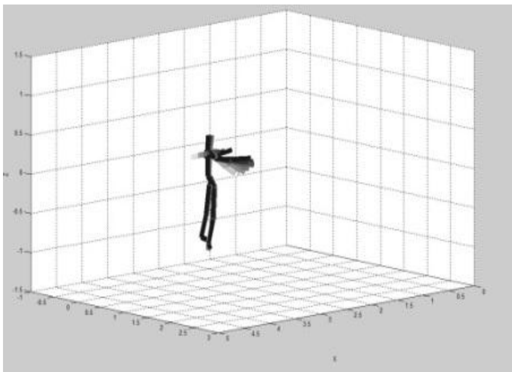

Figure 11. Motion History Image for Movement D

We choose the first 20 samples of each kind of movements for training to get the standard movement templates and the traditional hidden Markov model (3DHMM) and three-dimensional motion history image approach (3DMHI) to test the other 20 remaining samples.

4.2. The Analysis of the Data Results

In order to verify the robustness of the ambient light factor, here we do the experiment under different lighting conditions. Table 1 is the recognition accuracy in normal light and the weak light for every action. We can see that recognition accuracy rate declined sharply by the traditional method in low light conditions. The three-dimensional motion history image approach with 3D depth data will capture the trajectory of human motion very well even under weak light environment. The experiment shows that the new method has a better robustness.

Table 1. Recognition Rate

Movement Types	Recognition Rate under Ordinary Light Environment		Recognition Rate under Weak Light Environment	
	3DHMM	3DMHI	3DHMM	3DMHI
Movement A	0.85	0.91	0.60	0.89
Movement B	0.84	0.90	0.54	0.88
Movement C	0.80	0.89	0.55	0.86
Movement D	0.83	0.89	0.53	0.87

Besides as for different skin colours and clothes, the above method still can automatically adapt. There is no need to adjust the colour reference value manually according to the target body's clothes or skin colours. Last but not the least, under my experiment condition, the three-dimensional motion history image approach can keep an average of 15 frames per second in the experiment theoretically. However the traditional hidden Markov model method just can keep theoretically 9 frames per second. So the three-dimensional motion history image approach can be basically applied to the real-time monitoring.

5. CONCLUSION

Combined with the three-dimensional depth data and motion history image, this paper extends the traditional method to three-dimensional space so as to realize the recognition of human body movements trajectory. This paper overcomes the shortcomings of the traditional method, for example the light, the obstruction and the background, and meets the real-time and accuracy requirements. Finally through the experiments, we verify that the new method in the paper has a more accurate recognition rate and better robustness.

ACKNOWLEDGEMENT

This work was supported by National Nature Science Foundation of P.R. China (No.61272357, 61300074), the Fundamental Research Funds for the Central Universities (FRF-TP-09-016B), the new century personnel plan for the Ministry of Education (NCET-10-0221) and by China Post-doctoral Science Foundation (No. 20100480199).

REFERENCE

Daniel Weinland. Automatic discovery of action taxonomies from multiple views [C]. IEEE Computer society conference on computer vision and pattern recognition, 2006.

Edmond Boyer. Motion history volumes for free viewpoint action recognition [J]. IEEE International workshop on modeling people and human interaction, 2005.

Guofan Huang, Xiaoping Cheng. An automatic recognition approach of human gestures [J]. Journal of southwest china normal university, 2010, 35(4):136~140.

Junxue Liu, Zhenshen Qu. Real-time detecting and tracking of multiple moving object based on improved motion history image. Computer applications, 2008, 28(6):198~201.

Remi Ronfard. Free viewpoint action recognition using motion history volumes [J]. Computer vision and image understanding, 2006, 104:249~257.

Sheng Xu, Qicong Peng. Three-dimensional object recognition based combined moment invariants and neural network. Computer engineering and applications, 2008, 44(31):78~80.

Xiaoniu Liu, Kejie Yuan. Weight moment method based on hu invariant moments and applications. Journal of dalian nationaloties university, 2010, 12(5):470~472.