

Research Article

Human Action Recognition Technology in Dance Video Image

Lei Qiao¹ and QiuHao Shen ²

¹Jiaozuo University Art College Dance Teaching and Research Section, Jiaozuo 454000, China

²School of Humanities and Law, Zhejiang A&F University, Zhejiang 311300, China

Correspondence should be addressed to QiuHao Shen; 2019012389@zwu.edu.cn

Received 31 August 2021; Revised 8 October 2021; Accepted 16 October 2021; Published 3 November 2021

Academic Editor: Bai Yuan Ding

Copyright © 2021 Lei Qiao and QiuHao Shen. 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.

In order to effectively improve the recognition rate of human action in dance video image, shorten the recognition time of human action, and ensure the recognition effect of dance motion, this study proposes a human motion recognition method of dance video image. This recognition method uses neural network theory to transform and process the human action posture in the dance video image, constructs the hybrid model of human motion feature pixels according to the feature points of human action in the image coordinate system, and extracts the human motion features in dance video image. This study uses the background probability model of human action image to sum the variance of human action feature function and update the human action feature function. It can also use Kalman filter to detect human action in dance video image. In the research process, it gets the human multiposture action image features according to the linear combination of human action features. Combined with the feature distribution matrix, it processes the human action features through pose transformation and obtains the human action feature model in the dance video image to accurately identify the human action in the dance video image. The experimental results show that the dance motion recognition effect of the proposed method is good, which can effectively improve the recognition rate of human action in dance video image and shorten the recognition time.

1. Introduction

The advent of the intelligent era promotes the rapid development of computer image processing technology. Human action recognition technology has always been a hot topic in the field of computer vision [1]. The purpose of studying human action recognition technology is to effectively extract the human motion features of video image, analyze human action features in video through effective preprocessing of video image, extract human action features from video image and classify them, and finally realize effective recognition of image features. This recognition technology has been widely used [2, 3]. It is of great significance to analyze the human action in the dance video image by using the characteristics of computer vision, realize the recognition of human action, and correct the wrong action in time to achieve high-quality development.

At present, scholars in related fields have conducted research on video image recognition technology and

achieved some research results. Guo et al. [4] proposed a human action feature recognition technology based on image feature similarity. Firstly, they use the image analysis method to make action recognition of the video image, then carry out the dimension reduction, and obtain a group of new human action representation models at the same time, use the image feature similarity technology for secondary recognition analysis, and calculate the similarity of the same two images through the adaptive analysis of the image, and finally obtain the recognition results of human action features by the weighted processing method. This method can effectively improve the recognition accuracy of similar actions in human action video. Yu and Min [5] proposed a human action recognition algorithm based on improved time network. Firstly, they extract human action features based on improved time network, construct human recognition model through neural network, use CNN framework for grid fusion, and analyze the characteristic of neural grid. Then, they use the same structure as spatial network for

weighted summation, obtain a set of new feature vectors, and iterate the processing results. Finally, they get a new set of human action features and obtain the classification recognition results through the recognition of the two groups of human motion features. This method has good recognition effect. Based on the above analysis, this paper proposes a method of human action recognition in dance video image, which provides a certain reference for further improving the recognition rate of human action in dance video image and shortening the recognition time of human action.

2. Design of Human Motion Recognition Technology in Dance Video Image

2.1. Extraction of Human Action Features from Dance Video Image. In the design process of this technology, the human action features in the dance video image are extracted by using the neural grid theory, and the feature points of the extracted feature image are classified and processed. Combined with the above theory, the following research carries out the posture transformation of the human action target in the dance video image, which can be expressed as

$$A_{(i,j)}(k,l) = B(i,j) \cdot C(k,l). \quad (1)$$

In the above formula, $A_{(i,j)}(k,l)$ represents the filtering result of the human action image in the dance video image after processing, $B(i,j)$ represents the human dance target action image, and $C(k,l)$ represents the filtering combination result. By processing the above results, the definition of $p+q$ order matrix human action image $D(x,y)$ in dance video image is obtained, which is expressed as

$$E_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q D(x,y). \quad (2)$$

In formula (2), the center coordinate of the dance video image is (\bar{x}, \bar{y}) .

Through the iterative processing of human action image in dance video image [6–8], the feature point (i,j) in the coordinate system of human dance action image is obtained, and the pixel probability within the moment t is

$$p(G_{ij}^t) = \sum_{s=K,L} p_s(G_{ij}^t | \theta_{ij,s}^t). \quad (3)$$

In this formula, G_{ij}^t represents the pixel value of the human action in the dance video image at time t . It is assumed that the actual probability of the pixel (i,j) of the human motion image is $p_s(G_{ij}^t | \theta_{ij,s}^t)$ at any time t . The mixed model of human action feature pixel $p_s(G_{ij}^t | \theta_{ij,s}^t)$ in the dance video image at time t is as follows:

$$p_b(G_{ij}^t) = \sum_k w_{ij,b,k}^t (G_{ij}^t, \mu_{ij,b,k}^t). \quad (4)$$

In the above formula, k represents the number of recognition models of human action image in dance video image. The weight of feature vector of human action image in dance video image is $w_{ij,b,k}^t$ at time k . When $\sum_k w_{ij,b,k}^t = 1$, the characteristic vector value of human action image is

$\mu_{ij,b,k}^t$, and the variance matrix of human action features in dance video image is \mathbf{G} at time t [9–11].

Through weight analysis, the fitness value of individual action image in the dance video image can be obtained, and suitable human action feature distribution model can be found in the dance video image. At any time t , in the dance video image, the probability value of the action image pixel (i,j) of the human action is expressed as follows:

$$\eta(Y, \mu, \mathbf{v}) = \frac{1}{(2\pi)^{n/2} |\mathbf{v}|^{1/2}} \exp\left(-\frac{1}{2}(Y - \mu)^T \mathbf{v}\right). \quad (5)$$

In this formula, Y and η represent the feature vectors of the coordinate point (i,j) of the human action image in the dance video image, in which the probability value of the coordinate point is μ , and \mathbf{v} represents the variance matrix of the human action image in the dance video image.

According to the pixel information of the human action image in the obtained dance video image, the judgment pixel can be obtained, which is as follows:

$$p(H_B) = \frac{\lambda_a}{5}, \quad (6)$$

$$p(H_F) = 1 - p(H_B).$$

In formula (6), $p(H_B)$ represents the characteristic distribution function of the pixel value HB of the human action image, $p(H_F)$ represents the characteristic distribution function of the pixel value HF of the body action image. When $p(H_B) \geq p(H_F)$, the human action image feature can be used as foreground pixel. When $p(H_B) < p(H_F)$, the human action image can be used as the background pixel [12–14].

After obtaining the background of the human action image in the dance video image, comparing the human action image in the dance video image with the standard motion image, the features of the human action in the dance video image can be obtained, which can be expressed as

$$p_u(y) = I_h \sum_{i=1}^n k \left(\left\| \frac{y - x_i}{h} \right\|^2 \right) \delta[b(x_i) - u]. \quad (7)$$

In the above formula, I_h represents the normalization processing result of the human action image in the dance video image, u represents the characteristic value of the human action image, y represents the center pixel (i,j) of the human action image in the dance video image. The feature of the human action in the dance video image can be obtained by formula (7), so as to realize the extraction of the human action features.

2.2. Detection of Human Action in Dance Video Image. Assuming that the variance of the gray value distribution of the human movements in the dance video image is σ , all the gray values meet the expected value μ of the human action image and the Gaussian distribution of the human action in dance image at this time [15–17]. Then, the background probability model of human action image in dance video image can be obtained by using the following formula:

$$p(x|w_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right). \quad (8)$$

In formula (8), $p(x|w_i)$ represents the normal distribution of human action image features, and w_i represents the feature vector of gray value of human action. Assuming that $K = 2, 3, \dots, L$, the variance summation of the human action feature function in the dance video image can be expressed as

$$\begin{aligned} \mu'_{K+1} &= \frac{K-1}{K}\mu'_K + \frac{1}{K}\mu_K, \\ \sigma'^2_{K+1} &= \frac{K-1}{K}\sigma'^2_K + \frac{1}{K}\sigma_K^2. \end{aligned} \quad (9)$$

Then, in frame $K = L + 1$, if updated the variance of the human action feature function in the dance video, it can be expressed as

$$\begin{aligned} \mu'_{K+1} &= (1-\rho)\mu'_K + \rho\mu_K, \\ \sigma'^2_{K+1} &= (1-\rho)\sigma'^2_K + \rho\sigma_K^2. \end{aligned} \quad (10)$$

In the above formula, $\mu_K = 1/N_K \sum_{j=1}^{N_K} I_j$ represents the statistic mean value of human action image in the dance video image of frame K , $\sigma_K^2 = 1/N_K \sum_{j=1}^{N_K} (I_j - \mu_K)^2$ represents the variance of human motion feature vector, μ'_K represents the Gaussian model feature function, σ'^2_K represents the Gaussian model variance of human action, N_K represents the sample number of human action, I_j represents the gray value of human action feature points, and ρ represents the feature update rate.

In the analysis of human action in dance video image, in order to detect the relationship between human action features and basic feature points, this study carries out the target comparison of human action image in dance video image by using Kalman filter [18]. Assuming that the action amplitude of the human action image in the first dance video image is X_1 , then the feature of detecting the human action in the dance video image can be expressed as

$$\hat{X}_{K+1} = MX_K. \quad (11)$$

In formula (11), M represents the structure parameter in the dance video image, and X_K represents the detection variance value of human action in the K -th dance video image. Combined with the above formula, the detection of human action in dance video image can be completed.

2.3. Recognition of Human Action in Dance Video Image. In order to recognize human action in dance video image, the mathematical model is established as follows:

By connecting the human action samples in the dance video image, the recognition model of the human action in the dance video image can be obtained. The coordinate point of this model is p , and $T = (Y_1, Z_1, A_1, \dots, Y_p, Z_p, A_p)^U \in S^{3p}$. The number of human action samples in this model is n , and the human action component in the dance video image is T_j , and then $j = 1, 2, \dots, n$. T represents the behavior classification of human action. It can

be obtained through normalization processing of all human actions in dance video image [19, 20] and also can be used to describe the linear combination of human action features, which is as follows:

$$T = \sum_{j=1}^n x_j T_j. \quad (12)$$

Using the above analysis method, both the human action matrix of the dance video image and the human action image features can be obtained:

$$U = \bar{U} + Tu. \quad (13)$$

In the above formula, \bar{U} represents the mean value for classification of human action behavior in dance video image, and u represents the human action matrix in dance video image. According to the human action feature vector [21], the multiposture features of the human body can be obtained:

$$U^s = \bar{U}^s + \mathbf{G}^s U. \quad (14)$$

In formula (14), \mathbf{G}^s represents the offset matrix of human action in the dance video image; U^s and \bar{U}^s are the state variables in the changing process of human dance posture. Based on the feature distribution matrix K , the human action feature model of dance video image can be obtained by carrying out pose transformation of human dance action features. The formula is as follows:

$$(U^l)^s = \mathbf{K} \left[(\bar{U}^l) + (G^l)U^s \right]. \quad (15)$$

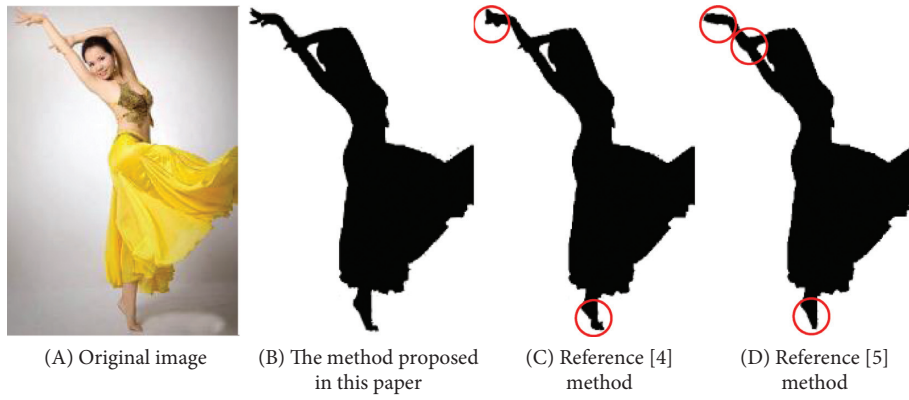
In the above formula, U^l represents the feature points of human action in dance video image [22]. By extracting the state variables of human dance action image, the feature vector of human dance action and the feature components \bar{U}^l and G^l of human dance action can be obtained. Through the above steps, the recognition of human action in dance video image can be realized.

3. Experimental Analysis

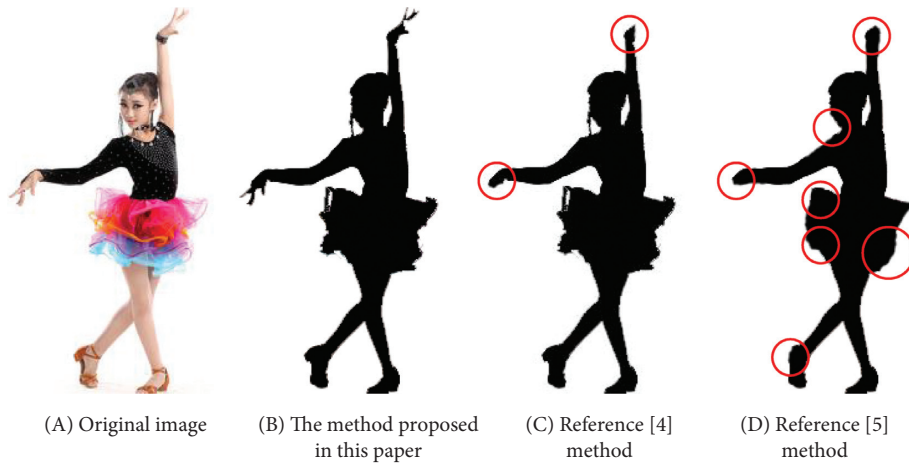
In order to verify the effectiveness of human action recognition technology in dance video image in practical application, this experiment chooses the AIST++ dance video dataset and uses MATLAB simulation software as the experimental platform to carry out simulation verification.

In the AIST++ dance video dataset, the most common dance actions mainly include belly dance action, Latin dance action, ballet dance action, national dance action, and folk dance action. This experiment respectively uses the method proposed in this paper and the methods in literature [4] and literature [5] for recognition and compares the recognition effects of dance action with different methods, as shown in Figure 1.

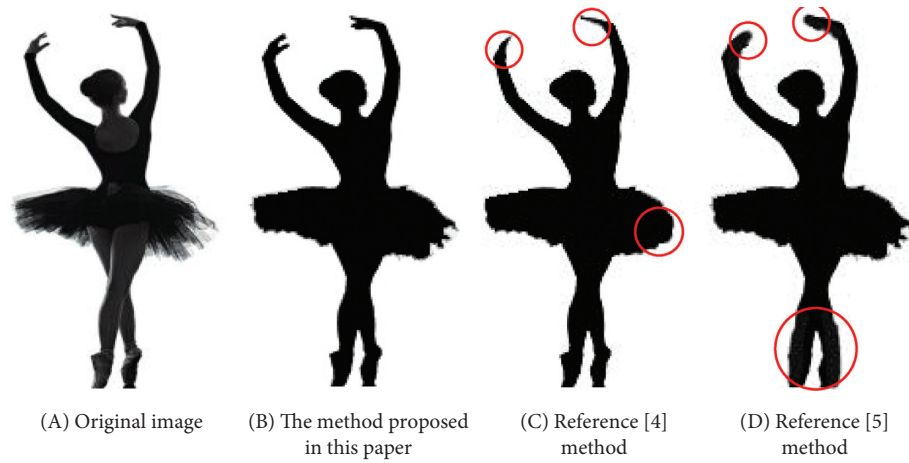
According to Figure 1, for the belly dance action, the dance action recognized by the methods in literature [4] and literature [5] cannot accurately identify the image details, while the method proposed in this paper can accurately



(a)



(b)



(c)

FIGURE 1: Continued.

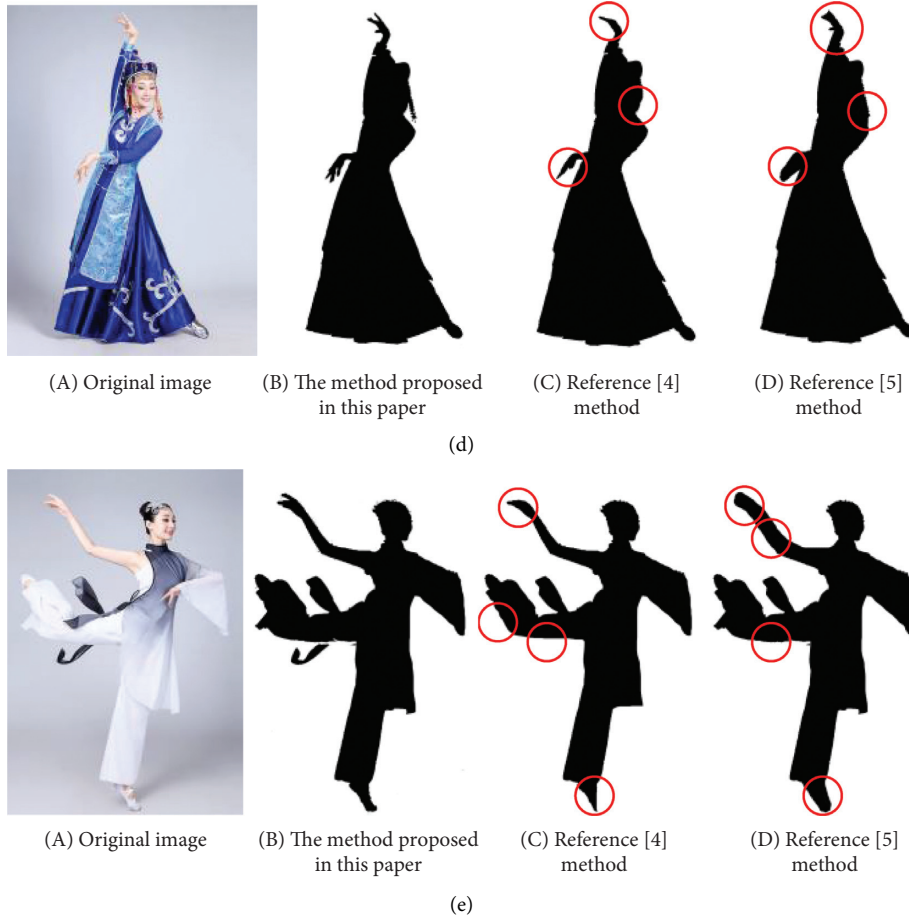


FIGURE 1: Recognition effect of different methods on dance action. (a) Belly dance action. (b) Latin dance action. (c) Ballet dance action. (d) National dance action. (e) Folk dance action.

identify the image contour. For Latin dance action, the dance action contour edges recognized by the methods in literature [4] and literature [5] are missing, while the dance action contour edges of the method proposed in this paper are complete. For the ballet dance action, the hand and foot areas of dance action recognized by the methods in literature [4] and literature [5] are deformed, and the method proposed in this paper can effectively restore the original image shape. For the national dance action, the methods in literature [4] and literature [5] still cannot effectively identify the details of the hand area of dance action, but the method proposed in this paper can accurately restore the original shape of the original image on the hand area. For the folk dance action, both the methods in literature [4] and literature [5] have deviations from the original image, while the method proposed in this paper avoids the image recognition deviations and accurately identifies the image edges. Based on the above analysis, it can be seen that the recognition effect of the proposed method is better, because the proposed method uses Kalman filter to detect human action in dance video image, which can effectively remove image noise, so as to ensure the recognition effect of dance action.

During the experiment, the neural network method is to extract the dance video image in the AIST++ dance video

dataset frame by frame, and the descriptor matrix is taken as the training sample. According to the dance action samples of the collected dance video image, the experiment aims at five dance action types in AIST++ dance video dataset, and it also tests the recognition rate of the dance action type by the method proposed in this paper. The results are shown in Figure 2.

According to Figure 2, in the AIST++ dance video dataset, when the method proposed in this paper is applied to identify dance action types, the recognition rates of five dance actions are high. In addition, among them, the recognition rate of belly dance action is the highest, which can reach 92%, and the recognition rate of folk dance action is the lowest, which is 85%. Then, the calculation shows that the average recognition rate of the proposed method is 88.7% when identifying five dance action types in AIST++ dance video dataset. Therefore, the recognition rate of dance action types by the proposed method is high.

In order to further verify the recognition effect of the proposed method on human action in dance video image, this experiment selects the data in 400 MB AIST++ dance video dataset, uses the proposed method for recognition, and obtains the recognition rate of human action in dance video image by the proposed method, as shown in Figure 3.

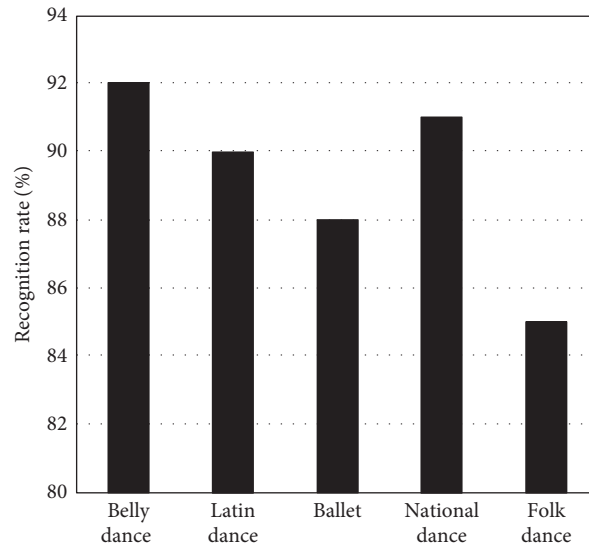


FIGURE 2: The recognition rate of dance action type by the proposed method.

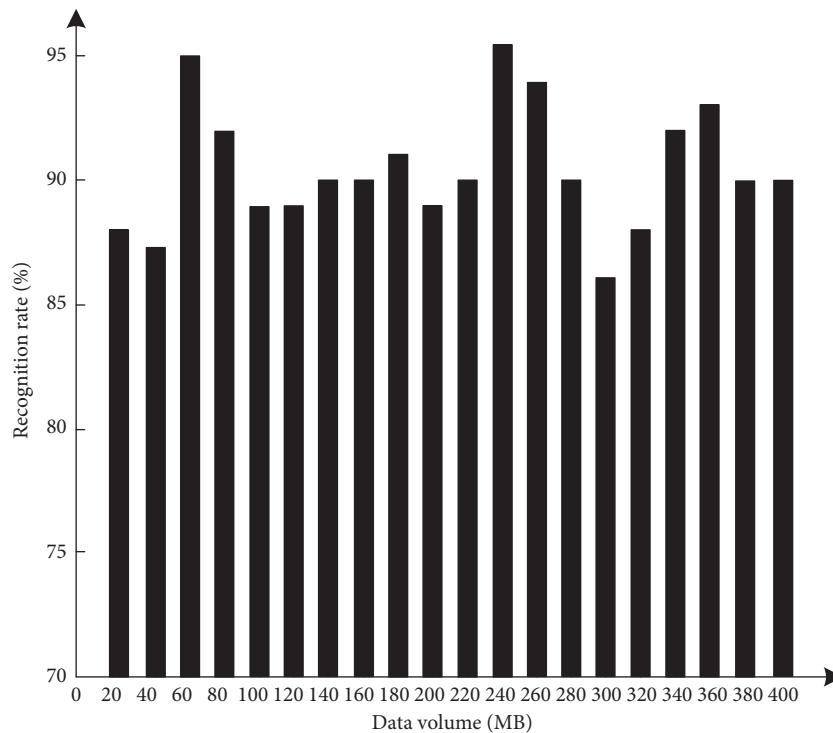


FIGURE 3: The recognition rate of human action in dance video image by the proposed method.

As it can be seen from the results in Figure 3, when the proposed method is used to identify human action in the AIST++ dance video dataset, the recognition rate is high. Among them, when the data amount of AIST++ dance video dataset is 240 MB, the highest recognition rate reaches 95.5%. However, when the data amount in the AIST++ dance video dataset is 300 MB, the lowest recognition rate is 86%. According to the calculation, when the data volume in AIST++ dance video dataset is 400 MB, the average recognition rate of human action in dance video image by the proposed method is 90.4%. Therefore, the proposed method

can effectively improve the recognition rate of human action in dance video image.

On the above basis, the recognition time of human movements in dance video image is further verified. The 400 MB AIST++ dance video data are selected and the recognition time of human movements in dance video image is obtained by using the method proposed in this paper, as shown in Figure 4.

According to the data in Figure 4, the recognition time of human action in the dance video image by the proposed method also increases with the increase of the data amount

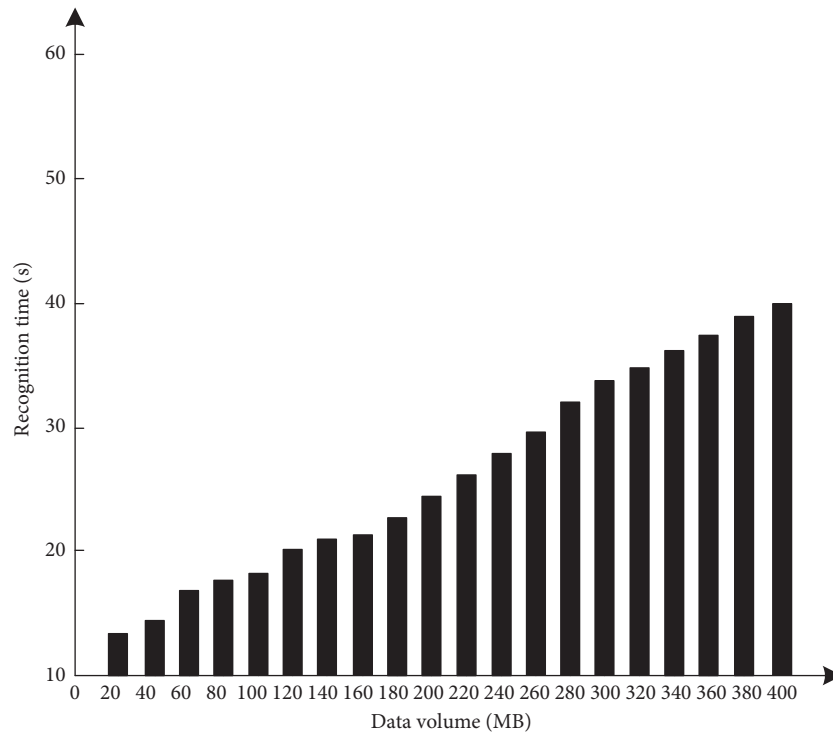


FIGURE 4: The recognition time of human action in dance video image by the proposed method.

in the AIST++ dance video dataset. When the data amount in the AIST++ dance video dataset is 400 MB, the recognition time of human action in dance video image by the proposed method is only 40 s. Therefore, the proposed method can effectively shorten the recognition time of human action in dance video image.

4. Conclusion

This paper studies the human action recognition technology in dance video image, detects the human action in the dance video image by extracting the features of human action, and finally realizes the human action recognition in the dance video image according to the principle of human action recognition in the dance video image. What is more, the research results show that the human action recognition technology designed in this paper has high recognition rate, and it can effectively shorten the recognition time, which further verifies the practicality of this technology [23–26].

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

- [1] H. Wu and Z. Cheng, “Action recognition algorithm based on complexity measure and multi-scale motion coding,” *Optical Technique*, vol. 44, no. 04, pp. 427–434, 2018.
- [2] H. Li, “Research on motion recognition method in dance video image,” *Video Engineering*, vol. 42, no. 07, pp. 34–37+52, 2018.
- [3] Z. Wu and Z. Zheng, “Motion recognition algorithm based on deep learning and motion information,” *Computer Engineering and Design*, vol. 39, no. 8, pp. 2668–2674, 2018.
- [4] Z. Guo, X. Cao, and Y. Hu, “Human motion recognition algorithm based on feature optimization and image similarity,” *Science Technology and Engineering*, vol. 19, no. 18, pp. 228–233, 2019.
- [5] H. Yu and Z. Min, “Human motion recognition based on improved CNN framework,” *Computer Engineering and Design*, vol. 40, no. 07, pp. 2071–2075, 2019.
- [6] W. Ding, K. Liu, F. Tang, and X. Fu, “Application of linear dynamic system inversion model in human behavior recognition,” *Journal of Image and Graphics*, vol. 24, no. 9, pp. 1450–1457, 2019.
- [7] T. Exner, C. A. Beretta, Q. Gao et al., “Lipid droplet quantification based on iterative image processing,” *Journal of Lipid Research*, vol. 60, no. 7, pp. 1333–1344, 2019.
- [8] F. Kokkinos and S. Lefkimiatis, “Iterative joint image demosaicking and denoising using a residual denoising network,” *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 4177–4188, 2019.
- [9] T. L. Gomes, R. Martins, J. Ferreira, R. Azevedo, G. Torres, and E. R. Nascimento, “A shape-aware retargeting approach to transfer human motion and appearance in monocular

- videos,” *International Journal of Computer Vision*, vol. 129, no. 2, pp. 2057–207, 2021.
- [10] P. B. Zhang and Y. S. Hung, “Articulated deformable structure approach to human motion segmentation and shape recovery from an image sequence,” *IET Computer Vision*, vol. 13, no. 3, pp. 267–276, 2019.
- [11] G. Singh, M. Chowdhary, A. Kumar, and R. Bahl, “A personalized classifier for human motion activities with semi-supervised learning,” *IEEE Transactions on Consumer Electronics*, vol. 66, no. 4, pp. 346–355, 2020.
- [12] S. Jiang, E. Chen, and M. Zheng, “Human motion recognition based on ResNeXt,” *Journal of Engineering Graphics*, vol. 41, no. 2, pp. 277–282, 2020.
- [13] A. Kushwaha, A. Khare, O. Prakash, and M. Khare, “Dense optical flow based background subtraction technique for object segmentation in moving camera environment,” *IET Image Processing*, vol. 14, no. 5, pp. 3393–3404, 2020.
- [14] Z. Lin, H. Qin, and S. C. Chan, “A new probabilistic representation of color image pixels and its applications,” *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 2037–2050, 2019.
- [15] H. Luo, K. Tong, and F. Kong, “The progress of human action recognition in videos based on deep learning: a review,” *Acta Electronica Sinica*, vol. 47, no. 5, pp. 1162–1173, 2019.
- [16] Y. Xinbo, H. Wei, L. Yanan et al., “Bayesian estimation of human impedance and motion intention for human-robot collaboration,” *IEEE transactions on cybernetics*, vol. 51, no. 4, pp. 1822–1834, 2019.
- [17] J. S. Park, C. Park, and D. Manocha, “I-planner: intention-aware motion planning using learning-based human motion prediction,” *The International Journal of Robotics Research*, vol. 38, no. 1, pp. 23–39, 2019.
- [18] Y. Liu, S. Qiu, and L. Sun, “Human motion recognition method based on multi-view self-paced learning,” *Computer Engineering*, vol. 44, no. 2, pp. 257–263, 2018.
- [19] D. Hu, H. Ke, and W. Zhang, “Research on human body attitude recognition based on kinect and ROS,” *Chinese High Technology Letters*, vol. 30, no. 2, pp. 177–184, 2020.
- [20] S. Chen, W. Wei, B. he, S. Chen, and J. Li, “Action recognition based on improved deep convolutional neural network,” *Application Research of Computers*, vol. 36, no. 3, pp. 945–949+953, 2019.
- [21] X. Shen, S. Yu, and Y. Dong, “Human motion recognition method based on deep learning,” *Computer Engineering and Design*, vol. 41, no. 4, pp. 1153–1157, 2020.
- [22] R. Zhang, Q. Li, and J. Chu, “Human action recognition algorithm based on 3D convolutional neural network,” *Computer Engineering*, vol. 45, no. 1, pp. 259–263, 2019.
- [23] Y. Zhu, X. Huang, and J. Huang, “Research on human motion recognition based on 3D CNN,” *Modern electronic technology*, vol. 43, no. 18, pp. 150–152+156, 2020.
- [24] B. He, W. Wei, and B. Zhang, “Improved deep convolutional neural network for human action recognition,” *Application Research of Computers*, vol. 36, no. 10, pp. 3107–3111, 2019.
- [25] S. Liu, X. Bai, M. Fang, L. Li, and C. C. Hung, “Mixed graph convolution and residual transformation network for skeleton-based action recognition,” *Applied Intelligence*, pp. 1–12, 2021.
- [26] X. Ma, K. Uu Rbil, and X. Wu, “Denoise magnitude diffusion magnetic resonance images via variance-stabilizing transformation and optimal singular-value manipulation[J],” *NeuroImage*, vol. 215, Article ID 116852, 2020.