Dance Movement Recognition Based on Gesture

Ping Lei[1,a], Nana LI[2,b,*], Haidong Liu[1,c]

[1] *School of Physical Education, Chengdu Normal University, Chengdu, 611130, Sichuan, China*
[2] *School of Physical Education, Henan Polytechnic University, Jiaozuo, 454003, Henan, China*
[a]*zumba2021@126.com,* [b]*gzyyxlnn@126.com,* [c]*tgzyyxlhd@126.com*

Abstract

Aiming at the low accuracy of traditional dance movement recognition methods, a movement recognition algorithm based on human posture estimation is proposed. Firstly, PAFs algorithm is adopted to recognize the spatial skeleton nodes of the human body model and the connection of human body joints, thus the human movement skeleton is obtained. According to the movement skeleton, the human body posture can be estimated. After the posture information is preprocessed and features are extracted, LSTM time series algorithm is used to classify and recognize the dance movements. The results show that the algorithm can clearly identify the dance movement skeleton nodes. For different movement categories, the recognition accuracy and recall rate of different movement categories are above 85%, and the recognition accuracy of curtsey movement is up to 95.2%. It can be seen that the recognition accuracy of this algorithm is significantly improved and different dance movement categories can be accurately recognized.

Keywords: *PAFs algorithm; human pose estimation; LSTM time series*

1. INTRODUCTION

With the support of information technology, it is possible to use computer vision technology to recognize human movements. This technical research result has played an important role in many fields. With the development of computer vision technology, people have developed video-based human motion recognition technology [4]. That is comprehensive application of image processing, recognition and classification technology to process the video data imported by users. Then, extract and analyze the feature information in the video data which can help us recognize the human movements presented in the video. Obviously, the video-based human action recognition process includes two main points. One is to extract feature information from video data, and the other is to identify and analyze the extracted feature information [5]. At present, human motion recognition technology has been applied in many fields, but so far there is no example of using human motion recognition technology to recognize dance movements. From an application point of view, the human body motion recognition technology can be used to process the recorded dance video and accurately recognize the dance moves in the video. Then compare it with the standard dance motion to find out the dancer's irregular motion and give us correct correction guidance. So the overall effect of dance teaching is improved. The posture-based dance action recognition technology studied in this paper first uses the human posture estimation algorithm to extract the dancer's spatial skeleton node in the dance video, and then uses the time sequence processing algorithm to identify the skeleton sequence on the time axis, which can be used for the entire dance Human movements are recognized.

2. OVERALL IDEA OF DANCE MOVEMENT RECOGNITION

The dance movement recognition algorithm is the core of the dance movement recognition technology. The algorithm includes two stages. The first stage is the human pose estimation process based on the human body pose estimation algorithm (PAFs), and the second stage is based on the long and short cycle memory neural network (LSTM) sequence classification process [1]. Of course, before starting the dance movement recognition algorithm, the input video needs to be extracted and cropped. And the output of the dance movement recognition algorithm is the action category contained in the video. The above ideas can be illustrated in Figure 1.

A. El-Hashash et al. (Eds.): IEIT 2022, ACSR 100, pp. 448–452, 2023.
https://doi.org/10.2991/978-94-6463-058-9_73

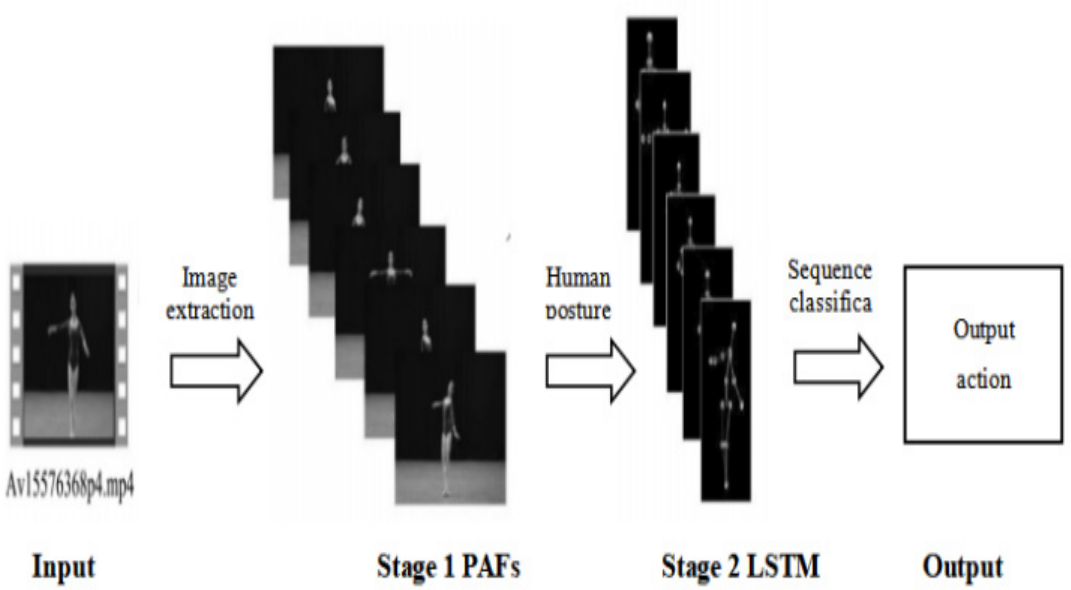


Figure 1. The execution process of the dance movement recognition algorithm

3. HUMAN POSE ESTIMATION THE PAFS ALGORITHM IS A BOTTOM-UP MODEL

PAFs algorithm is a down-top algorithm. It first identifies the spatial skeleton nodes of the human body model, and then links each node to obtain the human body action skeleton. The opposite of the PAFs algorithm is a top-down algorithm, that is, personnel detection is performed first, and then the joint points of each individual are identified. In contrast, the top-down algorithm will cause a delay in computing due to an increase in the number of people [6]. The PAFs algorithm effectively avoids such problems, and its parallel computing speed will not be affected by the increase in the number of people.

The idea of the PAFs algorithm is to pass the picture through the backbone network once and then the heat map for six times, thereby identifying all the individual joint points in the picture. Then link all the joint points to construct the human skeleton.

L(p) indicates the direction of the pixel on the human skeleton, and S(p) indicates the response of the key point. The backbone network VGG16 contains two branches, which are used for regression L(p) and regression S(p) respectively [3]. After the regression analysis is completed, the loss operation is performed, combining L, S and input to form a new input item, and returning to L, S again. The loss module uses the L2 norm to obtain the true values of L and S by reading the joint points in the annotation file, and then solve the loss value.

When performing S regression analysis, different types of joint points correspond to their unique channels, and the method of using multiple Gaussian distributions to take max is usually used to determine the true value of S.

When performing L regression analysis, the following formula (1) can be applied to calculate the PAFs on the c-th limb of the k-th person in the picture.

$$L_{ck}^{*}(P)=\begin{cases}\upsilon & if\ on\ \lim b\ c,k\\ o & \text{otherwise}\end{cases} \tag{1}$$

In the formula, $\upsilon=\frac{x_{j2},k-x_{j1},k}{\left\|x_{j2},k-x_{j1},k\right\|_2}$, $x_{j,k}$ refers to the position of the j-th key point of the k-th person [2]. A reasonable threshold is set to check whether the pixel point p falls on the human limb, which corresponds to equation (2):

$$0\le \upsilon.(p-x_{j1,k})\le L_{c,k}\ and\left|\upsilon_1\bullet(p-x_{j1,k})\right|\le\sigma_1 \tag{2}$$

In the formula (2), σ_1 refers to the width of the limb and $L_{c,k}$ refers to the length of the limb.

Perform homogenization of the limbs of the same category of different individuals in the picture to ensure that the output channel of L is consistent with the number of limbs, corresponding to equation (3):

$$L_c^{*}(p)=\frac{1}{n_c(p)}\sum{}_k L_c^{*}(p) \tag{3}$$

After establishing the heat map and determining the location of the key points x_j, apply the following point

integral formula (4) to estimate the correlation between the key points.

$$E = \int_{u=0}^{u=1} L_c(p(u)) \bullet \frac{x_{j2} - x_{j_1}}{\left\| x_{j2} - x_{j_1} \right\|_2} du \tag{4}$$

In formula (4), $p(u) = (1-u)d_{j1} + ud_{j2}$.

The human body pose estimation phase completes tasks such as extracting key nodes and calculating the weight of each node connection, which prepares for the subsequent time series recognition phase [7]. The essence of the second phase is to solve the graph optimization problem, that is, to use the Hungarian algorithm to achieve the optimal matching of adjacent nodes. So that we can obtain a humanoid posture skeleton consistent with the actual situation.

4. DANCE ACTION RECOGNITION BASED ON TIME SERIES

The recurrent neural network model (RNN) has advantages in processing sequence data. But when the model uses the BPTT algorithm to update the weights of each layer of the network, in addition to the input samples, it will also input the output of the previous round [8]. This sharing mechanism can exert beneficial effects in a short period of time, but over time it will consume a lot of memory and even cause a "gradient explosion." The LSTM algorithm belongs to the category of the RNN model, but it is different from the traditional RNN model by virtue of its forgetting function [9]. The LSTM model contains 1 tan layer module chain and 4 layer module chains. Among them, the 4 layer module chains act as the memory unit and gate structure of the LSTM model to control the addition (memory) or deletion (forgetting) of information. Therefore, this research applies LSTM to dance recognition.

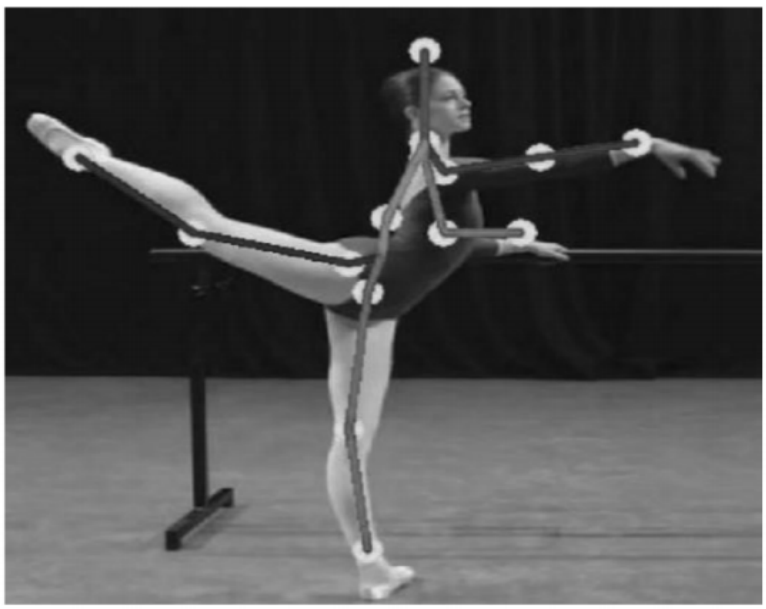
5. TEST AND ANALYSIS

5.1. Test Data Set

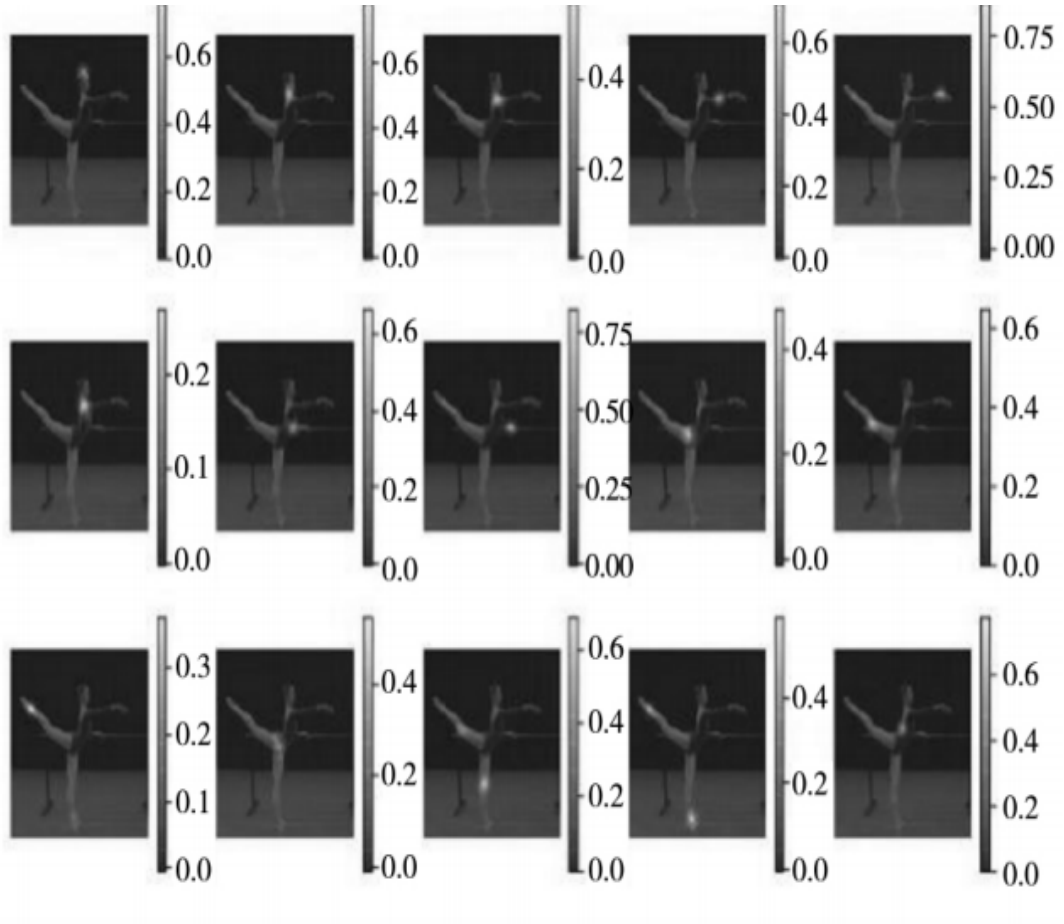
In this experiment, relevant dance video materials were collected from the Balletto dance video database to verify the dance action recognition algorithm. The video material collected in this experiment contains 10 ballet dance moves, such as bray dance step, big kick, circle, fast rotation, arabesque, toe rags, whip circle, curtsey, hand and foot coordination, cat jumping.

5.2. Effect Analysis of Dance Action Recognition Algorithm

After preparing sufficient video material, the application effect of the PAFs algorithm is tested first. The test results are shown in Figure 2(a) below. It showed that the PAFs algorithm has identified 15 joint points on the dancer, but the recognized joint points of the head and waist are deviated from the actual situation. This reveals that the PAFs algorithm's estimation ability for limb joints with large scale changes still needs improvement. In addition, according to the analysis of the human body posture estimation heat map 2(b), the PAFs algorithm outputs the heat maps of the joint points of the two legs at the same time. This is mainly because the large-scale dance moves are quite different from the normal human body shape. It causes interference when the PAFs algorithm returns to the joint points. On the whole, the PAFs algorithm can basically estimate the dancer's skeleton node more accurately.



(a)Skeleton diagram of human pose estimation



(b)Thermal map of human posture estimation

Figure 2. Skeleton diagram of human pose estimation and thermal map

The dance video is extracted from the frame by frame, thereby building a sequence of pictures, which is used as the input of the PAFs algorithm. We perform pose estimation on the extracted pictures to generate a sequence of spatial skeleton nodes. Obviously, the performance of the PAFs algorithm will have a direct impact on the spatial skeleton node sequence, and then affect the application effect of the LSTM algorithm.

Table 1. Accuracy table of dance action recognition algorithm

Action category	Accuracy rate/%	Recall rate/%
Hands and feet	90.0	90
Big kick	85.7	90
Arabesque	90.0	85
Curtsy	95.2	95
Toe step	90.9	85
Cat jumping	86.4	85
Whip in circles	95.0	80
Fast spin	85.0	90
Go round	94.7	75
Bray steps	90.9	85

According to the analysis in Table 1 above, the LSTM algorithm has a recognition accuracy of 95.2% for curtsy, which ranks first among all dance movements. And the LSTM algorithm has an 85% recognition accuracy for fast rotation, which is the last among all dance movements. In addition, The LSTM algorithm has the highest recall rate (95%) for curtsy and the lowest recall rate (75%) for turns. After statistical calculation, the accuracy and recall rate of the LSTM algorithm for identifying 10 ballet movements are 90.4% and 86.0% respectively. It can be seen that the accuracy of the LSTM classifier in identifying ballet movements is in line with expectations.

6. CONCLUSION

The pose-based dance action recognition method proposed in this paper combines the PAFs algorithm and the LSTM algorithm. First, the dance video input by the user is cropped into pictures according to frames, and then the body pose estimation is performed using the PAFs algorithm to generate a sequence of spatial skeleton nodes. It is substituted into the LSTM algorithm as an input item to predict the human movement in the dance video. According to the performance test results, the dance movement

recognition method proposed in this paper can more accurately identify and predict dance video movements, and can play a role in assisting dance teaching in reality. However, the dance action recognition method proposed in this article has a low recognition accuracy for human joint points with large scale changes, which is also an important direction for follow-up research.

REFERENCES

[1] Deng Lian (2020). An immersive virtual display application integrating interactive projection technology. Science and Technology Innovation, vol. 35, pp: 101-104.

[2] Johnson Zachary V.,Arrojwala Manu Tej Sharma,Aljapur Vineeth,Lee Tyrone,Lancaster Tucker J.,Lowder Mark C.,Gu Karen,Stockert Joseph I.,Lecesne Rachel L.,Moorman Jean M.,Streelman Jeffrey T.,McGrath Patrick T. (2020). Automated measurement of long-term bower behaviors in Lake Malawi cichlids using depth sensing and action recognition. Scientific Reports, no.1.

[3] Kong Xiangkui, Xiang Hua (2020). Research on movement correction of laser sensor sensing information. Laser Journal, vol. 41, no. 11, pp: 179-182.

[4] Qian Huifang, Yi Jianping, Fu Yunhu (2020). Overview of human action recognition based on deep learning. Computer Science and Exploration, vol.12, no.6, pp: 1-20.

[5] Stergiou Alexandros (2021). Poppe Ronald. Learn to cycle: Time-consistent feature discovery for action recognition. Pattern Recognition Letters, vol.141.

[6] Zhou Hongyu, Yan Chunfeng, Song Xu, Liu Guoying (2020). Action recognition algorithm based on weighted three-view motion history image and time series segmentation. Journal of Electronic Measurement and Instrument, vol. 12, no.16.

[7] Zhao Qilu,Dong Junyu. Self-supervised representation learning by predicting visual permutations (2020). Knowledge-Based Systems, vol.210.

[8] Zhang Hengxin, Ye Yingshi, Cai Xianzi, Wei Fuyi (2020). Efficient motion recognition algorithm based on human joint points. Computer Engineering and Design, vol. 41, no. 11, pp: 3168-3174.

[9] Zhao Xihua (2020). Automatic scoring system for aerobics combining Bayesian optimal classifier and target graph monitoring. Journal of Shijiazhuang University, vol. 22, no.6, pp: 111-116.