Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 4687465, 11 pages
https://doi.org/10.1155/2023/4687465

Research Article

Human Motion Recognition in Dance Video Images Based on Attitude Estimation

Nian Li[1] **and Shekhar Boers**[2]

[1] *Lu Xun Academy of Arts, Yan'an University, Yan'an Shaanxi 716000, China*
[2] *The King's School, BP1560, Bujumbura, Burundi*

Correspondence should be addressed to Nian Li; linian5699@sina.com

Received 10 June 2022; Revised 22 July 2022; Accepted 26 July 2022; Published 11 May 2023

Academic Editor: Kalidoss Rajakani

With the deep integration of science and technology and culture, the estimation of human movements in dance video images will become an important application field of computer vision technology, which can be used not only for professional dancers' movement correction, dance self-help teaching, and other application scenarios but also for athletes' movement analysis. Therefore, it will greatly promote the implementation of teaching students in accordance with their aptitude by applying information technology to estimate dancers' movements and postures in real time and obtaining information of classroom dance teaching status in time. In this paper, human motion recognition in dance video images is studied based on an attitude estimation algorithm. When the number of experiments reaches 20, the average value of deep learning algorithm and particle swarm optimization algorithm is 76.23 and 75.23, respectively, while the average value of attitude estimation algorithm in this paper is 77.95. Therefore, the average results of attitude estimation algorithm in this paper are slightly higher than those of other algorithms.

1. Introduction

Dance is one of the important manifestations of culture. The number of Chinese dance classes is usually large. Teachers can only roughly obtain students' movement changes and emotional changes through students' body movements and facial expressions. It is difficult to accurately understand students' real-time mastery of dance movements [1]. As an art form, dance has a far-reaching impact on our life. For example, we can use dance to express our feelings and communicate with others. However, learning dance is not an easy thing. With the development of the deep integration of science and technology and culture, the estimation of human body movements in dance video images will become an important application field of computer vision technology. It can be used not only for professional dancers' movement correction, dance self-help teaching, and other application scenarios but also for athletes' movement analysis. In terms of the massive video data generated, how to obtain the valuable information and realize the understanding of the content has become a concern. Most videos depict human behavior [2, 3]. Human motion recognition in videos has always been a research hotspot in the field of computer vision. However, there is no ready-made standard for the fine-grained division of human actions. Therefore, in this paper, we divide the analysis of human action from coarse to fine according to the process of cognition. Specifically, we divide the cognitive process of human action into three levels: the first level is the discrimination of coarse-grained categories. The second level is to identify the subclasses under the coarse category. The third cognitive level is to analyze the specific action composition under the subcategory, which is a more detailed analysis [4].

Therefore, it will greatly promote the implementation of teaching students in accordance with their aptitude by applying information technology to estimate dancers' movements and postures in real time and obtaining information of classroom dance teaching status in time. The human body is a complex deformed body, which is composed of several body parts, which are connected by joints. Therefore, the posture parameters of the human body mainly have two forms: one is the joint angle between each body part, and the other is the coordinate position of each body part in a three-dimensional space [5, 6]. Human posture estimation is a hot and difficult point in the field of computer vision, and it has a wide application prospect in intelligent monitoring, advanced human-computer interaction, image and video retrieval, virtual reality, motion analysis, and other fields. In order to fully consider the influence of occlusion on location detection and human structure, a modeling method of occlusion location based on occlusion level is proposed, and human body location detectors under various occlusion conditions are established and the relationship between adjacent locations is described [7]. Because it depends on the target detection algorithm and the single pose estimation algorithm with good performance, the accuracy of human pose estimation is high. However, the performance of this kind of method is seriously affected by the quality of the target detection frame, and even the most advanced target detector will have detection errors, resulting in redundancy, missed detection, and false detection of the human detection frame [8].

Through the pose estimation algorithm, the human motion recognition technology is applied to the dance video image motion recognition, accurately identifying the dance motion, comparing it with the standard motion, and identifying and correcting the dancer's nonstandard motion, which is a new method to assist dance teaching. The bottom-up method is mainly divided into two parts: joint point detection and joint point clustering. It uses the single person pose estimation algorithm to detect all joint points in the dance video image and then clusters the joint points of different human bodies to aggregate the joint points belonging to the same human body to realize multi person pose estimation [9, 10]. Through the fusion of high-level and low-level multiscale features, improve the robustness of pose estimation to scale changes, analyze the geometric relationship between human bone joint points through pose estimation algorithm, design a hierarchical pose estimation model based on the geometric relationship of joint points, carry out multilevel joint point estimation, and improve the accurate estimation of the position of dancers' joint points. Finally, the effectiveness of the proposed pose estimation algorithm is verified on public data sets and self-built dance data sets. Human motion recognition in dance video images based on pose estimation is an interdisciplinary research topic, which involves the knowledge of artificial intelligence, machine learning, and other cutting-edge disciplines. It not only has important academic research value in theory but also has broad application prospects in the fields of intelligent video surveillance, human-computer interaction, video retrieval, and auxiliary monitoring [11].

The following innovations are put forward in this paper:

(1) This paper describes the algorithm flow of human posture estimation in detail. The pose estimation algorithm of human body is easily interfered by the background noise of the image. In order to reduce the influence of the background and improve the accuracy of the part recognition, we usually first extract the parts of human body by image segmentation. The human body image can be divided into several image blocks by part segmentation, and each block usually has similar color or texture features

(2) The MRNet network is proposed. In art works, tea is the same as the core idea of Confucianism. Many Chinese painters are influenced by Confucianism in tea culture. Many of them are ambitious, but they have no choice but to have a bad career. Their patriotism and passion have nowhere to be released. Finally, they choose to "be independent" in tea products

The overall structure of this paper consists of five parts. Section 1 introduces the background and significance of human motion recognition in dance video images. Section 2 mainly describes the literature review and the research work of this paper. In Section 3, the principle and algorithm of human posture estimation are further discussed. In Section 4, the experiment is carried out and the results are analyzed. Section 5 is a summary of the full text.

2. Journals Reviewed

Yang and Lyu classified and analyzed the existing human motion recognition methods in dance video images and comprehensively described the steps and details in the process of motion recognition [12]. Swain et al. proposed to summarize human motion analysis from two aspects: model establishment and motion estimation. In addition, a variety of model free motion analysis methods are also discussed [13]. Zhang proposed a variant version of Mei, which characterizes human actions in dance video images through three-dimensional contours in space-time volume. However, the motion representation based on the human body contour in the dance video image is more sensitive to occlusion and angle change [14]. Iqbal and Sidhu reviewed the development level and common processing methods of human motion analysis from four aspects: motion detection, moving target classification, pedestrian tracking, and behavior understanding and description. Because this paper is mainly aimed at video-based motion analysis, the attitude estimation problem is included in the tracking part [15]. Barton a suggested that different students learn through different dance video image learning methods. Most students prefer a specific way to receive, process, and interact with information. The learning style that students prefer in human motion recognition in dance video images is the so-called learning style [16]. Polechoński et al. proposed that in general, the recognition of human actions in dance video images mainly includes three modes: segmented action mode recognition, continuous action mode recognition, and real-time action data stream recognition [17]. Sparacino et al. proposed and summarized more than 150 papers related to human motion analysis published before 2021. This paper describes the motion analysis based on computer vision from four aspects, including initialization, tracking, pose estimation, and motion recognition. The direct and indirect models are divided into three categories [18]. Ardizzone and Celebi proposed that the type of pattern and the beginning and end frames in these dance video images are unknown. Therefore, for human motion recognition in continuous dance video images, it is not the same as the recognition of segmented motion patterns; only the unknown motion is directly matched with the training data [19]. Luo and Ning introduced the idea of time template into the field of human motion recognition in dance video images. They integrated the human contour changing with time and proposed the human motion energy map [20]. Liu proposed a variety of methods and main steps of attitude estimation based on component structure, which has attracted much attention in recent years. The reason of using component-based method is given, and three processes of attitude estimation are described in detail: representation method, reasoning, and training. Finally, three applications of component-based method are described: pedestrian detection, pose estimation, and tracking [21].

In view of the content of the above scholars' research on human motion recognition in dance video images, this paper puts forward a pose estimation algorithm to study human motion recognition in dance video images. The pose estimation means that the pose parameters of each part of the human body are inferred from the input image sequence, and after the weights of key nodes and the connections of each point are obtained, the pose estimation of the human body becomes a graph optimization problem. Compared with the top-down algorithm, this algorithm first detects people and then returns to each person's joint points, which will not increase the calculation time due to the increase of the number of people in the picture. Because of the greater freedom of dance movements, compared with the natural movement of the human body, dance movements change more. Conventional movements include diving, twisting, kicking, crouching, bending, lying, and stretching, while more complex movements include swinging, sliding, walking, flying, spinning, and rolling over. These movements contain many degrees of freedom, and simple modeling methods and movement recognition methods are difficult to accurately express dancers' human movements. Therefore, using attitude estimation algorithm and motion recognition are two challenging research contents in human behavior analysis. Its main task is to make the computer automatically perceive where people are in the scene and judge what people are doing. The posture estimation algorithm is used to optimally match the adjacent nodes. For example, when calculating the connection between the head node and the neck node, the optimal matching is performed according to the correlation between the head node and other nodes in the graph, so as to obtain the connection between the head and the neck. Exploring the mechanism of brain visual information analysis and processing from the perspective of action can further deepen the mastery and understanding of human visual system and also provide new ideas and methods for further exploring human perception and psychological activities.

3. Algorithm Model

3.1. Principle and Algorithm of Human Pose Estimation. Human pose estimation refers to the process of obtaining the pose parameters such as the position or joint angle of each part of the human body in each frame image in a given video. It is a research hotspot in the field of computer vision. In addition, the existing human posture estimation methods are mainly aimed at traditional data sets, such as MS COCO, MPII, and LSP, which include simple human posture, such as standing, and walking. However, there are complex and changeable dance movements, strong coherence, and serious occlusion in dance pose estimation. There are many interference factors such as illumination change and camera angle change in dance classroom scenes, which greatly increases the difficulty of dance pose estimation [22]. The acquisition of posture parameters can provide support for reconstructing human motion, assists in realizing computer perception of where people are and analyzing what people are doing in the scene, and is widely used in motion recognition, human-computer interaction, video understanding, and other fields. In order to fully consider the overall information of human image and improve the robustness of super-pixel segmentation, the relationship function is calculated through the distance between super pixels. Compared with adjacent super pixels with different categories, adjacent super pixels with the same category have more similar color features and closer geometric distance. Firstly, the expression method of color similarity is as follows:

$$s_c\left(S_i, S_j\right) = \sum_{k=1}^{n} \min\left(Y_I^K, Y_j^k\right). \tag{1}$$

Among them, the color similarity $s_c(S_i, S_j)$ is measured by the intersection of the color histograms Y_i^k and Y_j^k of S_i and S_j, respectively.

The intersection of HOG histograms of two superpixel blocks is expressed as

$$s_t\left(S_i, S_j\right) = \sum_{k=1}^{n} \left(t_i^k, t_j^k\right). \tag{2}$$

The gradient similarity $s_t(S_i, S_j)$ is represented by the intersection of HOG histograms t_i^k and t_j^k.

The scale is represented by the ratio of two superpixels in the image:

$$s_s\left(S_i, S_j\right) = 1 - \frac{\text{size}(S_i) + \text{size}\left(S_j\right)}{\text{size}(I)}, \tag{3}$$

where $s_s(S_i, S_j)$ is the scale information of superpixels and size (S_i), size(S_j), and size(I) represent the superpixel blocks S_i, S_j, and the number of pixels in the whole image, respectively.

Human posture estimation is easily interfered by image background noise. In order to reduce the influence of background and improve the accuracy of part recognition, various parts of human body are usually extracted by image segmentation. The human body image can be divided into several image blocks by part segmentation, and each block usually has similar color or texture features. Each part of a person in an image usually has similar color and texture features, while the features of human body parts in different images are quite different, so it is difficult to use a unified model to represent the same parts in different human body images [23]. Therefore, obtaining the prior information of the human body in the image and the prior color distribution of the parts can correctly guide the image segmentation process. The dance movement recognition algorithm in this paper is a two-stage model. After image extraction and clipping, the video enters the first stage, which is mainly composed of human posture estimation algorithm. The second stage is composed of long and short period memory neural network. Finally, that action category in the video image is output, as shown in Figure 1.

Each joint part of human body is divided into several square windows, one window corresponds to one part, and the scores of all parts are obtained by the following formula:

$$P(I,p)=\sum_{i\in V} w_i^{t_i}\cdot\varphi(I,p_i), \tag{4}$$

where $\phi(I,p_i)$ is the characteristic function of the i part at the position p_i in the image I and $w_i^{t_i}$ is the parameter obtained when the i part belongs to t_i type.

The deformable model is used to describe the correlation degree, and the calculation method is shown in the formula.

$$D(I,p)=\sum_{ij\in E} w_{ij}^{t_i,t_j}, \tag{5}$$

where $w_{ij}^{t_i,t_j}$ is the parameter taken when i and j of adjacent parts belong to t_i and t_j types, respectively.

In order to more accurately obtain the direction of each part of the human body, quantify the direction relationship of adjacent parts, and represent the distribution of child nodes relative to parent nodes, a consistency model is introduced, and the formula is as follows:

$$C(t)=\sum_{i\in V} b_i^{t_i}, \tag{6}$$

where $b_i^{t_i}$ represents the parameters obtained when the i part belongs to t_i type and $b_{ij}^{t_i,t_j}$ represents the parameters obtained when the adjacent parts i and j belong to t_i and t_j type, respectively.

This paper adopts PAFS human posture estimation algorithm. PAFS first returns to the joint points of the human model and then links the joint points to obtain the skeleton, which belongs to the bottom-up algorithm. Good achievements have been made in the estimation of human posture in video. However, due to the diversity of human posture changes and the changes of human body shape, clothing, and viewing angle, it is difficult for the component model to capture all apparent changes. In addition, the model only introduces the consistency constraint between adjacent frames in time domain, without the constraint of long-term consistency, which is prone to the error accumulation of component state estimation [24, 25]. Because the pose estimation task is a pixel level joint point estimation problem, it needs to use low-level and high-level features to locate joint points of different scales. High-level features are conducive to the location of large-scale joint points, while low-level features are very important to the location of small-scale joint points. Aiming at the problem of drastic scale change of bone joints in dance, this paper constructs a sequential multiscale feature fusion model to improve the robustness of pose estimation to scale change.

3.2. MRNet Network. The task of human posture estimation needs the feature information contained in feature maps with different resolutions. MRNet uses mixed hole convolution and weighted way of two-way feature pyramid to retain more low-resolution feature map information, which makes the result of human posture estimation more accurate. In order to solve the high-dimensional problem of human posture space and taking into account a large number of constraints in human motion, a manifold learning method with time neighborhood preserving embedding of MRNet network is proposed to discover the essential distribution structure of three-dimensional human motion in low-dimensional space. MRNet in the process of information exchange from high-resolution to low-resolution characteristic map and mixed cavity convolution is used instead of ordinary convolution downsampling. Using ordinary convolution for downsampling will result in the loss of internal data structure and spatial hierarchical information. The mixed hole convolution method can overcome this problem, improve the receptive field, and avoid the grid effect caused by the traditional hole convolution method. MRNet network can extract the multiresolution features of the input image, and it has a strong feature representation ability, which can achieve good results in target detection, recognition, image segmentation, and joint point estimation of human body. MRNet will generate a low-resolution subnet in the process of deepening the network, and after the subnet is generated, connect the feature graphs with different resolutions to exchange information. Net generates backbone network in a similar way to MRNet, and the network structure is shown in Figure 2.

In the process of human posture joint point estimation, HRNet network does not make full use of its extracted multiresolution features, only uses the high-resolution features for joint point heat map estimation, and discards other medium and low resolution features, resulting in the loss of information in the process of feature representation and affecting the accuracy of joint point estimation. According to the estimated values of the brightness P of the potential target and the brightness q of the target template, the discrete estimated value of the distance between the potential target with the center position of y and the target template can be obtained according to the following formula:

$$\rho(y)=\rho\left[\widehat{p}(y),\widehat{q}_u\right] \tag{7}$$

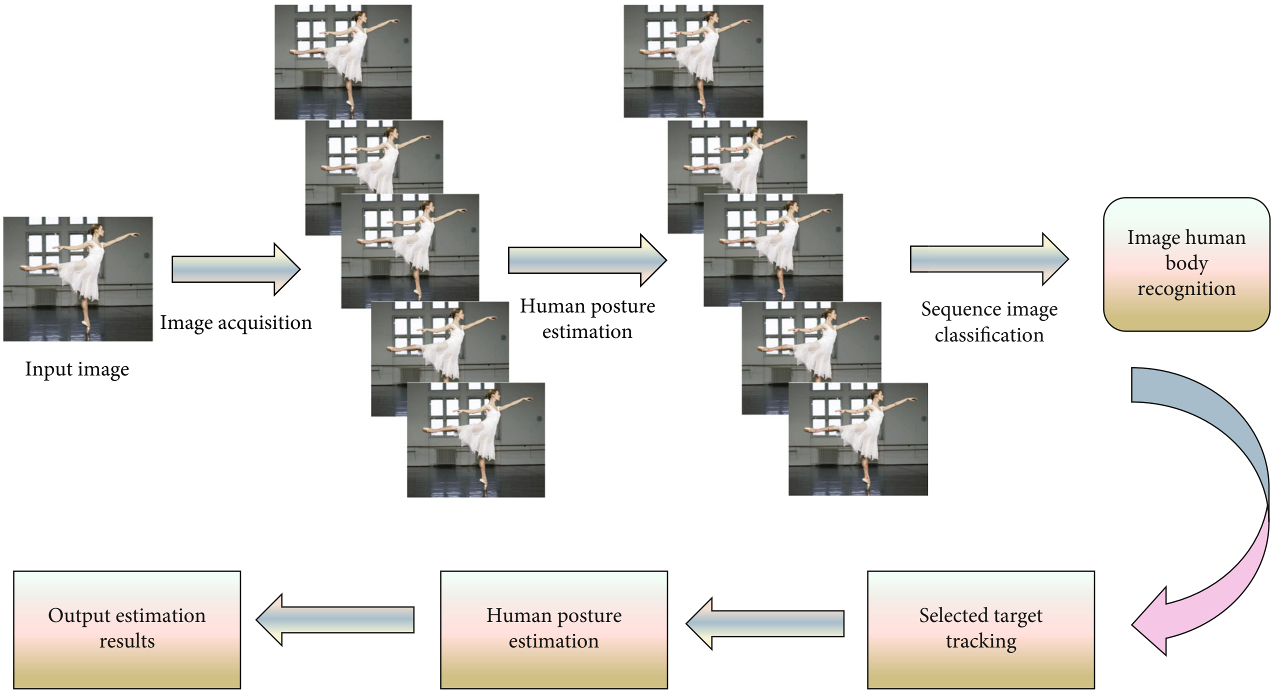

Figure 1: Flow chart of human posture estimation algorithm.

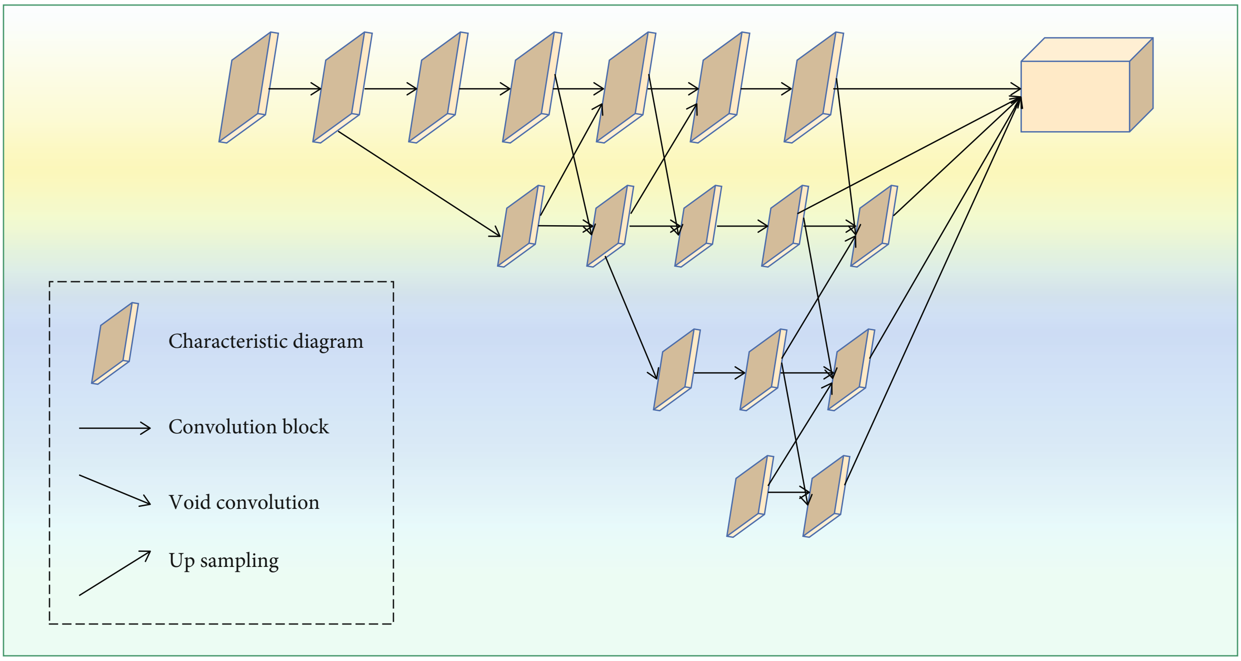

Figure 2: MRNet network structure.

The feature vector u represents the color of the target, q_u represents the probability distribution of the target template, and the gap between the two distributions is defined as

$$d(y) = \sqrt{1 - \rho[\widehat{p}(y), \widehat{q}]} \quad (8)$$

The statistic d is optimal and insensitive to the change of target scale, so it is very effective for random distribution density. d distance and Fisher linear discriminant are better.

Because the tracking target is affected by the background and some occlusion, the possibility of surrounding pixels is low. Therefore, the farther away from the center, the smaller the weight assigned to the pixel. This processing can improve the robustness of the estimated value. If x and y are standard coordinates normalized by h_x and h_y, the target probability density is

$$\hat{q} = c\sum_{i=1}^{n} k, \tag{9}$$

where c is the normalization constant, i.e.,

$$c = \frac{1}{\sum_{i=1}^{n} k\left(\|x_i^*\|^2\right)}. \tag{10}$$

MRNet network has memory function. It is a kind of neural network specialized in processing sequence data. However, when the traditional MRNet model uses the human posture estimation algorithm to update the weight of each layer of the network, MRNet will have two inputs at each time. While inputting samples, the output of the previous time will be used as input at each time.

3.3. Dance Video Image Motion Preprocessing Technology. Dance video image action recognition is mainly carried out through computer vision system, but considering the problem of computer computation, it should be preprocessed before computer recognition to extract effective image information and reduce the computational burden of computer. The posture of human body in the image is very different; even the posture of the same human body in different images is also different. This has a significant impact on the research of attitude estimation. Multiperson pose estimation needs to consider the acquisition of multiple targets in the same image, which is easy to affect each other. Starting from a frame of standard action video and learner video, the human key point information extracted from standard action and learner action images is globally aligned. When correcting actions, we should also start from the continuity of actions. For learners, some actions are too slow or too early, which also need to be corrected. Compare multiframe actions in the network to analyze whether the action timing is correct. Firstly, the pose of each frame in the video is detected. Then, the global motion information is introduced to spread the optimal pose detection results in each frame image to the whole video, and generate a trajectory for each human part to form the original state space. The state space is filtered layer by layer by the stacked trajectory fragment model and gradually reduced. Finally, the optimal state of the component is obtained through the reasoning of the lowest model. Human action recognition in video is a hot spot in computer research at present. Its purpose is to extract and analyze the actions in video through various image processing, recognition and classification technologies, so as to judge the actions of people in video, so as to obtain useful information, which has a very wide range of uses. Most of the images in dance video are color. If they are not converted and directly input into the computer vision system, the amount of information input will be increased, resulting in the increase of subsequent calculation. Because there are many changes in local posture, it is difficult to model the human body with diversified posture. The third method is the PS model method based on image analysis. In the PS model method based on image analysis, the image analysis method is used to extract the human image features, and the human pose estimation is completed according to a large number of image semantics and PS model fusion.

The application of human motion recognition technology to dance video can effectively recognize dance movements and postures. By comparing video movements with standard movements, dancers' dance postures can be evaluated and suggestions for modification can be given. It is an advanced auxiliary training method. According to the analysis of the research status of human posture estimation, there are some difficult problems in the process of acquiring posture, such as the interference of background clothes, the estimation of various human postures, the occlusion of two-dimensional images, and the optimization of posture reasoning. Grayscale processing of color images in video can reduce the subsequent computation and improve the efficiency of motion recognition in computer vision system. It makes full use of all kinds of semantic features of human images and has strong robustness to some interference factors, such as clothes. However, some additional image semantic analysis is needed, which increases the complexity of human posture reasoning. After getting the binary image of the current moment of the action video, it is necessary to separate the moving region from the scene, which involves the segmentation of the moving region. The binary image processing function of the software can find the edge of the moving human body contour by setting appropriate threshold. The dancers' posture changes greatly, the dancers' movements are complex and changeable, and the camera angle changes, which increase the difficulty of posture estimation. In order to better analyze the network architecture designed in this paper and reveal the performance of each part of the network, each part of the algorithm is separated from the whole algorithm, and the influence of each module on attitude estimation is analyzed by comparing with the algorithm without each module. When thresholding a dance video image, it is very important to determine the appropriate threshold, because it is related to the location of the pixels in the image. Only when the threshold is reasonable, can a more accurate binary image be produced.

4. Experimental Results and Analysis

4.1. Standard Data Set. In this experiment, there are many standard data sets for two-dimensional human pose estimation in images. These data sets can be divided into whole body and half body data sets, as shown in Tables 1 and 2.

From Tables 1 to 2, it can be seen that the IP data set contains 304 whole body images of human body with marked joint positions, and the height of human body in the image is usually about 152 pixels. The IP data set has determined that the first 100 images are training samples and the last 206 images are test samples. The LSP data set contains 2000 human body images, of which the first 1000 are training samples and the remaining 1000 are test samples.

This section gives the running results of each algorithm on the data set IP. Set the number of super pixels to 100150200 according to the size of the test image. The comparison results

Table 1: Standard data set of whole body posture estimation.

Data set	Number of images	Training image	Test image	Image category
LSP	2,000	1,000	1,000	Sports
Sport	1,298	648	652	Sports
Fashion pose	7,304	6531	774	Fashion
J-HMDB	31,837	31,837	—	Every kind

Table 2: Standard data set of half body posture estimation.

Data set	Number of images	Training image	Test image	Image category
Buffy stickmen	747	471	275	TV play
We are family	524	351	174	Group photo
FLIC	7558	6,542	1,015	Feature film
Armlets	12588	9,592	2,995	PASCAL

between the attitude estimation algorithm and other algorithms are shown in Table 3.

It can be seen from Table 3 that the results of the data set IP show that the attitude estimation algorithm obtains the highest average PCP when 150 super pixels are set, and the PCP in the trunk and legs are higher than other algorithms. The main reason is that this method uses an additional 1000 images to complete the learning of the part model, and the part detector has high robustness to the arm part. This paper uses the specified 100 images in the IP data set as the training set, which has a certain disadvantage in the amount of training data.

4.2. Analysis of Experimental Results of Dance Movement Recognition. In this experiment, deep learning algorithm, particle swarm optimization algorithm, and posture estimation algorithm in this paper are used to study the mean change of human motion recognition in dance video images. This chapter is compared with other methods on LSP set, and the experimental results are shown in Figure 3.

As can be seen from Figure 3, when the number of experiments reaches 20, the average value of deep learning algorithm in human motion recognition in dance video image is 76.23, the average value of particle swarm optimization algorithm in human motion recognition in dance video image is 75.23, and the average value of posture estimation algorithm in this paper in human motion recognition in dance video image is 77.95. Therefore, the average results of the attitude estimation algorithms in this paper are slightly higher than those of other algorithms, which shows that the star model in this chapter is helpful to describe the distribution of human parts.

Based on the pose estimation algorithm, the occluded human body parts in the image are used as training data, and the experiments are carried out by using trunk, thigh, calf, and head. The features of the occluded parts are extracted, which to some extent interferes with the accuracy of the part detector. An experiment was conducted on 1000 test images on LSP, the standard data set of dance video images, and the results are shown in Figure 4.

It can be found from Figure 4 that the statistics of the data set LSP show that 47.1% of the dance video images have different degrees of occlusion on the human body. The head is rarely covered in all parts of the human body, so the recognition rate is high. The legs are relatively less blocked. The blocking of this part is mainly partial blocking, and the complete blocking is less, but the probability of being blocked is more than 12%. Although the trunk is covered more, the recognition result is less affected because the part is large. Because the arm is relatively small, it is greatly affected by occlusion, especially the occlusion rate of the lower arm reaches 36.4%, and the recognition result is also the worst. Therefore, occlusion brings great difficulty to solve the problem of human pose estimation in dance video images.

In this experiment, the depth learning algorithm, particle swarm optimization algorithm, and pose estimation algorithm are used to study the recognition results of human motion recognition in dance video images and occluded images in LSP data set. The recognition results of occluded images in LSP data set in this chapter are shown in Figure 5.

As can be seen from Figure 5, the accuracy of the attitude estimation algorithm in this paper is higher than that of deep learning algorithm and particle swarm optimization algorithm. This is mainly because this chapter specifically recognizes the occluded parts of human motion recognition in dance video images and better captures the apparent characteristics of the occluded parts. It is mainly to identify parts in different directions. It is difficult to obtain direction information when parts are blocked.

In this experiment, depth learning algorithm, particle swarm optimization algorithm, and pose estimation algorithm are used to study the recognition results of human motion recognition in dance video images and occluded images in LSP data set. On the data set IP of trunk, thigh, calf, upper arm, and lower arm, the results of this chapter's method and other pose estimation methods of the other two algorithms are compared as shown in Figure 6.

As can be seen from Figure 6, among the three different algorithms, the pose estimation algorithm obtains better

Table 3: Comparison of other algorithm results on data set IP with attitude estimation algorithm.

Method	Trunk	Tshankhigh	Shank	Upper arm	Lower arm	Average
Yang	82.8	69.1	63.8	55.2	35.3	61.24
Johnson	87.5	74.6	67.2	67.2	45.7	68.44
Attitude estimation algorithm (100 superpixels)	87.2	77.7	71.1	61.4	39.1	67.3
Attitude estimation algorithm (150 superpixels)	88.7	77.7	71.4	63.3	41.2	68.46
Attitude estimation algorithm (200 superpixels)	83.3	74.5	67.2	58.2	36.2	63.88

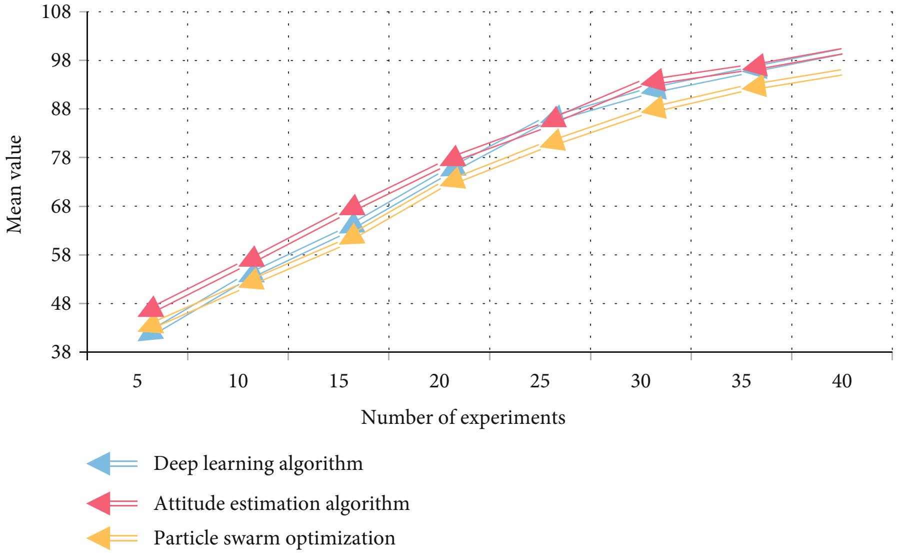

Figure 3: Average changes of different algorithms in human motion recognition in dance video images.

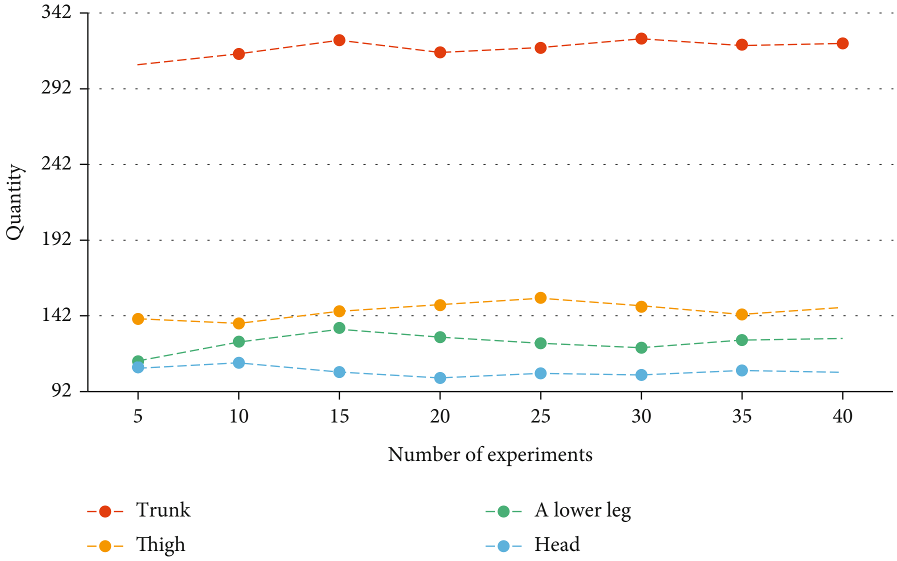

Figure 4: Variation curve of human body parts blocked.

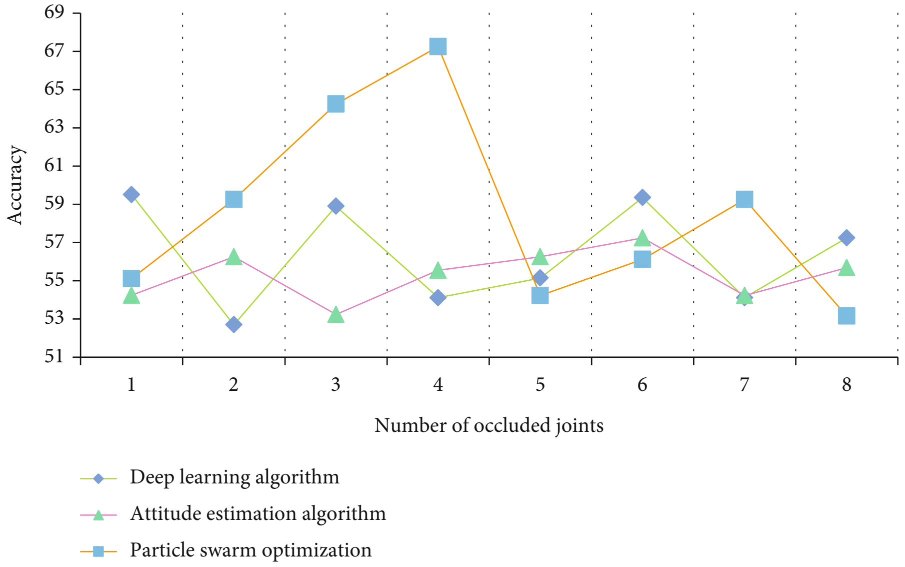

Figure 5: Recognition results of occluded images in LSP data set by different algorithms.

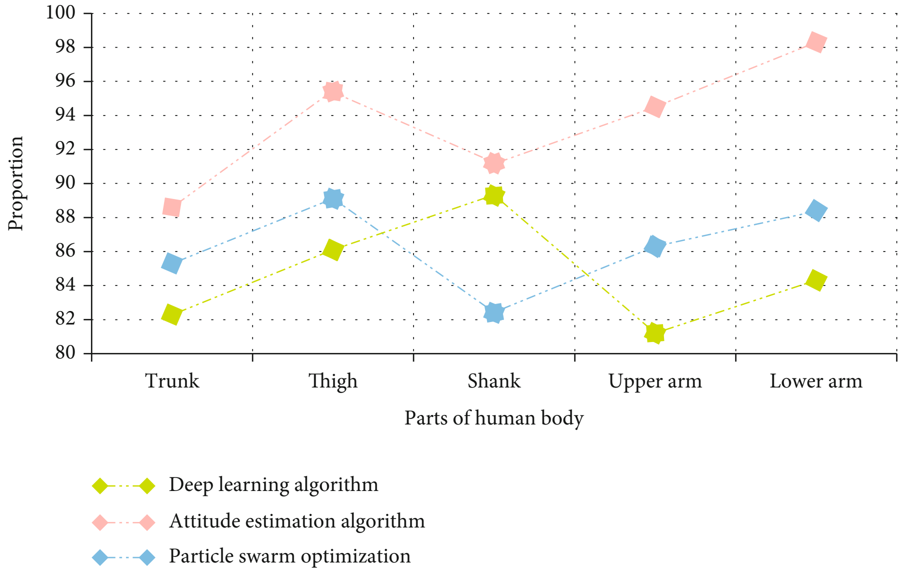

Figure 6: Comparison of results of different algorithms using strict PCP on data set IP.

results on the human motion recognition data set IP in the dance video image, because a large number of training data sets are used in addition to the better features obtained by the depth model. Deep pose takes a total of 12000 images in the LSP original and extended data set network as the training set and uses the model for the test of IP data set. The deep learning algorithm and particle swarm optimization algorithm only use 100 images specified in IP data set or 1000 images specified in LSP data set as training data, which are obviously at a disadvantage in the amount of human motion recognition training data in dance video images.

The deep learning algorithm, particle swarm optimization algorithm, and pose estimation algorithm in this paper are still used to compare the results of human motion recognition in dance video images. The results of strict PCP on the data set LSP are studied. The experimental results are shown in Figure 7.

As can be seen from Figure 7, the average result of the attitude estimation algorithm in this paper is higher than that of deep learning algorithm and particle swarm optimization algorithm. The recognition result of particle swarm optimization algorithm in lower leg is better than that of deep learning

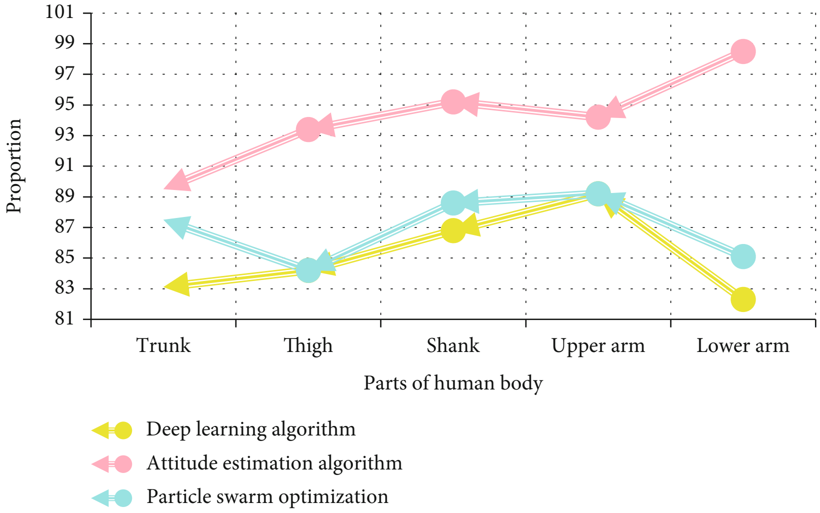

Figure 7: Comparison of results of different algorithms using strict PCP on data set LSP.

algorithm, especially the trunk recognition result is more than 4.6%. It shows that the occlusion relation model of this method can better extract the parts with relatively diverse positions. However, the recognition results of the methods in this chapter are higher than those of deep learning algorithm and particle swarm optimization algorithm, which shows that the occlusion level model can better obtain the occlusion information of relatively stable parts. The main reason is that the stable parts of human motion recognition in dance video images are highly correlated. The human body structure in this chapter is conducive to calculating the occlusion level of these adjacent parts.

5. Conclusions

Aiming at the diversity of human posture, a posture estimation algorithm is proposed to study the recognition of human motion in dance video images. When the number of experiments reaches 20, the average value of deep learning algorithm in recognition of human motion in dance video images is 76.23, and that of particle swarm optimization algorithm is 75.23. However, the average value of posture estimation algorithm in this paper is 77.95. Therefore, the average results of posture estimation algorithm in this paper are slightly higher than others. It can realize the accurate estimation effect of human body posture, and at the same time it plays an important role in the recognition and correction of human body movements in dance video images. Aiming at the problem that the pose estimation method in monocular images is easily interfered by occlusion, a pose estimation algorithm is proposed. Firstly, the occlusion level is defined as the degree of occlusion of human body parts, which is obtained by calculating the occlusion ratio and orientation of parts. Then, according to the occlusion level, a corresponding level of part detector is established for each part. It can be applied to dance teaching in combination with the requirements of practical application scenarios, realize real-time correction of dance movements, assist dancers in teaching and training, and have important significance for inheriting Chinese culture.

Data Availability

The figures and tables used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Shaanxi Provincial Department of Education’s special scientific research project “Research on Red Dance Creation in Yan’an in the New Era” (19JK0950). The authors would like to show sincere thanks for those techniques which have contributed to this research.

References

[1] N. Azman, K. Suzuki, T. Suzuki et al., “Effect of dance video game training on elderly's cognitive function,” *Transactions of Japanese Society for Medical and Biological Engineering*, vol. 55, pp. 526–529, 2017.

[2] A. Mejia-Downs, S. J. Fruth, A. Clifford et al., “A preliminary exploration of the effects of a 6-week interactive video dance exercise program in an adult population,” *Cardiopulmonary Physical Therapy Journal*, vol. 22, no. 4, pp. 5–11, 2011.

[3] Y. Kim and H. Choi, “Landscape of Korean dance in the 1960s through analysis of dance video from Garfias Collection of the National Gugak Center,” *The Journal of Dance Society for Documentation & History*, vol. 59, pp. 7–33, 2020.

[4] M. Murakami, J. K. Tan, H. Kim, and S. Ishikawa, “Human motion recognition using directional motion history images,” *Proceedings of International Conference on Artificial Life and Robotics*, vol. 25, pp. 779–782, 2020.

[5] K. Yeonho and K. Daijin, "Real-time dance evaluation by markerless human pose estimation," *Multimedia Tools and Applications*, vol. 77, pp. 1–22, 2018.

[6] M. Gholami, A. Rezaei, H. Rhodin, R. Ward, and Z. J. Wang, "Self-supervised 3D human pose estimation from video," *Neurocomputing*, vol. 488, pp. 97–106, 2022.

[7] R. Komiya, T. Saitoh, M. Fuyuno, Y. Yamashita, and Y. Nakajima, "Head pose estimation and motion analysis of public speaking videos," *International Journal of Software Innovation*, vol. 5, no. 1, pp. 57–71, 2017.

[8] R. Zhang, "Analyzing body changes of high-level dance movements through biological image visualization technology by convolutional neural network," *The Journal of Supercomputing*, vol. 78, no. 8, pp. 10521–10541, 2022.

[9] P. Swamy and B. A. Reddy, "Human pose estimation in images and videos," *International Journal of Engineering & Technology*, vol. 7, no. 3.27, pp. 1–6, 2018.

[10] H. Cheng, D. Yang, C. Lu, Q. Qin, and D. Cadasse, "Intelligent Oil Production Stratified Water Injection Technology," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 3954446, 2020.

[11] T. Mallick, P. P. Das, and A. K. Majumdar, "Posture and sequence recognition for Bharatanatyam dance performances using machine learning approach," 2019, http://arxiv.org/abs/1909.11023.

[12] X. Yang and Y. Lyu, "Dance posture analysis based on virtual reality technology and its application in dance teaching," *Educational Sciences: Theory & Practice*, vol. 18, no. 5, 2018.

[13] C. T. Swain, D. G. Whyte, C. L. Ekegren et al., "Multi-segment spine kinematics: relationship with dance training and low back pain," *Gait & Posture*, vol. 68, pp. 247–268, 2018.

[14] H. Cheng, P. Ma, G. Dong, S. Zhang, J. Wei, and Q. Qin, "Characteristics of Carboniferous Volcanic Reservoirs in Beisantai Oilfield, Junggar Basin," *Mathematical Problems in Engineering*, 2022.

[15] J. Iqbal and M. S. Sidhu, "Acceptance of dance training system based on augmented reality and technology acceptance model (TAM)," *Virtual Reality*, vol. 26, no. 1, pp. 33–54, 2022.

[16] A. Barton, "Re-presenting a dance moment," *IDEA JOURNAL*, vol. 17, no. 2, pp. 265–274, 2020.

[17] J. Polechoński, W. Mynarski, W. Garbaciak, A. Fredyk, M. Rozpara, and A. Nawrocka, "Energy expenditure and intensity of interactive video dance games according to health recommendations," *Central European Journal of Sport Sciences and Medicine*, vol. 24, pp. 35–43, 2018.

[18] F. Sparacino, C. Wren, G. Davenport, and A. Pentland, "Augmented performance in dance and theater," *International Dance and Technology*, vol. 99, pp. 25–28, 1999.

[19] E. Ardizzone and M. E. Celebi, "Image and video analysis, detection and recognition," *Journal of Electronic Imaging*, vol. 27, no. 1, pp. 051201.1–051201.2, 2018.

[20] W. Luo and B. Ning, "High-dynamic dance motion recognition method based on video visual analysis," *Scientific Programming*, vol. 2022, Article ID 6724892, 9 pages, 2022.

[21] L. Liu, "Moving object detection technology of line dancing based on machine vision," *Mobile Information Systems*, vol. 2021, Article ID 9995980, 9 pages, 2021.

[22] Y. Li, "Dance motion capture based on data fusion algorithm and wearable sensor network," *Complexity*, vol. 2021, Article ID 2656275, 11 pages, 2021.

[23] Z. Hou, Y. Guo, S. Liang et al., "Shale gas stimulation technology: large-scale triaxial physical simulation tests on longmaxi formation shale," *Soil Mechanics and Foundation Engineering*, vol. 58, no. 6, pp. 491–499, 2022.

[24] I. Ajili, M. Mallem, and J. Y. Didier, "Human motions and emotions recognition inspired by LMA qualities," *The Visual Computer*, vol. 35, no. 10, pp. 1411–1426, 2019.

[25] D. Zhang, "Intelligent recognition of dance training movements based on machine learning and embedded system," *Journal of Intelligent and Fuzzy Systems*, vol. 1, pp. 1–13, 2021.