# A Method of Recognizing Specific Movements in Children's Dance Teaching Video Based on Edge Features

Chunhui Liu[1(✉)] and Chao Long[2]

[1] School of Economics and Management, Hunan Software Vocational and Technical University, Xiangtan 411000, China
liuchunhui869@163.com

[2] College of Art, Hebei University of Economics and Business, Shijiazhuang 050061, China

**Abstract.** Dance videos have issues with self occlusion and high complexity of actions, which affect the effectiveness of action recognition. In order to improve the accuracy of action recognition, a video action recognition method for children's dance teaching based on edge features is proposed. Image preprocessing for children's dance teaching videos, including grayscale and enhancement of video images. The background subtraction method is used to detect moving objects in video images, and the Canny operator is used to detect the edges of moving objects, enhancing the continuity of the edges. After obtaining an image that only includes the edges of the object, further extract the contour features of the object. A recognition method based on Adaboost BP neural network has been constructed. Using the BP neural network as a weak classifier, the Adaboost algorithm is combined with the outputs of multiple BP neural networks to construct a strong classifier, avoiding falling into local optima. Using edge features as input to achieve specific action recognition for children's dance teaching videos. The experimental results show that the recognition method based on edge features has a high average recognition accuracy of 94.715%.

**Keywords:** Edge Feature · Children's Dance Teaching Video · Pretreatment · Adaboost-BP Neural Network · Specific Action Recognition Method

## 1 Introduction

Motion capture recognition is very challenging in computer research, mainly using classification recognition and image processing technology to analyze video data to achieve human motion recognition. This direction has high research value and has attracted a large number of scholars and researchers from scientific research institutions. Motion recognition technology is applicable to a variety of video scenes, and has been widely used in video retrieval, intelligent human-computer interaction, virtual reality, intelligence, motion aided analysis and other fields. However, at present, the application of this technology in dance video is still less, and due to the problems of self occlusion and high


L. Yun et al. (Eds.): ADHIP 2023, LNICST 550, pp. 214–229, 2024.
https://doi.org/10.1007/978-3-031-50552-2_14

complexity of actions in dance video, the research in this area needs to be further carried out. The successful application of motion recognition technology in other fields also provides sufficient basis for the research in this field [1]. For the current large number of dance video data, professionals often need to spend a lot of time to analyze the data by watching and listening, which requires a lot of manpower and material resources and is extremely inefficient. The application of motion recognition technology in the analysis of these materials and the realization of dance motion recognition can not only reduce the work intensity of data analysts, facilitate the retrieval of video data, but also improve the efficiency of the automatic choreography system. It is also of great significance for mining and protecting cultural heritage in the art field and dance teaching [2]. In addition, the research in this area also has some reference and guiding significance for video action recognition in different environments, and can enrich the research in the direction of action recognition technology.

Human motion recognition has always been widely concerned. With the continuous progress of feature extraction methods and classification algorithms, motion recognition methods also continue to develop. Most of the early action recognition methods based on traditional computer vision use manual features extracted from action sequences, and then use multi-layer perceptron or support vector machine to classify based on these features. For example, based on the recognition method of improved dense trajectories, this method extracts features through the trajectories of sampling points, and then classifies the encoded features using SVM. This method has achieved good results in action recognition, and is one of the best algorithms in the field of traditional machine learning action recognition. However, traditional machine learning methods rely too much on high-performance manual design features, which requires a large number of experiments and prior information and is inefficient. With the increasing size of data sets and the increasing computing power of computing devices, deep learning has made great progress in the field of motion recognition. For example, the identification method based on dual stream network. The dual stream network is divided into two subnetworks, spatial stream network and temporal stream network. The spatial stream network extracts the spatial information of the sequence based on RGB image frames, and the temporal stream network obtains the temporal information of the sequence based on the optical flow extracted from adjacent image frames, and classifies them respectively. Finally, the average of the softmax scores of the two networks is taken as the classification result. The dual stream network is simple and effective, but it also has disadvantages correspondingly. In the time flow network, optical flow is the vector of motion. It is obtained by gradient calculation of two adjacent frames. In order to represent the motion information, it is input as multi frame optical flow. However, there are limitations. The number of frames is too small to describe the motion information well, resulting in poor recognition effect. If the number of frames is too large, the calculation time will increase, the efficiency will decrease, and the performance will not necessarily improve, Therefore, the effect of the two stream network on the long-term action modeling is poor. Another type of action recognition method is based on bone joint point data. Bone data is not easily affected by the above factors. Compared with RGB and depth data, the amount of bone data used to represent human action sequences is relatively small. Therefore, research on this type of method has gradually increased in recent years. For example,

the method based on RNN converts bone data into feature vectors, and then models such as LSTM are used for modeling. Finally, the whole body bone features are sent into the classifier to get the final recognition results.

The main task of this paper is to study the methods of human motion recognition, and apply them to the recognition of specific dance movements, and propose a method of children's dance teaching video specific motion recognition based on edge features. The Canny operator is used to detect the edges of moving objects and enhance the continuity of the edges, thereby obtaining an image that only includes the edges of the object. This edge detection method can effectively extract the boundary information of objects. After obtaining the edge image of the object, a recognition method based on Adaboost BP neural network is used to further extract the contour features of the object. This method combines Adaboost algorithm and BP neural network, which can effectively improve the accuracy and robustness of the classifier. Innovatively using BP neural networks as weak classifiers and combining the Adaboost algorithm with the outputs of multiple BP neural networks to construct a strong classifier. This combination method can avoid falling into local optimum and improve the accuracy and stability of Object detection.

## 2 Research on Recognition of Specific Actions in Children's Dance Teaching Videos

The research on the combination of video motion recognition technology and dance art has just started in China. Through the application of human motion recognition technology to dance videos, dance movement posture can be effectively recognized. By comparing the video action with the standard action, we can evaluate the dancers' dance posture and give suggestions for modification, which is an advanced auxiliary training method.Under this background, this paper proposes a method based on edge features to recognize specific actions in children's dance teaching videos.

### 2.1 Pre Processing of Video Images for Children's Dance Teaching

(1) Video image graying

The grayscale transformation of an image refers to the method of changing the grayscale value of each pixel in the source image point by point according to the transformation function to achieve a certain target condition [3]. It can be expressed as

$$y(i,j) = I\left[f(i,j)\right] \tag{1}$$

where, $f(i,j)$ Is the grayscale function of the original image, $I[]$ Is a transformation function; $y(i,j)$ Is an output image function. The piecewise linear transformation can enhance the contrast of the image, highlight the areas of interest in the image, and effectively solve the problem of poor quality of the collected image. It is one of the

commonly used gray transformation methods. The mathematical expression of the three-stage linear transformation method used in this paper is:

$$y(i,j) = \begin{cases} \frac{\chi}{\alpha} \times f(i,j) & 0 \le f(i,j) \le \alpha \\ \frac{\delta-\chi}{\beta-\alpha} \times [f(i,j)-\alpha] + \chi & \alpha \le f(i,j) \le \beta \\ \frac{\psi-\delta}{\psi-\beta} \times [f(i,j)-\beta] + \delta & \beta \le f(i,j) \le \psi \end{cases} \tag{2}$$

where, $\psi$ It is the maximum gray scale of the video image. By adjusting the position of the inflection point of the polyline and the slope of the segmented straight line, that is, the control parameters $\alpha$, $\beta$, $\chi$, $\delta$ The expansion or compression of any gray range can be realized by taking the value of

(2) Image enhancement processing

In order to facilitate the subsequent analysis and processing by the computer, the video image of children's dance teaching is converted into a simpler gray image, but the gray image after conversion often has the problem of poor contrast, so the enhancement of gray image is crucial. Dynamic histogram equalization is developed on the basis of histogram equalization. Its equalization idea is to construct a cumulative distribution function through the probability density function of the original histogram, then use the cumulative distribution function and mapping interval to calculate the intensity value of the output image, and finally remap the pixel value of the original image [4]. The difference between the two is that dynamic histogram equalization divides the overall histogram into multiple sub histograms, and assigns a new mapping interval to each sub histogram. For interval: $[H_0, H_{L-1}]$ The probability density function and cumulative distribution function of a sub histogram of are as follows:

Probability density function:

$$\eta(k) = \frac{n_k}{A} \tag{3}$$

Cumulative distribution function:

$$\mu(k) = \sum_{q=H_0}^{k} \eta(q) \tag{4}$$

where, $k$ It refers to a certain gray level within the range, $n_k$ finger $k$ The frequency of gray level appearance, that is, the number of pixels, $A$ It refers to the total number of pixels in the range.

The specific process is as follows:

1) Input the video image of children's dance teaching.
2) Obtain the histograms of R, G and B channels of children's dance teaching video images respectively, which are recorded as $H_R, H_G, H_B$.
3) Calculation $H_R, H_G, H_B$ The distribution range of.
4) According to the distribution range, the improved segmentation method based on exposure value is used to calculate the segmentation points iteratively, and each histogram is divided into four sub histograms.

5) Use the set reconstruction parameters $b$ Calculate clipping threshold for each sub histogram $T_i$. The calculation formula is as follows:

$$T_i = \frac{A_i}{a_i} + b\left(f(H_i) - \frac{A_i}{a_i}\right) \tag{5}$$

Where, $T_i$ On behalf of the $i$ The total number of pixels in the sub histogram, $a_i$ Refers to the length of the interval, $b$ Is a reconstruction parameter whose value range is (0,1), $f(H_i)$ Is to get the first $i$ A function of the peak value of the sub histogram. Therefore, the adjustment range of the clipping threshold is between the average value and the peak value [5].

6) Statistics of exceeding clipping threshold in each sub histogram $T_i$ Total number of pixels $C_i$ And calculate its pixel redistribution value $D_i$.

$$D_i = \frac{(1-b)C_i}{a_i} \tag{6}$$

Among them, $C_i$ Refers to the $i$ The total number of pixels trimmed from the sub histogram. Reconstruction parameters are also used here $b$, so that the clipping threshold is associated with the reallocated value, that is, the more pixels are clipped, the larger the reallocated value is.

7) Rebuild each sub histogram, i.e.

$$E_i = \begin{cases} T_i, n_{ki} > T_i - D_i \\ n_{ki} + D_i, \text{otherwise} \end{cases} \tag{7}$$

Where, $E_i$ Refers to the reconstructed No $i$ Sub histogram, $n_{ki}$ Is the first $i$ In the sub histogram $k$ The number of pixels on each grayscale.

8) Construct the corresponding cumulative distribution function for each reconstructed sub histogram $\mu(k)$.

9) A new mapping interval is allocated according to the total pixel proportion of each sub histogram after reconstruction.

$$H_i = \frac{Lg(E_i)}{\sum_{g(E_i)=1}^{4} g(E_i)} - 1 \tag{8}$$

where, $H_i$ Is the first $i$ The segmentation point of the sub histogram mapping interval, $g_q$ It is the first time after reconstruction $q$ Sub histogram $E_i$ The total number of pixels for.

10) Equalize each sub histogram independently and remap each channel.

$$G(k) = H_0' + \left(H_{L-1}' - H_0'\right)\mu(k) \tag{9}$$

where, $\left[H_0', H_{L-1}'\right]$ It refers to the new mapping interval.

11) The three channels after remapping are fused to obtain enhanced results.

12) End.

### 2.2 Edge Feature Extraction

In the research of action recognition, the first step is usually feature extraction. Feature extraction refers to extracting the feature information used to describe the target action in the video from the action data set, which is an essential step for the research of action recognition. From this point of view, the extracted features play a vital role in the accuracy of action recognition results and the robustness of action recognition methods. In this paper, after fully considering the characteristics of dance movements, edge features are extracted from dance videos. Before edge feature extraction, the moving human body needs to be extracted. A video image generally contains two parts: moving area and still area. The purpose of moving object detection is to successfully separate the two parts and extract the moving area [6]. Background subtraction is a common method for moving object detection in video images. This method is a method based on background modeling. First, a stable background is established through parameter updating, and then the original video object is compared with the background to get the detection result, i.e. foreground target. The flow of this algorithm is shown in Fig. 1 below.

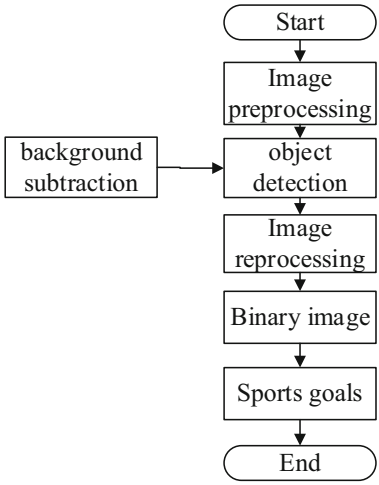


**Fig. 1.** Flow chart of background subtraction method

The above flow chart can be explained by formulas (10)~(11).

$$Q_t = |l_t - e_t| \tag{10}$$

$$\hat{Q}_t = \begin{cases} 1, Q_t > \varepsilon \\ 0, \text{otherwise} \end{cases} \tag{11}$$

In the above formula, $l_t$ Represents the original video frame, $e_t$ Is the background frame, $Q_t$ by $l_t$ and $e_t$ The difference image obtained by subtraction. Equation (11) indicates $Q_t$ The threshold value is set to $\varepsilon$, $\hat{Q}_t$ It is the result of binarization.

There is no doubt that the key of background subtraction based on background model is the establishment of background model. The common background modeling method is the average method. The principle of the average method is simple. It is to accumulate and sum multiple video image sequences containing moving objects in turn, and then get the background image [7] by averaging. The mathematical formula is shown in Eq. (12).

$$B(i,j) = \frac{\sum_{i=1}^{N} J(i,j)}{N} \tag{12}$$

where, $N$ Represents the number of frames of the video image; $J(i,j)$ Represents the grayscale value of the image, $B(i,j)$ Is the desired background image, changing pixels $(i,j)$ A stable and reliable background image can be obtained by using the value of.

In the above research, the moving human body has been successfully extracted, and the detection results obtained are all expressed in binary images. To further analyze the human behavior, it is necessary to describe a person's behavior as accurately as possible, which requires that some features that can fully represent the behavior be extracted first. Different images have different features. There are many features that can be extracted from an image, such as edge features, motion features, texture features, color features, light features, etc. In order to accurately describe the behavior of moving objects, edge features are extracted in this paper. The edge of an image generally refers to the places where the gray level, texture and other characteristics of pixels appear to be distributed in a jumping manner in an image. There are steps in the gray level changes of the image subject and image background pixels in these places, which are reflected in the function. The function image will show dramatic changes. Therefore, the traditional edge detection operators take this as the starting point. Calculate the first derivative or the second derivative [8] of the image gray level change. The first derivative operator considers that when the first derivative reaches the maximum value, it is the edge of the image, and the second derivative operator considers that when the second derivative of the function crosses the zero point, it is the edge of the image. The common first-order derivative operators are Roberts, Sobel and Prewitt operators; Common second-order derivative operators include Canny operator, Laplacian operator and Log operator. Canny operator to be introduced in this paper is non differential operator.

Canny gave three basic edge detection rules in 1986 as the basic idea of Canny operator, which is an optimized edge detection method up to now. The three basic principles are as follows:

① Signal to noise ratio rule: ensure the accuracy of the original image edge to prevent false edges;

② Positioning accuracy rule: the edge of the original image should be as close as possible to the tracked edge image;

③ Single edge response rule: the edge response should be unique to prevent multiple responses and resist virtual responses as much as possible.

The above three rules were first proposed by Canny, and Canny operator solved this problem completely in the form of mathematical expressions. The application of Canny operator in image edge detection has a very significant effect, which solves the problem of two-dimensional differential losing edge direction information, making Canny operator

one of the most widely used algorithms in edge detection. The detailed operation process of Canny operator is as follows:

In the first step, Gaussian filter operator is used to smooth the noisy image. Because the edge and noise information in the image are mostly concentrated in the high-frequency part, the noise information is easy to be detected as a false edge in the image; To eliminate noise interference, the traditional Canny algorithm uses a two-dimensional Gaussian filter $\phi(i,j)$ The image edge is smoothed by convolution. Set the original image as $f(i,j)$, the smoothed image $O(i,j)$ It can be expressed as:

$$O(i,j) = \phi(i,j) * f(i,j) \tag{13}$$

In the formula "$*$" Represents the convolution operation, and the Gaussian filter function formula is as follows:

$$\phi(i,j) = \frac{\exp\left(-\frac{i^2+j^2}{2\gamma^2}\right)}{2\pi\gamma^2} \tag{14}$$

where, $\gamma$ Is the standard deviation of the Gaussian function.

The second step is to obtain the gradient amplitude and gradient direction of all pixels in the image. After the smooth image is obtained through the Gaussian filter, the pixels of the image are in the horizontal direction $x$ And vertical direction $y$ Solve the partial derivative, and use the first order finite difference to calculate the gradient amplitude $Z(i,j)$ And gradient direction $\vartheta(i,j)$:

$$Z(i,j) = \sqrt{\left[\phi_x(i,j)\right]^2 + \left[\phi_y(i,j)\right]^2} \tag{15}$$

$$\vartheta(i,j) = \arctan\frac{\phi_x(i,j)}{\phi_y(i,j)} \tag{16}$$

The calculation formula for gradient amplitude and bearing transformation using rectangular coordinates is:

$$\phi_x(i,j) = \frac{f(i+1,j) - f(i,j) + f(i+1,j+1) - f(i,j+1)}{2} \tag{17}$$

$$\phi_y(i,j) = \frac{f(i,j+1) - f(i,j) + f(i+1,j+1) - f(i+1,j)}{2} \tag{18}$$

At this time, the edge strength of the image can be reflected, and the direction perpendicular to the edge can be reflected.

The third step is to perform non maximum suppression operation on the gradient amplitude of pixels in the image. Non maximum suppression is a kind of edge thinning technology. The role of non maximum suppression is to "thin" edges. After gradient calculation of the image, the edge extracted only based on gradient value is still fuzzy. For Criterion 3, there is and should be only one accurate response to the edge. Non maximum suppression can help suppress all gradient values other than the local maximum to 0. The algorithm for non maximum suppression of each pixel in the gradient image is:

(1) Compare the gradient intensity of the current pixel with two pixels along the positive and negative gradient directions;
(2) If the gradient intensity of the current pixel is the largest compared with the other two pixels, the pixel will remain as an edge point, otherwise the pixel will be suppressed.

In the fourth step, the double threshold algorithm is used to detect edges and connecting edges. Select high threshold first $R_{\max}$ And low threshold $R_{\min}$, then start scanning the map $f(i,j)$ Each pixel of the candidate edge points marked in the candidate edge image $(i,j)$ Detect if the pixel $(i,j)$ Amplitude of $Z(i,j)$ greater than $R_{\max}$, then this point is the edge point; If pixels $(i,j)$ Amplitude of $Z(i,j)$ lower than $R_{\min}$, then the point is not an edge point; gradient magnitude $Z(i,j)$ The points between high and low thresholds are regarded as suspected edge points, which need to be further judged according to edge connectivity. If there are edge points in the adjacent pixel points of the point, the point is regarded as an edge point for connection; Otherwise, the point is a non edge pixel and is discarded [9].

After the image containing only the target edge is obtained, the contour features of the object can be further extracted. The contour features that can be extracted are as follows:

1) Contour area: traverse all pixels of the image. If the pixel is in the contour, the value is increased by 1, and the final value is the contour area;
2) Contour Perimeter: calculate the spacing between adjacent points of the contour and accumulate it. The accumulated value is the contour perimeter;
3) Hu moment similarity: the second and third order normalized central moments of the object contour are linearly combined to obtain the Hu moment invariant of the contour, and the Hu moment similarity is obtained by comparing the Hu moment invariant of the target and the model;

### 2.3 Specific Action Identification

After the timely detection and effective feature extraction of moving objects are completed, the next problem is how to judge and classify the processed video image according to these effective features. The commonly said "identifying something" actually includes two points of recognition and distinction, that is, first recognize the object, that is, carefully observe one, two or more of its features, and then use these features to distinguish it from other objects in a pile of objects according to certain discrimination rules [10].

The full English name of BP neural network is back propagation neural network, that is, back propagation neural network. On the premise that the number of neurons in the hidden layer can be adjusted at will, it can approach any nonlinear mapping, and has a certain fault tolerance ability, so it is suitable as a classifier. Although BP neural network has many advantages [11], it is difficult to establish a better neural network system, which is caused by its own shortcomings. It mainly includes:

1) It is difficult to determine the number of iterations of the algorithm, and may oscillate around the local minimum value, which will lead to slow convergence of the algorithm. This will lead to too many iterations and low learning efficiency.
2) The gradient descent method is used to minimize the error function. The error function may contain multiple local poles [12]

3) The number of hidden layer nodes is difficult to determine. Must be selected by experiment or experience.
4) The newly added training samples will affect the original training samples, that is, the weight distribution of the network is modified.

Aiming at the problems of BP neural network, a massive image recognition method based on Adaboost-BP neural network is proposed. BP neural network is used as a weak classifier [13], and Adaboost algorithm is used to combine the output of multiple BP neural networks to build a strong classifier, as shown in Fig. 2 below.

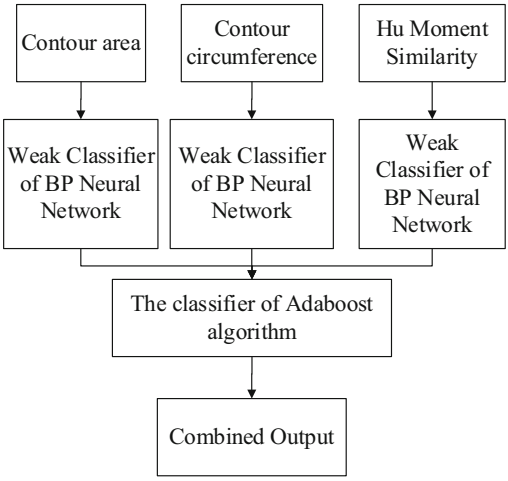


**Fig. 2.** Strong classifier model of the Adaboost BP neural network

AdaBoost (abbreviation of adaptive boosting) is a framework algorithm. AdaBoost learning algorithm can cascade a series of weak classifiers to form a strong classifier. The types of multiple weak classifiers used by AdaBoost classifier are consistent. Different classifiers are obtained through serial training, and later classifiers will be obtained through training according to the wrong data of the previous classifier. The output of AdaBoost classifier is the result of weighted summation of multiple weak classifiers [14]. The weight of each classifier is determined by its classification success rate. The higher the classification success rate, the greater the corresponding weight. Advantages of AdaBoost algorithm: low generalization error rate, easy implementation, and can be applied to most classifiers without parameter adjustment.

The construction of Adaboost-BP classifier needs two stages: training stage and recognition stage. The training process is as follows:

Input: training sample set $V$, the number of samples is $n$, one for each sample $m$ Dimensional eigenvector $U$ Representation, vector $U$ The element in represents an edge feature of the sample and the number of categories is $\lambda$.

Step 1: Because the range of each feature of children's dance teaching video image is different, there is no comparability between features. In order to eliminate this difference, all characteristic values can be converted to the range of [0,1]. Therefore, before inputting

the image feature vector of children's dance teaching video, it is necessary to normalize the image features:

$$\hat{v}_{ij} = \frac{v_{ij} - \min v_i}{\max v_i - \min v_i} \tag{19}$$

Among them, $v_{ij}$, $\hat{v}_{ij}$ It is the first time before and after processing $j$ No. of samples $i$ Features. max $v_i$, min $v_i$ It is the No $i$ The maximum and minimum values of the features.

Step 2: Set parameters of BP neural network weak classifier.

Step 3: Set the number of Adaboost iterations and training times $\varpi$ The initial value is 0; Parameters of BP network training stop: number of iterations, convergence accuracy; The number of incremental iterations each time when the condition is not met $N_u$. Let the initial weight vector of the sample be $w_0$, weight of each sample $w_{i,0} = \frac{1}{n}$.

Step 4; Create section $\varpi$ Weak classifiers $CL_\varpi$. Training section $\varpi$ BP neural network, the initial weight of neural network is set randomly.input $\hat{v}_{ij}$ To the neural network, and then calculate the error between the neural network output and the expected output $err_j$, Calculate $err_j$ Weighted error rate of $\Delta e$.

$$\Delta e_\varpi = \sum_{j=1}^{n} err_j \cdot w_{i\varpi} \tag{20}$$

Among them, $err_j$ On behalf of the $j$ Is the sample output correct? If it is wrong, then $err_j = 1$, conversely $err_j = 0$. If $\Delta e_\varpi < 0.5$, then use the trained neural network as the first $\varpi$ Weak classifiers $CL_\varpi$ Otherwise, the neural network will be trained every iteration on the original basis $N_u$ And check the weighted error rate of e, until the weighted error rate meets $\Delta e_\varpi < 0.5$ before stopping iteration. Finally, the weighted error rate $\Delta e_\varpi$ As the AdaBoost framework $\varpi$ Iterations $CL_\varpi$ Error rate.

Step 5: Calculation $u_\varpi$ As $CL_\varpi$ Weight of.

$$u_\varpi = \frac{\ln\left(\frac{1-\Delta e_\varpi}{\Delta e_\varpi}\right)}{2} \tag{21}$$

Step 6: Update each sample to $w_{i,\varpi+1}$ The weight of is calculated according to the formula, normalized, and then executed $\varpi = \varpi + 1$, and skip to step 5 for the next iteration.

Step 7: BP weak classifier is established. judge $\varpi <$ Maximum Iterations max $\varpi$ Whether it is true or not. If it is not true, the iteration is completed and the strong classifier has been established. Exit the algorithm. Otherwise, skip to Step 4.

For the established Adaboost-BP strong classifier, the joint output of each output node is:

$$Y_i = \arg\max \sum_{\varpi}^{\max \varpi} \log \frac{1}{u_\varpi} (CL_\varpi.out_\varpi = \varsigma) \tag{22}$$

among $\varsigma$ If it is 0 or 1, that is, the weight sum of the output is 1 and the weight sum of the output is 0. Select the maximum weight and the corresponding output as the output of the strong classifier.

After completing the training of Adaboost-BP classifier, it can be used for test sample recognition.

# 3 Method Test

## 3.1 Datasets

At present, the research on the combination of motion recognition technology and dance has just started, and the available dance data set is still relatively small. The open motion capture data set of Carnegie Mellon University, but the data set contains very little dance data, which cannot be specifically used for dance motion recognition research; The Dance DB dance data set published by the Virtual Reality Laboratory of the University of Cyprus can meet the requirements of dance action recognition research. Therefore, two dance data sets were used in the experiment, namely, the Dance DB data set and the folk dance data set produced by my laboratory. In the Dance DB dataset, each dance category uses emotion markers; The Folk Dance dance data set is divided into four groups in total. Each group contains a number of subdivisions of dance actions. The action categories are relatively rich, and each group of dance actions is relatively complex and challenging.

(1) DanceDB

There are 48 dance videos in the DanceDB dance dataset currently published by the Virtual Reality Laboratory of the University of Cyprus. The background and camera perspective in each dance video are fixed. The frame rate of the image is 20fps, and the size of each frame is 480 * 360. Although the data set currently contains a relatively small number of categories, there are challenges such as easily mixing moving objects and backgrounds in the video. It is an excellent dance action data set published in the field of dance action analysis research. Therefore, it can be used to measure the effectiveness of the algorithm proposed in this paper. There are 12 kinds of dance actions in the DanceDB dance data set, each of which is marked with an emotion tag as the category of this kind of dance action. The dance action categories in this dataset are: Afraid, Angry, Annoyed, Bored, Excited, Happy, Miserable, Pleased, Relaxed, Sad, Satisfied, Tired.

(2) FolkDance

The FolkDance dance data set is a dance data set produced by the laboratory itself. It uses the motion capture equipment Vicon to collect professional dance action videos. During the production of the entire data set, four groups of folk dance actions are designed according to the data set production plan and the final scheme discussed with dance experts. In view of the fact that the research on dance action recognition is still in its infancy, the production of the FolkDance dataset currently only considers the situation of single dance, and does not consider the changing stage background, props and other factors. In the specific process of dance video capture, we invited several dance majors to perform dance according to the grouping settings, while using the Vicon device to collect dance video data. A total of 84 dance videos were recorded, and the background and camera angle of view in each video were set to be fixed. The image in the video is uniformly set to a frame rate of 20 prints per second, and the size of each frame is 480 * 3600. This data set contains many types of dance actions, and the dance actions are complex, which is challenging for dance action recognition. Therefore, this data set can be used to verify the effectiveness of the dance action recognition algorithm proposed

in this paper. The FolkDance dance data set mainly includes four groups of dances, namely, step double flower combination, lining flower combination, towel flower combination and flower combination.

### 3.2 Test Method

In order to verify the feasibility of the dance action recognition algorithm in this paper, we used cross validation to evaluate the algorithm in the experiment. Cross validation is a statistical method of cutting data samples into subsets. Its idea is to divide the original data set into training set and test set. Usually, the training set is used to train the classifier. After the training is completed, the test set is used to test the model obtained through training, and to evaluate the performance of the classifier, that is, the feasibility of the algorithm. K-fold cross validation is a common method of cross validation.

K fold cross validation: divide the data set into K groups, select one of them as the test set, and the remaining K-1 groups as the training set. Repeat the cross validation for K times, and select one group from them as the test set each time. Finally, the recognition accuracy of K times is taken as the final recognition result through the average cross validation. Generally, the value of K in the experiment is l0. On DanceDB, one person's dance data set is selected as the test set each time, and the rest three people's dance data set is selected as the training set. Repeat four times, and finally take the average result of the four times as the final result; For the FolkDance data set, the data of one person is also selected as the test set each time, and the data of the other two persons is used as the training set. Repeat three times, and finally average the final results of the three times.

### 3.3 Edge Contour Features

Use the research in chapter 1.2 to extract image edge contour features. Taking one of the video image samples as an example, the image edge contour features are shown in Fig. 3 below.

(a) Original drawing (b) Edge contour feature drawing

**Fig. 3.** Edge contour feature extraction results

The outline area of the figure is 24.63 cm2; The contour perimeter is 32.27 cm;Hu moment similarity is 0.745.

### 3.4 Method Test Results

Combining all recognition results, calculate the action recognition accuracy of the two dance video data sets, and then calculate the average value. The results are shown in Table 1 below.

**Table 1.** Accuracy rate of action recognition

| Method | Data set | Recognition accuracy/% | Average recognition accuracy/% |
|---|---|---|---|
| A Recognition Method Based on Edge Features | DanceDB | 93.65 | 94.715 |
| | FolkDance | 95.78 | |
| Identification Method Based on Improved Dense Trajectory | DanceDB | 88.66 | 85.22 |
| | FolkDance | 81.78 | |
| Identification method based on dual flow network | DanceDB | 85.97 | 84.375 |
| | FolkDance | 82.78 | |
| Recognition method based on bone joint point data | DanceDB | 78.62 | 82.795 |
| | FolkDance | 86.97 | |

According to Table 1, the recognition accuracy of the article's method is relatively high, with an average of 94.715%. However, the average recognition accuracy of the improved dense trajectory based recognition method is 85.22%, the average recognition accuracy of the dual flow network based recognition method is 84.375%, and the average recognition accuracy of the bone joint number based recognition method is 82.795%. This indicates that the edge feature based recognition method proposed in this article has high recognition accuracy and can effectively achieve specific action recognition.

## 4 Conclusion

The recognition and analysis of dance movements has a very broad application prospect, and can play an important role in such aspects as dance video understanding, dance distance teaching and cultural protection. Nowadays, more and more advanced motion recognition algorithms and computing devices also help to achieve all of this. However, the research in this area is still very rare. This paper studies dance motion recognition based on edge features, and mainly completes the following work:

(1) Aiming at the research and analysis of the characteristics of dance movements at the same time, this paper proposes an effective method to extract edge features, analyzes the general steps of edge detection, and then takes Canny operator as an example, mainly introduces the steps of edge detection using Canny operator and the specific operations of each step Hu moment similarity characterizes the appearance and contour features of dance movements in the video.

(2) The AdaBoost framework algorithm is used to improve the BP neural network algorithm, and the AdaBoost enhanced BP neural network algorithm is proposed, which overcomes the problems that the traditional BP neural network is easy to fall into the local minimum and the convergence speed is slow.
(3) Another major contribution of this paper is the collection and production of dance data sets. For dance action recognition research, dance data sets play a very critical role in the research. We have specially produced a folk dance dataset. During the production process, we developed a detailed data set recording scheme, and discussed with professional dance experts about the production of dance data sets. In the later specific recording process, we used the Vicon motion capture system to invite different dance majors to record dance videos according to the dance group motion design. At the same time, considering that the research on dance action recognition is still in its infancy, and the dance action is too complex, our data set is recorded in a fixed scene and with a single person performing dance. At present, we have completed a total of three person times, four groups, and 84 dance action videos, as well as other single person time and multi category data sets for other dance research.

Although the method in this paper has finally achieved good action recognition results, there are still many areas worthy of improvement. In view of the complexity of dance movements, the dance data set we have produced at this stage only considers the situation of single person dance, and does not consider factors such as changing stage scenes. In the future, we will focus on more challenging research on dance action recognition such as scene changes and multi person dance, extract dance action features more suitable for complex backgrounds to better represent dance actions, and then make research results on dance action recognition that are more in line with actual needs such as real dance choreography.

**Aknowledgement.** Characteristics and Innovation of Grassroots Party Building in Xiangtan under the Background of Rural Revitalization (Project No. 2023C54).

## References

1. Wang, F., Hu, R., Jin, Y.: Research on gesture image recognition method based on transfer learning. Procedia Comput. Sci. **187**(10), 140–145 (2021)
2. Jin, S.: Image recognition method for fault service action of tennis based on feature matching[J]. International Journal of Biometrics **13**(2/3), 150 (2021)
3. Sun, K., Zhang, B., Chen, Y., et al.: The facial expression recognition method based on image fusion and CNN. Integr. Ferroelectr. **217**(1), 198–213 (2021)
4. Yang, X., Liu, D., Liu, J., et al.: Follower: A Novel Self-Deployable Action Recognition Framework[J]. Sensors **21**(3), 950 (2021)
5. Toldinas, J., Venkauskas, A., Damaeviius, R., et al.: A novel approach for network intrusion detection using multistage deep learning image recognition. Electronics **10**(15), 1854 (2021)
6. Daradkeh, Y.I., Tvoroshenko, I., Gorokhovatskyi, V., et al.: Development of effective methods for structural image recognition using the principles of data granulation and apparatus of fuzzy logic. IEEE Access **9**(99), 13417–13428 (2021)
7. Chen, M., Wang, X., Luo, H., et al.: Learning to focus: cascaded feature matching network for few-shot image recognition. Sci. Chin. Inf. Sci. **64**(9), 192105 (2021)

8. Xiong, J., Yu, D., Liu, S., et al.: A review of plant phenotypic image recognition technology based on deep learning. Electronics **10**(1), 81 (2021)
9. Jin, L., Liang, H., Yang, C.: Sonar image recognition of underwater target based on convolutional neural network. Xibei Gongye Daxue Xuebao/J. Northwest. Polytechnical Univ. **39**(2), 285–291 (2021)
10. Tian, L., Xu, H., Zheng, X.: Research on fingerprint image recognition based on convolution neural network. Int. J. Biometrics **13**(1), 64–79 (2021)
11. Lyu, Z., Yu, Y., Samali, B., et al.: Back-propagation neural network optimized by K-fold cross-validation for prediction of torsional strength of reinforced concrete beam. Materials **15**(4), 1477 (2022)
12. Reza Kashyzadeh, K., Amiri, N., Ghorbani, S., et al.: Prediction of concrete compressive strength using a back-propagation neural network optimized by a genetic algorithm and response surface analysis considering the appearance of aggregates and curing conditions. Buildings **12**(4), 438 (2022)
13. Cong, Y.L., Hou, L.T., Wu, Y.C., et al.: Energy consumption prediction and diagnosis of heating ventilation and air conditioning system based on bidirectional LSTM method. In: 2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI), pp. 633–636. IEEE (2022)
14. Jindal, H., Yadav, A., Sehgal, A., et al.: Geospatial landslide prediction–analysis & prediction from 2018-2022. J. Pharmaceutical Negative Results 2589–2599 (2023)