Research Article

Characteristic Analysis Technology of Moving People Based on Image Recognition

Yuanhong Xu

Department of Physical Education, Hangzhou Normal University, Hangzhou 311121, China

Correspondence should be addressed to Yuanhong Xu; 20021390@hznu.edu.cn

Received 11 April 2022; Revised 15 May 2022; Accepted 17 May 2022; Published 4 July 2022

Academic Editor: Liping Zhang

Copyright © 2022 Yuanhong Xu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, the visual analysis of human motion has become a frontier direction in the field of computer vision. It recognizes, identifies, and tracks people in image sequences, as well as understands and describes their actions. Using computer vision technology to study the field of image processing and pattern recognition, as well as extracting and effectively identifying human motion features from video images, has become a hot topic of concern with the rapid development and popularization of information technology. This study presents a method for analyzing the characteristics of moving human bodies based on image recognition, introduces the extraction function method, and analyzes the characteristics of the extraction function to improve the ability to analyze the characteristics of moving human bodies. To achieve fine segmentation, the hierarchical clustering algorithm is used to segment the periodic motion in each motion. Different benchmark databases and self-built data were used in the experiments. Experiments show that the algorithm can achieve good classification and recognition results while maintaining low computational complexity and extracting less feature data. It can also organically integrate the static and dynamic features of human walking.

1. Introduction

The visual analysis of human motion is a new frontier in computer vision that has gotten a lot of attention in recent years. It recognizes, identifies, and tracks people in image sequences, as well as understands and describes their actions. Moving target extraction is a low-level process that falls under the category of image processing. Target recognition, tracking [1, 2], image analysis, and comprehension are high-level processes that fall under the artificial intelligence umbrella [3, 4]. Today’s research focuses on using computer vision technology to analyze and study problems in the fields of image processing and pattern recognition, as well as extracting and recognizing human motion features from video images. A video can concretely and vividly express the information contained in things and objectively describe things as the information carrier of computer vision technology [5, 6] analysis and processing. The dialogue between people and computers will be more intuitive and concise if the computer can capture human motion through the video input device and recognize or even understand these behaviors. It is necessary to keep the human body in the camera’s field of vision throughout the interaction process to understand people’s behavior and then analyze the movement of the human body.

Human motion recognition can be divided into several different levels, and the classification standard is widely accepted: action primitive, action, and behavior. Zhou et al. [7] proposed a sliding window strategy with increasing length, which, combined with the measurement method of similarity between different samples in high-dimensional space, realized real-time motion segmentation [7]. Das and Arpita [8] estimated the head pose by learning the mapping function between face image features and head pose space using different nonlinear regressions [8]. Peng et al. [9] proposed a local gradient direction histogram head descriptor to reduce the deviation of face location caused by face detector [9]. At work, the author proposes to use kernel-based partial least squares regression method to select the most suitable one from a series of candidate face detection
windows to reduce the location deviation of face area. Although many researchers are trying different methods to improve the accuracy of head pose estimation, there are still many challenges in the research of head pose estimation, such as facial expression changes, partial facial occlusion, light, and other factors, which affect the performance of head pose estimation. Therefore, how to automatically and robustly estimate the head pose in video images is a challenging task.

It is necessary to create a suitable character model to accurately detect the characters in the image. As it is a joint object, most character models represent it with body parts; however, different models use different levels of detail [10]. This study develops a moving target detection and tracking system based on image recognition, with a focus on moving target image recognition. The main contributions of this study are as follows:

1. The motion segmentation algorithm is deeply studied. By analyzing the technical characteristics of the turning point segmentation algorithm and clustering segmentation algorithm, a new adaptive motion segmentation algorithm with unknown motion sequence segmentation is proposed.

2. In this study, an improved feature recognition method is proposed, which overcomes the shortcoming of the traditional method of using Fourier descriptors to represent contour features with a large amount of data. Then, the human identity is recognized by the nearest neighbor fuzzy classifier by combining with the limb joint angle features.

The content of this study is divided into five chapters, which are as follows: the first part mainly introduces the background of the topic selection and the significance and purpose of the research. The second part introduces the theoretical basis of the feature analysis of moving characters. The third part mainly introduces the motion segmentation algorithm and the improved feature recognition method, which provides the experimental basis for the following discussion. In the fourth part, the method of this study is realized through experiments. The fifth part presents the conclusion and prospect of this study.

2. Related Work

2.1. Research Status of Human Motion Recognition. Motion segmentation and motion recognition are the two main aspects of human motion analysis. To prepare for motion recognition and retrieval, motion segmentation involves identifying the transition frame from the captured motion data and separating the motion data of different actions on both sides. The term "motion recognition" refers to the process of labeling and categorizing motion segments. It is an important step in the human-computer interaction process. The computer finally achieves the goal of human behavior by recognizing actions.

Li et al. [11] put forward a semantic model of motion, which calculates the velocity and acceleration of feature points, quantifies them into corresponding characters, and then uses machine learning to segment the motion sequence [11]. Motion segmentation can also be achieved by clustering. Shao et al. [12] proposed a $K$-means extended clustering algorithm based on the kernel function, which used hierarchical alignment as a segmentation method for time domain clustering analysis, transformed the time domain segmentation problem into an energy minimization problem, and obtained the segmentation points with the help of the dynamic programming algorithm [12]. Gao and Shhardt [13] proposed the feature representation of ordinary 2D pictures by the optical flow method [13]. Tu and Kim [14] successfully extracted the 3D bone data of the human body from depth images with the decision random forest algorithm, which greatly reduced the dimensional redundancy of depth images and opened up a new research space for human motion recognition [14]. Chen et al. [15] used the principle of robot mechanics to transform the rotation and translation of each joint point into Lie group space for feature representation. It is proposed to identify the basic actions in people's daily life in real time according to the relative position characteristics of bone points [15].

2.2. Research Status of Moving Target Change Detection. The motion areas obtained by motion detection may contain different moving targets, and under different application backgrounds and image resolutions, people detection algorithms are also different. In the field of image cognition, it is necessary to identify the motion of various parts of the human body. In intelligent monitoring, as the image resolution is not high, we do not need to pay too much attention to the specific details of people but just need to judge whether it is the target of people.

Xu et al. [16] proposed that by reducing the correlation of different shape parameters, modeling the change of deformation with the Gaussian model in low-dimensional space can represent the global deformation, however, it is still insufficient for local deformation [16]. Xu and Ding [17] established a pendulum model with thigh and calf as a link, took the angle signal between the pendulum and the vertical direction as the walking movement feature, and then classified and identified the individual identity [17]. Sl et al. [18] used the regional dispersion, area, aspect ratio, and so on as characteristics and divided the moving targets into people, crowds, cars, and background trunks using the three-layer neural network method [18]. Feng et al. [19] used dispersion and area information to classify two-dimensional moving areas, mainly to distinguish people, cars, and chaotic disturbances, and the time consistency constraint made the classification more accurate [19]. By tracking the moving target of interest, Xw et al. [20] calculated the autocorrelation characteristics of the target with time, and the autocorrelation of the target is periodic because of the periodic movement of people. Therefore, the method of transforming time domain to frequency domain is used to analyze whether the target has periodic motion characteristics and identify whether the target is human [20]. Guo and Na [21] proposed a method of hierarchical classification using time co-occurrence matrix, which can be used to distinguish not only

...
objects but also behaviors [21]. Lu et al. [22] proposed the 
method of combining the color histogram model and gray
gradient model of the target to realize the real-time tracking
of the human head [22]. Binsaeedan and Alramlawi [23]
proposed a feature analysis method of moving people based
on image recognition, which proposed to extract the contour
of moving target by background subtraction and to build a
model of contour feature vector using the boundary moment
invariants and morphological features of the contour [23].

3. Research Method

3.1. Establishment of the Moving Target Model. Motion 
feature recognition involves a wide range in today’s scientific
research field, mainly involving a series of research fields,
such as image processing, multisensor technology, virtual
reality, pattern recognition, computer vision and graphics,
computer-aided design, visualization technology, intelligent
robot system, and so on. Real-time and effective detection
and feature extraction of moving targets are very important
for later target identification. In this study, an analysis
method of moving human features based on image recog-
nition is proposed. The extraction function method is in-
troduced, and the characteristics of the extraction function
are analyzed. The hierarchical clustering algorithm is used to
segment the periodic motion in each motion to achieve fine
segmentation. Experiments are carried out on different
benchmark databases and self-built data.

The goal of moving target detection is to see if any objects
in the sequence images are moving relative to the whole
scene. Moving target detection is the foundation for sub-
sequent processing, such as moving target recognition,
tracking, understanding, describing moving target move-
ments, and so on, and it has a significant impact on sub-
sequent processing. In this study, target recognition is
divided into two categories: human and nonhuman. Because
the moving target is inevitably occluded in an indoor en-
vironment and the shoulder and above areas of the human
body are not easily occluded and have relatively stable
shapes, this study uses the shape of the human head and
shoulder as a two-dimensional recognition model.

The steps of establishing the head and shoulder model
are as follows:

![Figure 1: Vertical projection histogram.](image)

Calculate the aspect ratio of the moving target.

$$r = \frac{x_{\text{max}} - x_{\text{min}}}{y_{\text{max}} - y_{\text{min}}}$$

If it is between [0.28 and 0.36], it means that the whole
human body may have entered the capture range of the
camera.

Calculate the vertical projection histogram of the moving
target, as shown in Figure 1, and find the local maximum
point $w$ near the top of the head.

Look for the global maximum point $w_i$ of the vertical
projection histogram, and calculate the approximate height
$h$ of the human body from the global maximum and the
aspect ratio of the human body in an upright condition.

We use the statistical information of high-level energy
distribution of images to describe binary face images. We use
$I(x, y), x \in [1, H], y \in [1, W]$ to represent the normalized
binary graph. We divide $I(x, y)$ into $m \times n$ blocks, and the
size of each block is $h \times w$, where $h = H/m, w = W/n$.
The block energy diagram $B(x, y), i \in [1, w], j \in [1, w]$ is ob-
tained by the following calculation formula:

$$B(i, j) = \frac{1}{h \times w} \sum_{x=1}^{(i-1)h+1} \sum_{y=1}^{(j-1)w+1} I(x, y).$$

In this study, we choose the radial basis function according to experience. We model the regression machine
mapping relationship according to the following formula.

$$f(x) = <w, x> + b,$$

where $w$ is the weight and $b$ is the bias term. The function is solved by the following optimization process:

$$\min_{w, b, \xi_i, \xi^*_i} \frac{1}{2} ||w||^2 + C \sum_{i=1}^l (\xi_i^* + \xi^*_i),$$

s.t. $$-\epsilon - \xi^*_i \leq y_i - <w, x_i> - b \leq \epsilon + \xi^*_i,$$

$$\xi^*_i, \xi^*_i \geq 0, \quad i = 1, 2, \ldots, l,$$

where $\epsilon$ represents the parameter of insensitive loss function, and the value of $\epsilon$ affects the number of support vectors. To measure the deviation degree of training samples out of $\epsilon$
insensitive band, we introduce a non-negative relaxation variable $\xi^t$, $\xi^-$. $C$ is a regularization parameter, which is mainly used to control the punishment of samples exceeding the error.

3.2. Human Motion Segmentation. To prepare for motion recognition and retrieval, motion segmentation entails determining the transition frame from captured motion data, dividing the motion data of different actions on both sides and then dividing the entire motion sequence into several segments. The first step in analyzing a motion sequence is to segment it. Turning point detection and clustering are two of the most commonly used methods. Based on previous research findings, an adaptive segmentation algorithm is proposed in this study. Figure 2 shows the specific procedure. The similarity matrix is established using the maximum mean square difference algorithm to calculate the similarity of each segment in the initial segmentation. The principal components of the similarity matrix are then extracted using the PCA (principal component analysis) algorithm to estimate the number of motion types contained in the motion sequence. Periodic motion detection is the third step. The periodic submotion of each motion is segmented, and the segmentation result is further optimized using the iterative motion sequence segmentation combined with the hierarchical clustering algorithm.

Initial segmentation is an important step of this segmentation algorithm, which lays an important foundation for determining the number of motion categories in the next step. To realize this function, we use the probabilistic PCA algorithm to detect turning points in motion sequences. First, we regard the previous $K$ frame as a Gaussian distribution and use the Mahalanobis distance to calculate the similarity between the distribution of segments from $K+1$ frame to $K+T$ frame and this Gaussian distribution.

The following equation is the expression of Mahalanobis distance. Markov distance is often used to measure the difference between classes.

$$M = \frac{1}{T} \sum_{i=T+1}^{K+T} (x_i - \overline{X}_N)^T S_{X_N}^{-1} (x_i - \overline{X}_N),$$

where $\overline{X}_N$ represents the average value of the previous $K$ frames, $S_{X_N}$ represents the covariance matrix, $X = \{x_{k+1}, x_{k+2}, \ldots, x_{k+T}\}$ represents the adjacent sequence of length $T$, and $H$ represents Hilbert control transposition.

After the initial segmentation, the initial motion sequence will be divided into several segments. In this study, the maximum mean difference algorithm will be used to describe the similarity between segments, and the principal component analysis will be carried out with PCA to estimate the corresponding number of motion categories.

Any segment can be represented by the feature vector corresponding to its similarity matrix. To find the corresponding eigenvector, we use formula (6) to perform singular value decomposition on the similarity matrix $L_{MMD}$. When each column vector in the matrix is projected into the feature space composed of the first $r$ feature vectors corresponding to the similar matrix, there is a certain mapping error, and the expression is as follows:

$$e = \sum \|x - x'\|^2 = \sum_{j=r+1}^{\infty} \lambda_j^2.$$  \hspace{1cm} (6)

The formula of information retention rate is as follows:

$$E_r = \frac{\sum_{j=1}^{\infty} \lambda_j^2}{\sum_{j=1}^{\infty} \lambda_j^2}.$$  \hspace{1cm} (7)

In the principal component analysis, the value of $r$ is determined by minimizing the corresponding value of $\tau$ when $E_r > \tau, \tau < 1$ holds.

Sometimes, people will run past the camera quickly, so that there will be virtual shadows in a certain frame of image, which will greatly affect the detection results. To avoid this phenomenon, some cameras will increase the exposure speed of the shutter. Therefore, under the same aperture, the amount of incident light will decrease, and the whole image will be dark. As a result, moving characters in low gray areas may be missed, resulting in detection errors.

To solve this problem, this study puts forward the concept of extraction function. Let the frame to be detected containing moving characters be $A(i, j)$, and the background of the gray image be $B(i, j)$, where $(i, j)$ represents the coordinates of pixel points, $0 \leq A(i, j) \leq 255$, and $0 \leq B(i, j) \leq 255$. Then, the extraction function is defined as follows:

$$E(i, j) = 1 - \frac{2 \sqrt{A(i, j) \times B(i, j)}}{A(i, j) + B(i, j)}.$$  \hspace{1cm} (8)
The binary output function $M(i, j)$ based on the extraction function $E(i, j)$ is defined as follows:

$$M(i, j) = \begin{cases} 1, & E(i, j) \geq r, \\ 0, & E(i, j) < r. \end{cases}$$ (9)

If $M(i, j) = 1$, it means that the current pixel $(i, j)$ belongs to a moving person. If $M(i, j) = 0$, it means that the current pixel $(i, j)$ belongs to the background area.

The adaptive matching window’s design actually intercepts a portion of the original image that contains all of the tracked moving target’s features. As a result, feature extraction and feature matching are only performed on the images in the matching serial port, rather than on the entire field of view image, which not only reduces the amount of data to be processed but also reduces other interference factors outside the adaptive window.

The size of the matching window is the most important factor affecting the speed and accuracy of matching moving objects. It is divided into two types: fixed size matching windows and adaptive matching windows that change in size as the moving target moves. As the speed of the moving target affects the prediction accuracy of the moving target during the moving process of the target, this study proposes an adaptive matching window that varies with the moving speed of the moving target to ensure that the calculation amount is as small as possible while still meeting the accuracy and improve the system’s anti-interference ability.

Let the size of the moving target image be $O_X \times O_Y$ and the size of the adaptive matching window be $D_X \times D_Y$. They have the following relationships:

$$D_X = \alpha \times O_X, \quad (10)$$

where $\alpha$ is a function of the speed of the moving target in the horizontal direction, and $\alpha = f(v_1) \geq 1$.

Similarly, for vertical direction,

$$D_Y = \alpha \times O_Y. \quad (11)$$

3.3. Moving Target Recognition. Motion recognition technology can also be used in video retrieval, reducing people’s workload significantly. Motion recognition technology has a wide range of applications in the field of auxiliary medical care. Detecting and recognizing the patient’s movements can help determine the severity of the injury. Furthermore, the distinction between calculation and standard movements can be an important foundation for the patient’s rehabilitation. Human motion recognition research in China is currently in the exploratory stage, and its enormous commercial and social potential value is yet to be realized.

The classification decision-making is the key link of the classifier system and the final link of the pattern recognition system, and the classification decision-making is the machine that finally realizes it. The use of classifiers to calculate the similarity between test samples and training samples according to certain decision rules and to classify the patterns to be tested into a certain category is referred to as classification decision. This study has spent a lot of time researching how to extract useful human motion features. We chose the template matching classifier with mature design, broad application range, and simple experimental operation to reduce the complexity of the recognition process and improve recognition speed.

This study adopts the method of machine learning based on HMM (hidden Markov model) for motion recognition. Motion recognition is divided into two stages: training stage and test recognition stage. In the training stage, the HMM model corresponding to each movement is trained. In the recognition stage, the probability that the motion sequence to be recognized belongs to each model is calculated, in which the motion corresponding to the maximum probability is the recognition result. The specific content of feature representation based on global information is shown in Figure 3.

First, a coordinate system is established with the root node as the origin, and the characteristics of each joint point in the whole motion segment are expressed as Euler angles in the coordinate system. Second, the included angle interval $[0, 180]$ between the joint point and the coordinate axis is divided into seven subintervals. Then, the Euler angles of each joint point are mapped to the corresponding subintervals, and the probability of the subintervals on the corresponding Gaussian distribution is calculated. Third, according to the principle of independence, the product of Gaussian probability weights of each joint point in each coordinate axis direction is taken as a feature component of the frame where the joint point is located.

This classifier does not require the uniform measurement standard and feature type of each input feature when parallel combination feature vectors are input, and it can also deal with a single feature. As a result, two different types of extracted motion features are fused in this topic, and this classifier is then used to achieve simple and efficient identity recognition of human motion features. The purpose of the nearest neighbor fuzzy classifier is to combine multiple classifiers for classification in a direct and simple manner. The following conditions must be met for fuzzy distribution to minimize the impact of feature extraction errors on recognition results: when the value of the feature difference is small, the curve gradually decreases, and when the value of the characteristic difference is large, the curve rapidly drops. The method of descending ridge distribution can then be used to calculate the similarity between the features of the same dimension. The formula of this fuzzy distribution function is as follows:

$$\mu(x) = \frac{1}{2} - \frac{1}{2} \sin \left( \frac{\pi}{a_2 - a_1} \left( x - \frac{a_1 + a_2}{2} \right) \right), \quad (12)$$

where $x$ represents the magnitude of the characteristic difference.

The membership degree of different features in the test sample relative to each training sample in the template library is obtained using the above formula, and then the membership degree matrix $\mu$ between the features of the same dimension is further constructed by the fuzzy distribution function.
When classifying and identifying, the membership degree of each different feature belonging to each category should be multiplied with the weight. Then, add them to get the membership matrix with weighting coefficient, which can make the recognition result more accurate and fully reflect the different contributions of different features to the recognition effect.

Combined with the fuzzy feature decomposition method, multiscale decomposition and information fusion are carried out on the moving character footprint image, and the pixel value \( H(v) \) of the moving character footprint feature distribution is as follows:

\[
H(v) = L_{xx}(x, \sigma) + \frac{1}{\sqrt{L_{yy}}} \exp \left( \frac{j\pi}{4} \right) \exp \left( -j\pi \frac{s(t)}{L_{yy}} \right). \tag{14}
\]

Among them, \( L_{xx}(x, \sigma) \) is the ambiguity function of multiscale transformation, and \( L_{xx}, L_{yy} \) is the wavelet high-frequency coefficient and low-frequency coefficient of the feature decomposition of a moving image.

The coding of this study adopts the principle of vector quantization, i.e., the pose data frame is mapped to the code word closest to it. Here, the code word is the class center obtained by clustering, and each code word exists in the form of characters in the codebook.

\[
c_{i,j} = \begin{cases} 
1, & \text{if } j = \arg \min_{j=1,2,\ldots,M} \| P_i - f_j \| \\
0, & \text{else}
\end{cases} \tag{15}
\]

Here, \( c_i \) represents the weight of the corresponding code sub-\( f_j \), \( c_i = 1 \), and the data frame will be represented by the characters corresponding to the code.

Calculate the probable probability \( P(O|\lambda_i), i = 1, 2, \ldots, \text{num} \) of this observation sequence similar to the HMM model of each movement, where \( \text{num} \) is the number of models, i.e., the number of types of movements.

Through the calculation of the following formula, \( P_i \) with the highest probability is obtained, and then \( \lambda_i \) is the HMM model of human motion closest to the motion to be identified.

\[
P_i = P_r(S|\lambda_i), \quad \text{Decision} = \arg \max_{i=1,2,\ldots,\text{num}} \{ P_i \}. \tag{16}
\]

4. Results Analysis and Discussion

In this study, we use the commonly used MAE to measure the performance of the head pose estimation algorithm. We
list the estimated performance data of different head pose estimation methods on the dataset in Table 1.

As given in Table 1, the MAE of literature [16] in Yaw direction is 6.55 and that in Pitch direction is 6.68, however, the methods proposed in this study all get lower MAE compared with it. It can be found that the performance of this method is similar to that of its estimation in YAW direction, however, the performance of estimation in the Pitch direction is better, and our method is superior to other methods in general.

To accurately evaluate the performance of pose estimation, many published methods of head pose estimation mainly use the marked face images for pose estimation and do not study the influence of face positioning offset on head pose estimation. Figure 4 shows the average MAEs of this experiment.

Compared with reference [16], their reported MAE is 10 in the YAW direction at 15% offset, while our method gets a lower MAE of 8.6 at the same offset. Through analysis, the method of face location and the description of block energy diagram proposed in this chapter are more robust to face location offset.

To test the robustness of this motion segmentation algorithm, we have done a series of experiments on synthetic data. We added different levels of noise to the synthesized data and made experiments on the synthesized data with noise. The expression of noise level: 0.1 means that one frame of noise data is inserted into every 10 frames of data. Each test data sequence is randomly synthesized by them. Therefore, each test sequence can be divided into several categories, as shown in Figure 5.

The motion segmentation algorithm in this study can accurately estimate the number of categories contained in the test sequence and has high segmentation accuracy when the noise level is low. The segmentation accuracy of the motion segmentation algorithm in this study decays slowly and stays at a reasonable level as the noise level increases. The segmentation accuracy of literature [13] and literature [21] is lower than that of the algorithm in this study. This algorithm can estimate the number of motion categories in a motion sequence without knowing the number of motion categories in the test sequence. This algorithm, on the other hand, can detect periodic motion in each motion. As periodic motions contain all of the action’s information, the action’s information can be obtained by analyzing one of the periodic motions, greatly reducing information redundancy.

As given in Table 2, compared with literature [13], literature [21] has higher segmentation accuracy, which proves the feasibility of this method again. In this experiment, the average accuracy of period segmentation is 0.81. As periodic segmentation is a fine segmentation based on the previous segmentation results, it is greatly influenced by it.

The main reason is that the video data stream is greatly influenced by the illumination, which has a great influence on the segmentation accuracy. With the change of the moving object’s moving direction relative to the camera, it will affect the motion segmentation. That is to say, the same movement may be regarded as different movements because of the different movement directions relative to the camera.

![Table 1: Comparison of MAE.](image)

| Method          | MAE (yaw) | MAE (pitch) |
|-----------------|-----------|-------------|
| Literature [16] | 6.55      | 6.68        |
| Literature [19] | 6.03      | 5.56        |
| Literature [20] | 8.66      | 12.24       |
| Literature [22] | 10.16     | 13.95       |
| Our             | 6.17      | 4.33        |

![Figure 4: Methods of MAE migration in this study.](image)

![Figure 5: Experimental results of segmentation with different algorithms.](image)

| Method          | Accuracy |
|-----------------|----------|
| Literature [13] | 0.93     |
| Literature [21] | 0.88     |
| Our             | 0.81     |

![Table 2: Experimental results of algorithm segmentation for bone data and video stream data.](image)
In the experiment, the transition interval that we regard as a new category is not included in the calculation of segmentation accuracy. In this database, the unit period of a single motion of objects generally lasts from 165 frames to 330 frames. Because of the reduction of their frames in this method, we set $n = 20, 30, 40$ in the experiment. We made experiments on different $n$ and analyzed the experimental results, as shown in Figure 6.

With the increase of $n$, the time for the algorithm to segment the same test sequence gradually increases. When $n = 20$ is used, the algorithm runs in the shortest time. The accuracy of period segmentation is obviously low because the motion sequence is long, and the calculation is based on manual segmentation, which inevitably leads to errors in statistics.

Periodic motion segmentation is only an extension of motion segmentation. In the experiment, the final segmented periodic motion contains between 20 and 40 frames, which can already reflect the inherent attributes of a motion.

Figure 7 shows the curve of the extraction function $E(i, j)$ in two cases of $B(i, j) = 4, 40$.

It can be seen that when the gray value changes very little, the change can be detected in the low gray level region, while it is likely to be ignored in the middle and high gray level region. It shows that the extraction function can adaptively adjust the sensitivity of image difference according to the gray level of pixels, which cannot be achieved by a simple frame difference algorithm.

We applied three classifier features to form three gait recognition algorithms. Then, we compared these three methods with other latest related algorithms, as shown in Figure 8.

The data analysis shows that the algorithm in literature [16] only recognizes a single motion feature, and the recognition rate obtained is not ideal. The recognition rate can be effectively improved by combining multiple features for recognition. Experiments show that the algorithm can achieve good classification and recognition results while maintaining low computational complexity and extracting less feature data. It can also organically integrate the static and dynamic features of human walking. The algorithm in this study assigns appropriate weights to two different features before performing fusion classification, which improves the recognition rate significantly. Experiments show that this algorithm can achieve better classification and recognition results by ensuring low computational complexity and extracting less feature data, as well as organically integrating the unique static and dynamic features of human walking.

5. Conclusion

This study proposes an image recognition-based feature analysis method for moving people for video images. The
motion sequence is divided into several segments in this study. The maximum mean difference algorithm is then used to measure segment similarity, and the matrix describing segment similarity is decomposed using the principal component analysis algorithm before determining the number of motion categories included in the motion sequence. Finally, to achieve fine segmentation, the hierarchical clustering algorithm is used to segment the periodic motion in each motion. After fusing the weights from the previous steps into a membership matrix, the decision-making principle is determined as the condition with the largest membership, and the result is obtained. Experiments show that for motion sequences with an unknown number of motion categories, the motion segmentation algorithm proposed in this study has a good segmentation effect and robustness.

Although a three-dimensional human model can be obtained through conversion, a significant amount of information is lost in the process, resulting in severe local distortion of the model. To improve model quality in the future, we will need to find plug-ins that better meet model format requirements.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by Zhejiang University Virtual Simulation Construction Project (Basketball referee simulation training).

References

[1] J. Zhang, W. Feng, T. Yuan, J. Wang, and A. K. Sangaiah, “SCSTCF: spatial-channel selection and temporal regularized correlation filters for visual tracking,” Applied Soft Computing, vol. 118, Article ID 108485, 2022.

[2] J. Zhang, J. Sun, J. Wang, Z. Li, and X. Chen, “An object tracking framework with recapture based on correlation filters and Siamese networks,” Computers & Electrical Engineering, vol. 98, Article ID 107730, 2022.

[3] J. Chen, C. Du, Y. Zhang, P. Han, and W. Wei, “A clustering-based coverage path planning method for autonomous heterogeneous UAVs,” IEEE Transactions on Intelligent Transportation Systems, pp. 1–11, 2021.

[4] W. Cai, M. Gao, Y. Jiang, and X. Gu, “Hierarchical domain adaptation projective dictionary pair learning model for EEG classification in IoMT systems,” IEEE Transactions on Computational Social Systems.

[5] J. Kong, H. Wang, C. Yang, and X. B. Jin, “A spatial feature-enhanced attention neural network with high-order pooling representation for application in pest and disease recognition,” Agriculture, vol. 12, no. 4, 2022.

[6] Y. Ding, Z. Zhang, X. Zhao, D. Hong, W. Cai, and C. Yu, “Multi-feature fusion: graph neural network and cnn combining for hyperspectral image classification,” Neurocomputing, 2022.

[7] L. Zhou, C. Zhang, Z. Wang, Y. Wang, and Z. M. Lu, “Hierarchical palmprint feature extraction and recognition based on multi-wavelets and complex network,” IET Image Processing, vol. 12, no. 6, pp. 985–992, 2018.

[8] P. Das and D. Arpita, “Shift invariant extrema based feature analysis scheme to discriminate the spiculation nature of mammograms,” ISA Transactions, vol. 103, pp. 156–165, 2020.

[9] Z. Peng, J. An, X. Gui, Z. Wang, and W. Zhang, “Location correlated differential privacy protection based on mobile feature analysis,” IEEE Access, vol. 7, Article ID 54483, 2019.

[10] I. W. Tcho, W. G. Kim, and Y. K. Choi, “A self-powered character recognition device based on a triboelectric nanogenerator,” Nano Energy, vol. 70, Article ID 104534, 2020.

[11] G. Li, J. Zhang, D. Chen, and Z. Liu, “Chinese flower-bird character generation based on pencil drawings or brush drawings,” Journal of Electronic Imaging, vol. 28, no. 3, p. 1, 2019.

[12] K. Shao, Q. Fan, Y. Zhang, F. Bao, and C. Zhang, “Noisy single image super-resolution based on local fractal feature analysis,” IEEE Access, no. 99, p. 1, 2021.

[13] X. Gao and Y. Shardt, “Dynamic system modelling and process monitoring based on long-term dependency slow feature analysis,” Journal of Process Control, vol. 105, pp. 27–47, 2021.

[14] L. Tu and J. M. Kim, “Discriminative feature analysis based on the crossing level for leakage classification in water pipelines,” Journal of the Acoustical Society of America, vol. 145, no. 6, pp. EL611–EL617, 2019.

[15] M. S. Chen, C. P. Hwang, T. Y. Ho, H. F. Wang, C. M. Shih, and H. Y. Chen, “Driving behaviors analysis based on feature selection and statistical approach: a preliminary study,” The Journal of Supercomputing, vol. 75, no. 4, pp. 2007–2026, 2019.

[16] M. Xu, J. Shen, Y. Huang, and Y. N. Xu, “Weighting inversion of dynamic light scattering based on particle-size information distribution character,” Acta Physica Sinica -Chinese Edition-, vol. 67, no. 13, 2018.

[17] X. Xu and J. Ding, “Decentralized dynamic process monitoring based on manifold regularized slow feature analysis,” Journal of Process Control, vol. 98, pp. 79–91, 2021.

[18] A. Sl. A. Ct, C. B. Yang, and T. Lan, “Dynamic statistical process monitoring based on online dynamic discriminative feature analysis,” Journal of Process Control, vol. 103, pp. 67–75, 2021.

[19] G. Feng, H. Li, J. Dong, and Z. Jiashu, “Face recognition based on volterra kernels direct discriminant analysis and effective feature classification,” Information Sciences, vol. 441, pp. 187–197, 2018.

[20] A. Xw, L. B. Ying, and O. A. Jie, “A meshfree collocation method based on moving Taylor polynomial approximation for high order partial differential equations,” Engineering Analysis with Boundary Elements, vol. 116, pp. 77–92, 2020.
component analysis,” *IEEE Transactions on Control Systems Technology*, vol. 28, no. 99, pp. 583–592, 2020.

[22] J. Lu, W. S. Lee, G. Hao, and X. Hu, “Immature citrus fruit detection based on local binary pattern feature and hierarchical contour analysis,” *Biosystems Engineering*, vol. 171, pp. 78–90, 2018.

[23] W. Binsaeedan and S. Alramlawi, "CS-BPSO: hybrid feature selection based on chi-square and binary PSO algorithm for Arabic email authorship analysis,” *Knowledge-Based Systems*, vol. 227, no. 2, Article ID 107224, 2021.