

A Multi-Feature Motion Posture Recognition Model Based on Genetic Algorithm

Yuanguo Liu, Ying Wu*

Sports Training Institute, Shenyang Sport University, Shenyang 110102, China

Corresponding Author Email: wuying@syty.edu.cn



<https://doi.org/10.18280/ts.380307>

ABSTRACT

Received: 5 February 2021

Accepted: 8 May 2021

Keywords:

motion posture recognition, multi-feature, genetic algorithm (GA), visual background extractor (ViBe) algorithm

The effect of motion posture recognition hinges on the accurate description of motion postures with effective feature information. This study introduces Wronskian function to improve the denoising ability of visual background extractor (ViBe) algorithm, and thus acquires relatively clear motion targets. Then, a multi-feature fusion motion posture feature model was developed based on genetic algorithm (GA). Specifically, GA was called to optimize and fuse the extracted feature information, while a fitness function was constructed based on the mean variance ratio, and used to select the feature information with high inter-class discriminability. Taking support vector machine (SVM) as the classifier, a multi-class classifier was designed by one-to-one method for the classification and recognition of motion postures. Through experiments, our model was proved highly accurate in motion posture recognition.

1. INTRODUCTION

To recognize human motion postures [1], it is important to analyze videos and images through image processing, so that the computer can parse the postures captured by cameras. In this way, human motion information can be obtained comprehensively and effectively, facilitating subsequent analysis on behaviors and actions. Motion posture recognition plays an important role in the intelligent analysis of videos, and provides a highly practical application field of computer vision.

Computer vision relies on cameras and computers instead of humans to perceive and process the visual information of objects, with the aid of advanced image processing techniques. The processed images facilitate subsequent analysis by human eyes or machines. Recognition is a classic problem in both image processing and machine vision. It intends to determine whether an image set contains a specific object, image feature, or motion state. According to the common definition, recognition aims to extract the features and spatial information from one or a class of objects, and identify the object(s) through the predefinition or learning of the features.

Human motion postures generally include walking, running, jumping, squatting, etc. These postures not only reflect the body state in sports, but also convey the purpose of behaviors and emotional responses. The recognition of human motion postures is premised on the detection of moving objects, and the feature extraction of motion postures. By analyzing the extracted features, it is possible to automatically classify and recognize human motion postures.

Over the years, the recognition technology of human motion postures has developed continuously, and penetrated more and more fields. In sports analysis, athlete videos are often processed by intelligent computer vision techniques, which yield valuable sports data. The objective and efficient analysis of these data help athlete quickly identify training problems, and master the essentials of sports actions. In addition, sports

video analysis assists referees in competitions, and supports strategy and tactic evaluation in team events. The relevant techniques can extract and track moving objects in videos, and recognize motion postures from videos.

This paper designs a multi-feature motion posture recognition model based on videos, with the aid of visual background extractor (ViBe) algorithm. The core functions of the model are detecting human motion postures, extracting features from motion postures, and fusing different features.

2. LITERATURE REVIEW

The first step of human posture recognition is to acquire the postures. Human postures can be recognized by two types of techniques: the contact recognition techniques requiring the subject to wear devices or get marked [2], and the non-contact techniques eliminating these requirements [3]. The commonly used contact techniques are mechanical, optical, or electromagnetic, while the prevalent non-contact technique is computer vision. Gao et al. [4] designed a mechanical motion capture system. In the system, the object needs to wear mechanical devices connected to sensors. The motions and body state of the object are measured by the sensors in real time. The system provides a low-cost solution with high accuracy and good real-time performance. However, the mechanical devices are very likely to impede the motions of the subject. Choe et al. [5] developed an optical motion capture system capable of collecting rich multi-dimensional information. The system can capture real-time motion postures quickly. Liu and Chan [6] proposed a wearable bioelectrical signal sensor to gather accurate and complete data, laying a solid basis for subsequent posture recognition.

With the aid of computer vision, human motion capture techniques can obtain video information via cameras, and identify human motion postures in videos or images. Swamy and Reddy [7] derived a human pose estimation framework,

which recognizes body parts accuracy by representing the body in multiple models. To estimate human postures, Huang et al. [8] put forward a novel bottom-up approach to learn scale sensing representation by pyramid with high resolution features. Their approach can be implemented in two stages: training through multi-resolution supervision, and reasoning through multi-resolution fusion. Through the two stages, the poses of multiple people can be estimated in the presence of scale change, and the key points can be located very accurately. St-Pierre [9] proposed distributed sensing coordinates for posture estimation. Specifically, the mean position of real Gaussian distribution is estimated from the maximum value on heatmap and its position, thereby minimizing the quantization error. He et al. [10] proposed self-supervised three-dimensional (3D) pose estimation model. In this model, 3D human poses are estimated based on transient images, i.e., 3D spatiotemporal histograms of photons, which are obtained by optical non-line of sight (NLOS) imaging system. Despite the above works, the computer vision-based human motion capture techniques should be improved in terms of cost, convenience, and transmission speed, such as to meet the requirements of various fields of production and life. Considering the advantages of videos captured by ordinary cameras, it is promising to capture and analyze human motions through intelligent video analysis.

Human motion postures can be described effectively based on feature information. The feature descriptor can characterize motion postures with a high accuracy. Of course, the feature information should not only reflect the similarity between motion postures in the same class, but also reveal the distinction between those in different classes. The feature extraction of human motion postures can be categorized into global feature extraction and local feature extraction [11]. Global feature extraction describes human regions of interest (ROIs) with contours, optical flows, and moments. They are usually obtained through moving object tracking or foreground detection. The global features are highly invariant, simple to compute, and offer intuitive and rich representations. Local feature extraction depicts the ROIs in the original video, without needing to tracking moving objects or model any trajectories. Local features are insensitive to complex background, illumination change, visual angle variation, or partial occlusion. However, they cannot provide the complete information of moving objects. Akiduki et al. [12] represented motion postures with a continuous state sequence, and developed a global feature extraction method for motion posture recognition: the human motions are classified by representing the state changes with probability relationship. Etemad and Arya [13] characterized motion postures with a set of templates, which includes the known video sequence of human motion postures and the features extracted from the sequence, matched the target motion posture templates with the known templates by similarity, and developed a global feature extraction strategy for recognizing human motion postures. To identify human motion postures, De Silva et al. [14] presented a local feature extraction method based on posture features and grammatical information, and relied on the method to acquire high-level semantics and obtain an abstract description of motion postures.

3. IMPROVED VIBE COUPLING WRONSKIAN FUNCTION

3.1 ViBe algorithm

ViBe algorithm Barnich and Van Droogenbroeck [15] models the background and detects foreground based on the spatiotemporal levels of pixels. The background model is established and updated through random update, and neighborhood propagation, which ensure the detection accuracy and speed of the algorithm. The algorithm can be divided into three parts: initializing background model, detecting foreground, and updating background model.

(1) Initializing background model

Each pixel of each frame is modeled iteratively through the following process. First, the background model is initialized with the first frame of the source video:

$$M(x) = \{v_1, v_2, \dots, v_N\} \quad (1)$$

where, $v(x)$ is the value of background pixel at x ; $M(x)$ is the set of background samples at x .

(2) Detecting foreground

Each new pixel value $v(x)$ is compared with the set of background samples $M(x)$ to see if it belongs to the background. If it is close to a sample value in the set, then the pixel must belong to the background. The judgement criteria can be described by:

$$v(x) \in \begin{cases} \text{foreground} & \text{if } |B_R(v(x) \cap M(x))| < T_{\min} \\ \text{background} & \text{otherwise} \end{cases} \quad (2)$$

where, $B_R(v(x))$ is the sphere with center $v(x)$ and radius R ; T_{\min} is the decision threshold. If the number of $M(x)$ in the sphere is fewer than T_{\min} , then point x must belong to the foreground; otherwise, that point must belong to the background.

(3) Updating background model

Different from conventional approaches, ViBe algorithm updates the background model randomly by a simple strategy:

Firstly, the sub-templates are updated by random. As new samples are continuously added to the background model, the previous samples are randomly removed from the model. Let N be the number of samples in background model $M(x)$ at a moment. Then, the probability for a previous sample to be retained at that moment is $(N - 1)/N$. With the elapse of time, the probability for a previous sample to be retained at $t + dt$ can be calculated by:

$$P(t, t + dt) = ((N - 1)/N)^{(t+dt)} \quad (3)$$

Formula (3) can be rewritten in the exponential form:

$$P(t, t + dt) = e^{-\ln((N-1)/N)dt} \quad (4)$$

Formula (4) shows that the number of samples in the background model decreases smoothly in their lifecycle, i.e., the retention probability decreases with time.

Secondly, if the sample size is small, a second sampling factor is introduced to lower the frequency of background updates, and extend the lifecycle of background samples. In other words, when there are a few samples, the time window is widened for the fixed size background model, without sacrificing algorithm accuracy.

The updates of background model can be mathematically described as follows: Firstly, a random frame is selected from N background models, and denoted as image I_G . Next, a random point x is taken from the frame. For the pixels in x and its eight neighborhoods, when a new frame I_T is imported, the pixel $I_T(x)$ corresponding to x in I_T is judged as the background, calling for the update of I_G . By this eight-neighborhood update strategy, any error sample can propagate through the model, such that the neighborhood pixels will not match the incorrectly updated samples. In this way, the subtle interference and noise can be removed from the video, making the detection more accurate.

3.2 Improved ViBe algorithm

One of the advantages of ViBe algorithm is the ability to initialize the background model rapidly, using only the first frame of the video sequence. However, this might lead to ghosting problem.

As mentioned before, the original ViBe algorithm can initialize the background model with only the first frame, and select the pixels through eight-neighborhood random sampling. The algorithm is robust in detecting moving objects from complex scenes, owing to the absence of parameter setting and small memory occupation.

However, ViBe algorithm could easily mistake a foreground pixel as part of the background, if the initial frame contains a moving object, or if the state of the moving object changes abruptly. Then, the detection results will suffer from the ghosting problem. The ghosting areas will remain in the background for a long time, and undermine the detection accuracy of moving objects.

To mitigate the problem, this paper tries to improve the ViBe algorithm with Wronskian function, and fill the ghosting frames by the seed filling algorithm. The improved ViBe algorithm could detect moving objects from the video sequence without being affected by the ghosting problem. The workflow of the improved algorithm is explained in Figure 2.

The moving object detector based on Wronskian function [16] is a vector image model. It is often adopted to detect the pixel variation between the same position of two adjacent frames. This paper assumes that each pixel in a frame is correlated with its adjacent pixels, and represents it by a vector (Figure 1) composed of the center pixel and its adjacent pixels.

To detect the change of pixel position between two frames, a linear independent test needs to be conducted on the support region (Figure 1). The pixels in the same positive must have changed, if they are linearly independent of the corresponding pixels in the adjacent frames. The linear correlation or independence of vectors can be determined by Wronskian function.

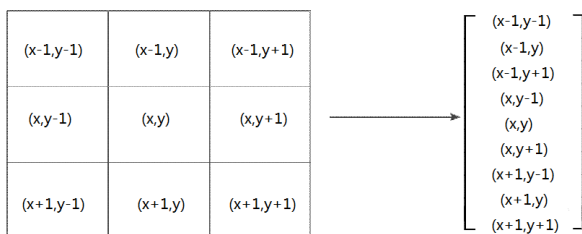


Figure 1. Support region of pixel (x, y)

During the detection of the variation between two adjacent frames, the determinant of the Wronskian matrix is zero, if the two pixels have a linear correlation:

when the two pixels are linearly correlated.

$$|Q| = \begin{vmatrix} G_t(x,y) & G_{t-1}(x,y) \\ G'_t(x,y) & G'_{t-1}(x,y) \end{vmatrix} = 0 \quad (5)$$

where, $G_t(x,y)$ and $G_{t-1}(x,y)$ are gray values at t and $t-1$, respectively.

The moving object detector based on Wronskian function was applied to the pixel region with spatial temporal attributes. At a specific position (x, y), the variation between the t -th frame and the $(t-1)$ -th frame can be detected by Wronskian function:

$$|Q| = \frac{1}{n} \sum_{i=1}^n \left(\left(\frac{F'_t(x,y)_i}{F'_{t-1}(x,y)_i} \right)^2 - \frac{F'_t(x,y)_i}{F'_{t-1}(x,y)_i} \right) \quad (6)$$

where, n is the number of pixels in the vector image; $1/n$ is a normalizer to ensure that the same threshold applies to different vector dimensions.

Seed filling algorithm chooses a point from a known image region as the seed (x, y), detects the color of the seed, and compares it with the boundary color and filling color. If the three colors are not the same, the point will be filled with the filling color. Then, the comparison and color filling will be carried out on the adjacent position. This process is repeated until all the pixels in the neighborhoods of the image region have been processed.

By filling area and search direction, seed filling algorithm can be divided into 4-interconnection algorithm and 8-interconnection algorithm [17]. This paper selects the more efficient 4- interconnection algorithm to fill the ghosting areas, for the following considerations: the recognition of human motion postures should be computationally efficient and good in real-time performance, and morphological processing could reduce the miss rate of actual filling more effectively than boundary crossing.

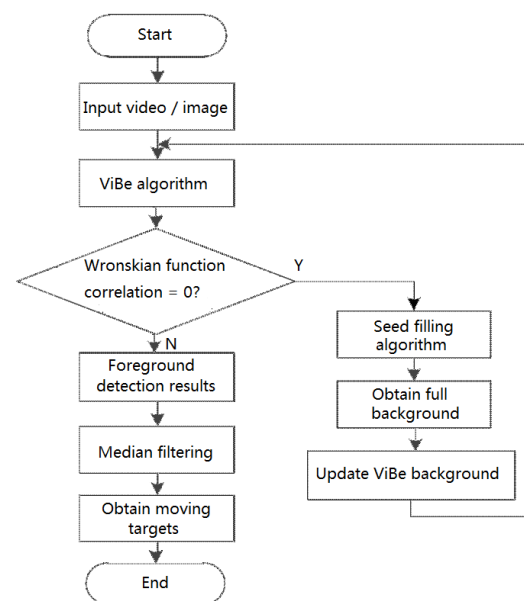


Figure 2. Workflow of the improved ViBe algorithm

As shown in Figure 2, the flow of the improved ViBe algorithm can be defined as follows:

Step 1. Input the source video or frame, and detect moving objects with the original ViBe algorithm.

Step 2. Judge the correlation and pixel value by Wronskian matrix.

The matrix evaluates the pixel variation with an interval of five frames suffering from ghosting problem. These frames are selected through repeated tests on different videos. Since the ghosting area is fast and still in each frame, the motion of each pixel can be judged by the Wronskian function, and the ghosting area can be identified based on pixel value.

Step 3. Fill the detected ghosting area by 4-interconnection algorithm, so that the whole video sequence is free of ghosting.

Step 4. Take the pixels in the ghosting area as background pixels to update the ViBe template, and thus suppress the influence of ghosting on moving object detection.

Through the above steps, it is possible to get a clear moving target, reduce the interference of noise and non-fuzzy edges, and provide accurate information for subsequent processing. The improved ViBe algorithm can accurately recognize moving targets in complex background. Figure 3 shows the recognition effect of the improved algorithm for moving skiers.



Figure 3. Recognition effect of moving skiers

4. MULTI-FEATURE MOTION POSTURE RECOGNITION MODEL

Feature fusion Yeh et al. [18] and Pong and Lam [19] aims to get the optimal combination of eigenvectors of multiple types and dimensions. The fused feature group should fully demonstrate the complementarity of information, and eliminate redundant data to enhance real-time performance. The key of feature fusion is to find a subset of highly distinguishable features from the set of multi-source features.

Genetic algorithm (GA) stands out as a feature fusion method powerful in classification and dimensionality reduction. Through binary coding of chromosomes, GA can filter effective features, and reduce the dimensionality of eigenvectors. The resulting new eigenvector group helps to improve the accuracy and efficiency of moving target recognition.

As a random search method, GA Tian et al. [20] mimics the process of natural selection and genetic mutation in biological evolution. The initial population is a set of possible solutions for the optimization problem, which take the form of individual binary coded genes. Through selection, crossover, and mutation, new generations are derived from the initial population. In each generation, the individuals are compared by fitness, and the elites are retained in the next generation.

Since feature fusion intends to get the optimal combination

of eigenvectors, it is important to find suitable criteria to evaluate each solution, and assess the classification ability of the corresponding eigenvectors in classification and recognition. To classify different motion postures, the maximum inter-class and minimum intra-class were chosen as the criteria.

$$S_{ab} = \frac{|m_a - m_b|}{\sqrt{\varepsilon_a^2 + \varepsilon_b^2}} \quad (7)$$

where, m_a and m_b are the mean values of features of classes a and b, respectively; ε_a^2 and ε_b^2 are the feature variances of classes a and b, respectively. Formula (7) is the basis for binary classification of multiple types of features.

Let K be the class of a motion posture; S_{ab} be the distance between two classes; $n = K(K - 1)/N$ be the number of class separation values; $V = [S_1, S_2, \dots, S_n]$ be the n-dimensional vector of each individual. Then, the fitness function can be established by:

$$f = \sum_{x=1}^n S_x / n\sigma_V^2 \quad (8)$$

where, σ_V^2 is the variance of vector V.

If the eigenvalues of a selected individual improve the performance of both types of separation values, the individual can represent different motion postures accurately. Therefore, the mean variance ratio was adopted as the metric. The greater the mean variance ratio, the better the overall performance of vector V. The smaller the ratio, the stabler the data in V, which increases the data values in V.

The selection operation randomly chooses individuals from the current generation by fitness and the preset strategy. The selected individuals are retained, while the other individuals are eliminated. Based on the proportion of fitness, roulette selection can be defined as:

$$P_i = F_i / \sum_{j=1}^N F_j \quad (9)$$

In this strategy, an individual with a large fitness is very likely to be retained in the next generation, and could be directly adopted to solve the maximization problem.

The crossover operation swaps the corresponding genes in two parent chromosomes, in order to produce new individuals. The gene swap is similar to the mating mode in nature, so that the population is naturally diverse. This paper selects single-point crossover with a probability of 0.7.

The mutation operation changes the gene in a certain position of an individual by the mutation probability. Under binary encoding, mutation means changing the zero or one on the gene string to one or zero, in order to generate new individuals. In this paper, the mutation probability is set to 0.05.

GA has two termination conditions: First, the preset maximum number of iterations is reached. Second, the mean fitness of adjacent generations reaches the preset threshold. Therefore, the termination of GA depends on the convergence degree of the population. This paper sets the maximum number of iterations to 300.

According to the feature extractor of our motion posture recognition technique, 39 human motion posture features were

extracted from the images on the key actions, such as walking, running, and jumping. The features include the eight-star model, Hu moment invariants, Zernike moment invariants, and wavelet moment invariants. The eigenvalues were standardized and normalized for feature fusion. Then, GA-based multi-feature fusion was carried out to optimize the combination of 19 features. The fused features have a large class gap in object recognition. This facilitates classification and recognition tasks, reduces the number of features and computing load, and improves the real-time performance of motion posture recognition.

5. SUPPORT VECTOR MACHINE (SVM)-BASED MULTI-FEATURE POSTURE RECOGNITION

This paper selects SVM for learning and training, and constructs a multi-class classifier through feature fusion, so

that human motion postures can be classified and recognized by both global and local features. In this way, the recognition becomes more accurate and rapid.

At present, SVM multi-class classifier can be designed by direct method, or by one-to-one method [21]. The latter is more accurate and faster in training. Hence, this paper selects one-to-one method, and chooses radial kernel function SVM to establish a multi-class classifier for the recognition of human motion postures. Figure 4 explains the implementation process of the multi-class classifier.

In the presence of N classes of training samples, both the two possible classes of training sets were constructed, and used to generate the corresponding binary SVMs. A total of $(N \times (N-1))/2$ binary SVMs needs to be constructed. Six binary SVMs are needed for four types of samples. During the training of the sub-classifiers of classes a and b, the samples of class a were taken as positive samples, and those of class b as negative samples.

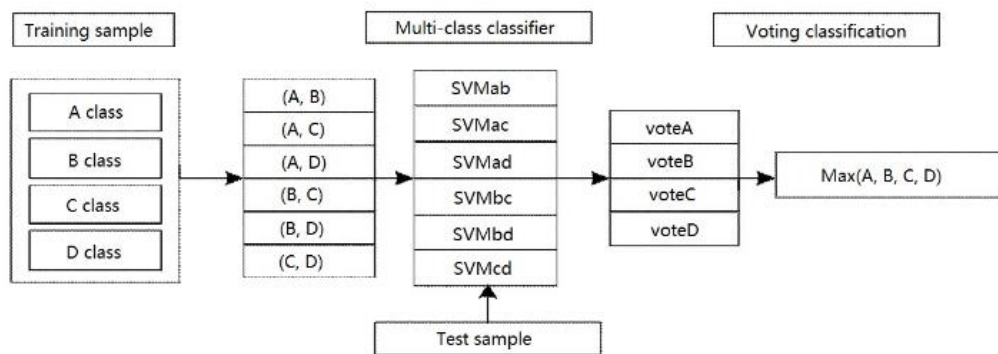


Figure 4. Implementation process of multi-class classifier

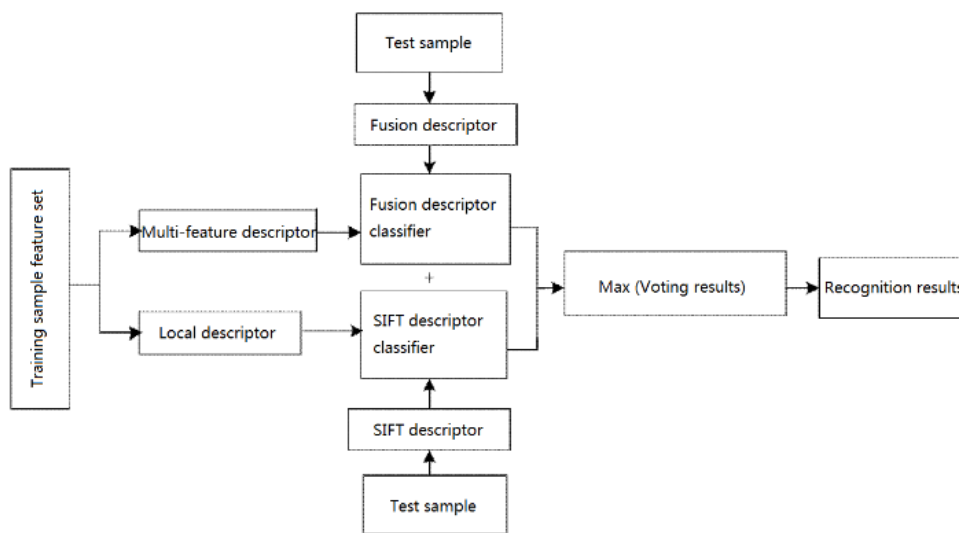


Figure 5. Structure of SVM-based multi-feature posture recognition

The test samples were imported to $(N \times (N-1))/2$ classifiers, respectively, because N classes were outputted. Thus, N classes of cumulative voting scores could be obtained. Then, each class corresponds to a sum of cumulative voting scores. After a test sample was imported, if a class was determined as the output, then the voting score of this class would be increased by 1 (sum = sum + 1). Finally, the maximum cumulative value was found, and the corresponding class was taken as the class of the test sample.

When the classification model is applied to recognize human motion postures, each feature of motion posture sample contains two independent features. Thus, it is necessary to establish a multi-class classifier of fused features. Figure 5 shows the structure of the SVM-based multi-feature posture recognition model.

The features of training samples were obtained from the standard video database. Two kinds of features could be extracted from each type of videos, namely, fused feature and

scale invariant feature transform (SIFT) feature. This paper trains the two features of training samples to generate two classification models, and imports the two features of the test samples into the two classifiers. Then, the two multi-class classifiers output their own voting scores. These results were fused into the final cumulative voting scores. Finally, the class with the maximum cumulative score was outputted.

6. EXPERIMENTS AND RESULTS ANALYSIS

The proposed multi-feature motion posture recognition model was applied to recognize three kinds of motion postures (walking, running, and jumping) in standard video databases, such as KTH database [22], Weizmann database [23], a self-built video database, and UCF-sport database (Table 1).

Tables 2 and 3 present the classification results on Weizmann and KTH databases, focusing on the three common

postures of walking, running, and jumping, respectively.

As shown in Tables 2 and 3, the proposed model greatly improved the recognition rate of the fusion multi-feature, because the features fused by GA has sufficient information and a prominent inter-class difference. But the SIFT feature was not well recognized on the two video databases, and the recognition rate of fusion + SIFT feature was not much different from that of fusion multi-feature. The reason is that the video background in the two databases is not diverse, failing to provide much information about the scenes.

Figure 6 shows the recognition effect of diving, golf, and gymnastics in UCF-sport database. Table 4 records the classification results of UCF-sport database. It can be seen that the recognition rate of Fusion + SIFT feature of UCF-sport video database was higher than that in Weizmann database and KTH database. A possible reason is the rich and diverse motion posture scenes in the former database, which provide more effective information.

Table 1. Information of video databases

Video database	Number of videos	Number of posture types	Resolution
KTH	603	6	160*120
Weizmann	80	8	180*150
UCF-Sport	150	10	720*480
Self-built video database	77	3	720*480

Table 2. Classification results on Weizmann database

Posture	SIFT feature	Multi-feature	Fusion multi-feature	Fusion + SIFT feature
Running	40.17%	93.71%	99.92%	100%
Walking	47.79%	88.72%	94.81%	98.37%
Jumping	62.82%	96.86%	100%	100%

Table 3. Classification results on KTH database

Posture	SIFT feature	Multi-feature	Fusion multi-feature	Fusion + SIFT feature
Running	45.22%	91.69%	95.72%	95.97%
Walking	57.31%	90.71%	92.57%	97.01%
Jumping	60.22%	92.81%	97.03%	99.01%

Table 4. Classification results of UCF-sport database

Posture	SIFT feature	Multi-feature	Fusion multi-feature	Fusion + SIFT feature
Diving	72.32%	90.13%	95.62%	98.99%
Golf	46.91%	70.82%	87.65%	96.91%
Gymnastics	88.31%	90.21%	95.09%	99.98%

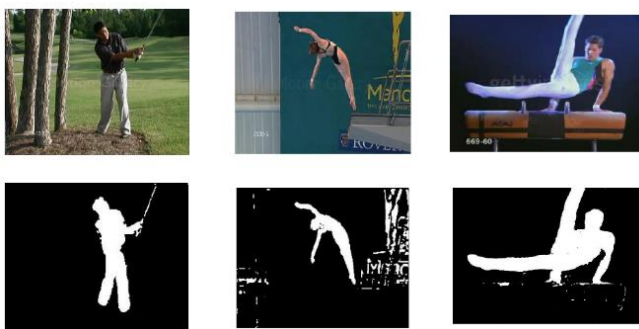


Figure 6. Recognition effect of diving, golf and gymnastics in UCF-sport database

7. CONCLUSIONS

This paper mainly studies the detection of moving objects, and the extraction of motion features, and designs a multi-feature fusion motion posture feature model based on GA. The model is dedicated to obtain clear moving objects, describe motion postures with a few effective features, highlight the disparity between postures, and improve the recognition performance in an all-round manner. Experimental results show that our model can recognize more than 95% of walking, running, and jumping postures accurately, and achieve a recognition rate of motion videos in UCF-sport video database up to 96%. Therefore, the proposed human motion posture recognition model, which relies on multi-feature fusion, is highly feasible.

ACKNOWLEDGMENT

This work was supported by Liao Ning Revitalization Talents Program (Grant No.: XLYC2007064), and in part by the Liaoning Natural Science Foundation Program Guidance Plan: Monitoring and analysis of physiological indexes of biathlon athletes by self-powered biosensor (Grant No.: 2019-ZD-0515).

REFERENCES

- [1] Sun, G., Sheng, B., Dong, L. (2018). New trend of image recognition and feature extraction technology introduction. *Traitement du Signal*, 35: 3-4.
- [2] Zhao, L., Chen, W. (2020). Detection and recognition of human body posture in motion based on sensor technology. *IEEJ Transactions on Electrical and Electronic Engineering*, 15(5): 766-770. <https://doi.org/10.1002/tee.23113>
- [3] Orenco, G., Lagati, A., Saggio, G. (2014). Modeling wearable bend sensor behavior for human motion capture. *IEEE Sensors Journal*, 14(7): 2307-2316. <https://doi.org/10.1109/JSEN.2014.2309997>
- [4] Gao, L., Zhang, G.F., Yu, B., Qiao, Z.W., Wang, J.C. (2020). Wearable human motion posture capture and medical health monitoring based on wireless sensor networks. *Measurement*, 166: 108252. <https://doi.org/10.1016/j.measurement.2020.108252>
- [5] Choe, N., Zhao, H., Qiu, S., So, Y. (2019). A sensor-to-segment calibration method for motion capture system based on low cost MIMU. *Measurement*, 131: 490-500. <https://doi.org/10.1016/j.measurement.2018.07.078>
- [6] Liu, K.C., Chan, C.T. (2017). Significant change spotting for periodic human motion segmentation of cleaning tasks using wearable sensors. *Sensors*, 17(1): 187. <https://doi.org/10.3390/s17010187>
- [7] Swamy, P., Reddy, B.A. (2018). Human pose estimation in images and videos. *International Journal of Engineering & Technology*, 7(3): 1-6. <https://doi.org/10.14419/ijet.v7i3.27.17640>
- [8] Huang, H., Kalogerakis, E., Chaudhuri, S., Ceylan, D., Kim, V.G., Yumer, E. (2017). Learning local shape descriptors from part correspondences with multiview convolutional networks. *ACM Transactions on Graphics (TOG)*, 37(1): 1-14. <https://doi.org/10.1145/3137609>
- [9] St-Pierre, N.R. (2016). Comparison of model predictions with measurements: A novel model-assessment method. *Journal of Dairy Science*, 99(6): 4907-4927. <https://doi.org/10.3168/jds.2015-10032>
- [10] He, C., Kazanzides, P., Sen, H.T., Kim, S., Liu, Y. (2015). An inertial and optical sensor fusion approach for six degree-of-freedom pose estimation. *Sensors*, 15(7): 16448-16465. <https://doi.org/10.3390/s150716448>
- [11] Lertniphonphan, K., Aramvith, S., Chalidabhongse, T.H. (2016). Human action classification using adaptive key frame interval for feature extraction. *Journal of Electronic Imaging*, 25(1): 013017. <https://doi.org/10.1117/1.JEI.25.1.013017>
- [12] Akiduki, T., Zhang, Z., Imamura, T., Miyake, T. (2012). Feature extraction of motion from time-series data by using attractors. *IEEJ Transactions on Electronics, Information and Systems*, 132(6): 975-982. <https://doi.org/10.1541/ieejieiss.132.975>
- [13] Etemad, S.A., Arya, A. (2014). Extracting movement, posture, and temporal style features from human motion. *Biologically Inspired Cognitive Architectures*, 7: 15-25. <https://doi.org/10.1016/j.bica.2013.10.001>
- [14] De Silva, P.R., Bianchi-Berthouze, N. (2004). Modeling human affective postures: An information theoretic characterization of posture features. *Computer Animation and Virtual Worlds*, 15(3-4): 269-276. <https://doi.org/10.1002/cav.29>
- [15] Barnich, O., Van Droogenbroeck, M. (2010). ViBe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing*, 20(6): 1709-1724. <https://doi.org/10.1109/TIP.2010.2101613>
- [16] Sun, G., Chen, C.C., Bin, S. (2021). Study of Cascading Failure in Multisubnet Composite Complex Networks. *Symmetry*, 13(3): 523. <https://doi.org/10.3390/sym13030523>
- [17] Bin, S., Sun, G. (2020). Optimal energy resources allocation method of wireless sensor networks for intelligent railway systems. *Sensors*, 20(2): 482. <https://doi.org/10.3390/s20020482>
- [18] Yeh, Y.R., Lin, T.C., Chung, Y.Y., Wang, Y.C.F. (2012). A novel multiple kernel learning framework for heterogeneous feature fusion and variable selection. *IEEE Transactions on Multimedia*, 14(3): 563-574. <https://doi.org/10.1109/TMM.2012.2188783>
- [19] Pong, K.H., Lam, K.M. (2014). Multi-resolution feature fusion for face recognition. *Pattern Recognition*, 47(2): 556-567. <https://doi.org/10.1016/j.patcog.2013.08.023>
- [20] Tian, G.L., Zhou, S., Sun, G.X., Chen, C.C. (2020). A novel intelligent recommendation algorithm based on mass diffusion. *Discrete Dynamics in Nature and Society*, 11: 1-9. <https://doi.org/10.1155/2020/4568171>
- [21] Qian, H.M., Mao, Y.B., Xiang, W.B., Wang, Z.Q. (2010). Recognition of human activities using SVM multi-class classifier. *Pattern Recognition Letters*, 31(2): 100-111. <https://doi.org/10.1016/j.patrec.2009.09.019>
- [22] Lindgren, B., Österlund, J.M., Johansson, A.V. (2004). Evaluation of scaling laws derived from Lie group symmetry methods in zero-pressure-gradient turbulent boundary layers. *Journal of Fluid Mechanics*, 502: 127-152. <https://doi.org/10.1017/S0022112003007675>
- [23] Nanni, L., Brahnam, S., Lumini, A. (2011). Local ternary patterns from three orthogonal planes for human action classification. *Expert Systems with Applications*, 38(5): 5125-5128. <https://doi.org/10.1016/j.eswa.2010.09.137>