Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 25, Issues 6, pp. 5351–5357, DOI 10.12694/scpe.v25i6.3320

SPORTS PLAYER ACTION RECOGNITION BASED ON DEEP LEARNING

FENG LI*

Abstract. A sports auxiliary evaluation system suitable for China's national conditions was established using big data and sports identification technology. First of all, this paper extends the data of normative behavior and constructs a normative library of scores and comparisons. The acquisition of 3D data is emphasized. The method based on Fourier descriptors is used to locate the motion accurately. In this way, better gait recognition results can be obtained. The Fourier characteristics before and after wavelet transform are compared with the actual object characteristics, and the results show that the proposed algorithm can extract the features with high precision. This scheme can obtain a more accurate identification effect. The system proposed in this paper provides a powerful means for judges to score.

Key words: Big data; Particle filter; Fourier descriptors; Feature extraction; sports

1. Introduction. Biometrics is a way to identify people based on their intrinsic physiological characteristics. The uniqueness and universality of biometrics make it a hot topic in current identity research. People must first obtain some of its characteristics, such as physiological and behavioral characteristics, to determine their true identity. Because it is not easy to camouflage, hide and be detected and identified over a long time based on human gait characteristics, it can effectively overcome the image blur caused by a long distance. Some scholars have proposed a Fourier descriptor image feature extraction algorithm, which can effectively solve the problem of matching between keyframes. However, according to the Fourier characteristics of the spatial position obtained by the human body when it moves, there are significant errors in practical application.

Recognizing human posture is a significant research direction with a broad development space in machine vision. Therefore, methods based on 3D human pose estimation and behavior recognition are widely used. The research on human motion capture, human modeling, pose estimation, and motion recognition has achieved fruitful results in the last ten years. In literature [1], an extensible graph model can effectively improve the identification efficiency of human behavior. Literature [2] provides a global descriptor that synthesizes the orientation and size of the movement of various parts of the human body. Scholars first build a three-dimensional model and then use the model to identify human behavior. Literature [3] has improved the characteristics of RGBD images. The descriptor encodes the extraction of three-dimensional directions from evenly spaced areas to realize action recognition. This method provides a new idea for machine vision research. Key research topics include advanced human-computer interaction, assisted living, gesture interactive games, intelligent driver assistance systems, movies, 3D TV and animation, physical therapy, autonomous intellectual development, intelligent environment, motor behavior analysis, video surveillance, video annotation, etc. Especially in the study of physical activity, many events, such as calisthenics, have high demands on the posture and movement of the human body.

The pose estimation method can be used to study sports performance and training in physical education. In the process of competition, it can also help the judges score [4]. Judges need to give accurate scores in the aerobics competition by evaluating the complex movements done by the players. The difficulty judge determines the difficulty of each action performed by the player. Because the problematic movements of groups A, C, and D are instantaneous, it is easy to make visual errors. Therefore, when there are different scoring levels in difficulty evaluation, the judging committee needs to use big data and behavior identification technology to design and implement a set of aerobics scoring systems. This paper proposes a gait feature extraction method based on particle filter tracking. The accurate determination of human behavior is realized through the acquisition of 3D

*Department of Public Sport, Xuhai College, China University of Mining & Technology, Jiangsu, 221008, China (lifeng_198208@163.com)

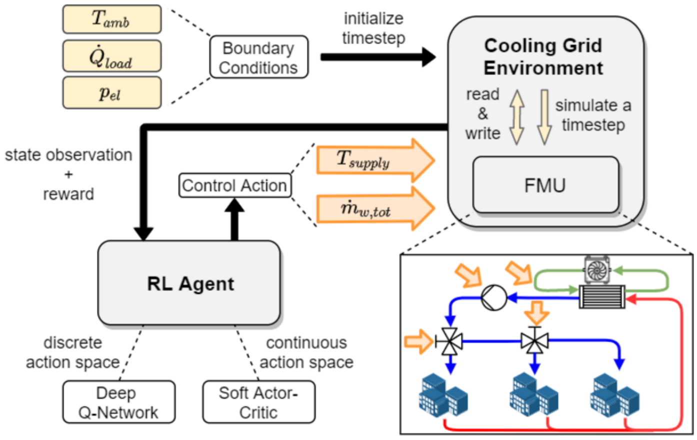

Fig. 2.1: *Schematic diagram of action representation method.*

movement data of human body, combined with the characteristics of human bone structure. The algorithm of human motion trajectory detection is studied [5]. The human body positioning correction based on background deduction is realized by particle filtering on the moving human body. Then, based on the improved location information, Fourier description is carried out to realize the extraction and matching of pedestrian features in the moving state.

2. Design of motion characteristic identification system.

2.1. Feasibility and performance of actions. A three-dimensional image is insensitive to the change in the surrounding environment, which is suitable for aerobics movement recognition [6]. This paper selects the movement recognition model of calisthenics in three-dimensional space. The specific action representation is shown in Figure 2.1.

Three-dimensional behavior data includes two aspects: time-space characteristics and motion trajectory tracking. The time-space representation of behavior is a local representation of behavior. It can map complex behavior [7]. The motion representation of three-dimensional data can recognize moving objects better, but the recognition of complex aerobic movement features needs further research. The behavior representation method based on a three-dimensional skeleton structure has the advantages of solid robustness, fast processing speed and small data scale and has been widely used in behavior recognition research. The geometric relation between the connecting points is obtained by the geometric description method to realize the identification of the object. Through the feature extraction of critical gestures, the gestures are matched to complete the behavior recognition [8].

2.2. Acquisition of motion parameters. According to the demand of human behavior, there have been many research results. It mainly includes tag type, laser ranging type, structured light sensor, Microsoft Kinect sensor, multi-camera, etc. The acquisition process is shown in Figure 2.2. The system is divided into depth mapping, human body parts, and 3D joint modeling.

Since the release of Kinect, in-depth imaging has come a long way. The Kinect camera in the article obtained 640x480 pixel images at 30 frames per second and a depth resolution accuracy of only a few centimeters. Unlike the conventional brightness sensor, it has the advantages of high calibration and positioning accuracy under low light conditions [9]. It can directly synthesize deep images of real people to build a large training dataset cheaply.

3. Particle filter tracking. This algorithm uses the median method to extract static background for multiple video sequences. Select and trace the first frame with a particle filter to obtain relatively accurate positioning information. Then, a particle filter tracks it to get more accurate positioning. Then, the Fourier characteristics of the object are analyzed. Finally, the features obtained directly by background subtraction

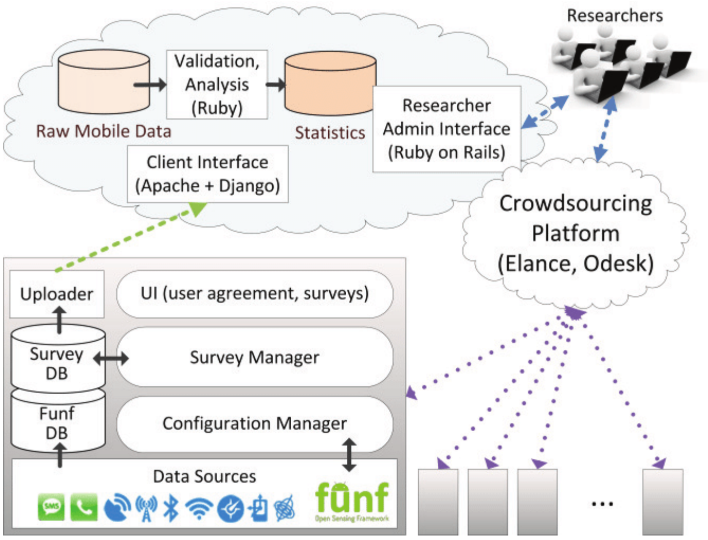

Fig. 2.2: *Action data acquisition architecture diagram.*

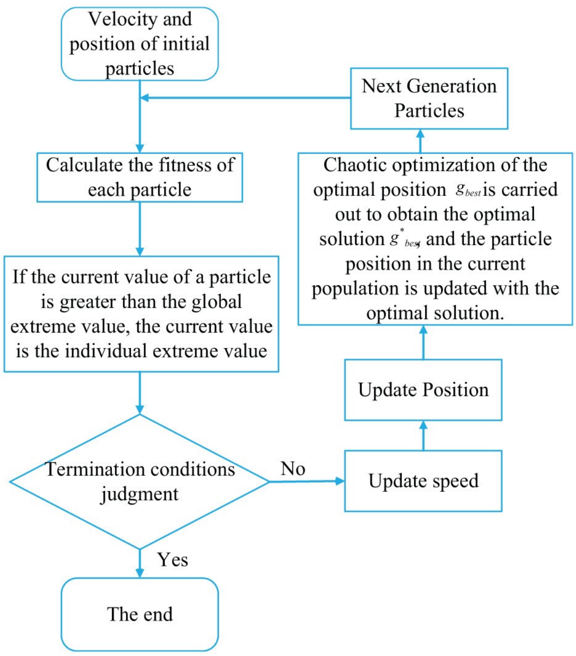

Fig. 3.1: *Flow chart of particle swarm optimization algorithm.*

before tracking and those obtained after tracking are compared with the fundamental features [10]. After tracking this frame and feature extraction according to the above steps, the best state in this frame is used as the tracking start screen for the next frame. The following image is repeatedly predicted until it is tracked to the final image. This allows for complete tracking and feature extraction. The algorithm flow chart is shown in Fig. 3.1.

Particle filter is a new method to process Bayesian filters based on Monte Carlo sampling. The main idea is to use a series of weighted correlated random particles to obtain the posterior probability of each state. The actual states of each state are obtained according to the sample values and weights [11]. The most crucial step in the particle filtering algorithm is to predict and update it. Due to the degradation of particles in the model iteration, many methods have adopted the re-sampling method to improve the tracking results. Behavior identification starts with building a model of an object. Let $\{u_t, t \in N\}$ be the state sequence of the target. A Markov process handles it

$$u_t = g_t\left(u_{t-1}, \delta_{t-1}\right)$$

g_t is the system status function. δ_t is for processing noise. Suppose it's a Gaussian white noise with an average of o. The tracking target predicts the estimated state u_t according to the observation formula. In the process of moving, the arms and legs will shift and rotate, so it is necessary to reconstruct the characteristics of the Fourier descriptors between each frame [12]. The least matching error is selected as the Fourier feature of the current frame. The metric formula is as follows

$$c_t = l_t\left(u_t, n_t\right)$$

c_t is the measurement function of the system. n_t is the measured noise. Suppose it's a Gaussian white noise with an average of o. Baves's theory holds that under observation data $c_{1:t} = \{c_i, i = 1, \cdots, t\}$ from the first time to time point t, the state u_t is estimated by observation data at time point t. The function $f\left(u_t \mid c_{1:t}\right)$ of probability density is calculated. Assuming that the initial probability density $f\left(u_0 \mid c_0\right) = f\left(u_0\right)$ is given, the solution $f\left(u_t \mid c_{1:t}\right)$ can be obtained by the two steps of prediction and updating.

Given that the probability distribution for time $t-1$ is $f\left(u_{t-1} \mid c_{1:t-1}\right)$, the prediction procedure is as follows according to the Chapman-Kolmogorov equation

$$f\left(u_t \mid c_{1:t-1}\right) = \int f\left(u_t \mid c_{t-1}\right) f\left(u_{t-1} \mid c_{1:t-1}\right) du_{t-1}$$

In the above form $f\left(u_t \mid u_{t-1}, c_{1:t-1}\right) = f\left(u_t \mid u_{t-1}\right)$ can be obtained by a Markov method. $f\left(u_t \mid u_{t-1}\right)$ can be determined from formula (3.2) and the common knowledge of process noise δ_{t-1}. According to the measurement result c_t of time t, the prediction formula modified by Bayes' rule can be obtained:

$$f\left(u_t \mid c_{1:t}\right) = \frac{f\left(c_t \mid u_t\right) f\left(u_t \mid c_{1:t-1}\right)}{f\left(c_t \mid c_{1:t-1}\right)}$$

The standardized constant $f\left(c_t \mid c_{1:t-1}\right) = \int f\left(c_t \mid u_t\right) f\left(u_t \mid c_{1:t-1}\right) du_{t-1}$ is determined by $f\left(c_t \mid u_t\right)$ derived from formula (3.2). During the correction period, measure c_t is used to correct the prior density, and then the posterior density of the current state is obtained. Since the particle's weight is constantly changing, the deterioration of the particle occurs when some properties are lost. For this reason, the sample must be re-sampled [13]. The lighter particles are removed and focused on the heavier ones.

4. Fourier descriptors. After segmenting the human form, it is used as a closed curve. Describe it as A complex number $c_i = u_i + jv_i$, where $i = 1, 2, \cdots, N$ is the number of circumference points. By using this method, 2D contour lines can be converted into one-dimensional high-precision vectors to reduce the computation [14]. The center coordinates of the gait path are shown as follows

$$u_c = \frac{1}{N}\sum_{i=1}^{N} u_i, v_c = \frac{1}{N}\sum_{i=1}^{N} v_i$$

Select the starting point (u_0, v_0), and select the starting point A on the contour line S according to the counter-clock, and record the points of (u_i, v_i) on the contour line S according to the counter-clock mode. Then, the distance vector is recorded as follows

$$R_i = \sqrt{\left(u_i - u_c\right)^2 + \left(v_i - v_c\right)^2}$$

Fig. 5.1: *Original grayscale image.*

Perform A discrete Fourier transformation on $R_i(i = 1, 2, \cdots, N)$ to obtain the Fourier descriptor for this distance vector:

$$\varphi(n) = \frac{1}{N} \sum_{i=1}^{N} R_i \exp\left(\frac{-j2\pi in}{N}\right), n = 1, 2, \cdots, N$$

Refer to $\varphi(1)$ and normalize the Fourier descriptor to get $\varphi^*(n)$. The object must be detected first To realize the effective recognition of the object. This algorithm is designed to separate different regions from a series of images accurately. The usual algorithms can be divided into three categories: interframe difference methods, algorithms based on sports field estimation, and algorithms based on background elimination [15]. The experiment used background subtraction. This method can usually obtain the most reliable image information, but it shows strong robustness to complex and changeable environmental factors in the external environment, such as light and external factors. This paper proposes a video-based target detection method to reduce moving targets in video. After the image background is modeled, the background is subtracted from the current image by image background subtraction. Then, a threshold is set, and the difference above this threshold is that the object is moving so that the object containing noise can be extracted. Then, the morphologic transformation is carried out to remove the noise and obtain a smooth image. This study selects the first 30 spectral components of Fourier descriptors. Its minor frequency component mainly characterizes it, and its most significant characteristic is its largest spectral component of order 30. So far, this paper has 30 data properties of Fourier descriptors.

5. System test and performance analysis. Kinect was used to perform experiments and performance evaluations on complex assisted gymnastics movements (group A, group B, group C, and group D) based on the MSRAction 3D database. Two kinds of images are denoised by two-layer, three-layer and four-layer Fourier space-time spatial pyramid, and then multi-dimensional information fusion and particle swarm pattern recognition are used [16]. The test results are listed in Table 5.1. A-, B -, C - and D- are the particular reducing actions of each group of difficulty movements, and A, B, C and D are the minimum standard actions of each group of difficulty movements. The test results and the background image obtained are shown in Figure 5.1.

Experiments show that this method can effectively improve the recognition rate of behavior. The feature information processing algorithm based on skeleton and depth information can significantly improve the recognition rate of human behavior and further verify the superiority of this algorithm in behavior classification. By studying the characteristics of the extracted Fourier descriptors, it is found that after the improved image processing, the obtained image is closer to the actual human body characteristics, which can effectively improve the image classification effect. Eighty screens were used in the experiment, each with 294x465 colors. In extracting Fourier descriptors, the human body is first tracked, and the experimental results of tracking are shown in Figure 5.2 and Figure 5.3.

Table 5.1: *System test result table.*

Difficulty group	Two-layer recognition rate /%	Three-layer recognition rate /%	Four-layer recognition rate /%
A-	81.03	87.84	98.97
A	81.55	91.44	101.13
B-	81.34	88.97	99.79
B	82.47	93.20	102.06
C-	81.24	88.66	98.87
C	81.34	91.75	101.44
D-	81.55	87.22	98.04
D	82.27	92.68	102.16

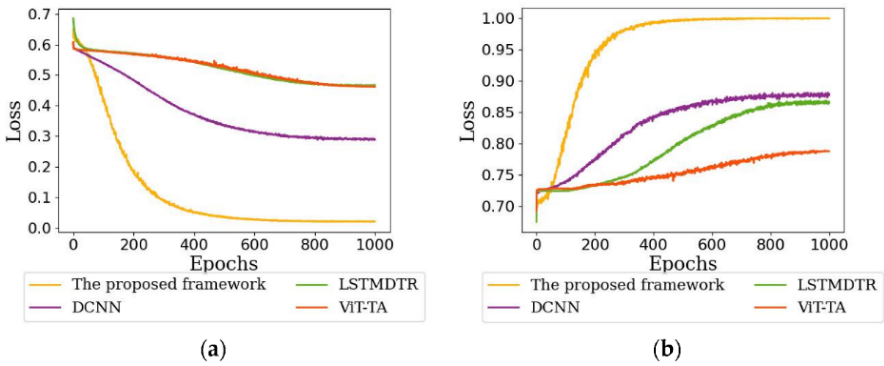

Fig. 5.2: *Background subtraction results.*

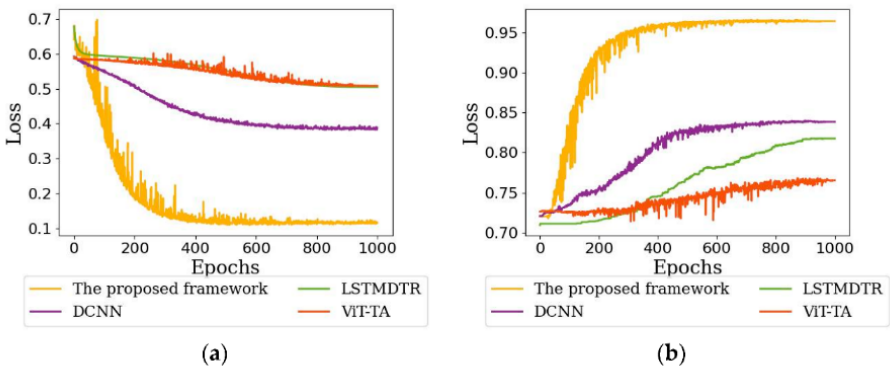

Fig. 5.3: *Particle filter tracking results.*

6. Conclusion. The words are too much because the current evaluation standard of aerobics is too general. A gymnastics performance evaluation system based on big data is developed. This project plans to use multiple gymnastics performance events to expand the human body comparison database and adopt the Fourier Pyramid method to filter and fuse the human skeleton and deep region features to realize behavior classification and recognition based on particle swarm. The experiment shows that this method can achieve a good recognition effect.

REFERENCES

[1] Li, M., Chen, S., Chen, X., Zhang, Y., Wang, Y., & Tian, Q. (2021). Symbiotic graph neural networks for 3d skeleton-based human action recognition and motion prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6), 3316-3333.

[2] Song, Y. F., Zhang, Z., Shan, C., & Wang, L. (2022). Constructing stronger and faster baselines for skeleton-based action recognition. IEEE transactions on pattern analysis and machine intelligence, 45(2), 1474-1488.

[3] Shi, H., Peng, W., Chen, H., Liu, X., & Zhao, G. (2022). Multiscale 3D-shift graph convolution network for emotion recognition from human actions. IEEE Intelligent Systems, 37(4), 103-110.

[4] Shu, X., Yang, J., Yan, R., & Song, Y. (2022). Expansion-squeeze-excitation fusion network for elderly activity recognition. IEEE Transactions on Circuits and Systems for Video Technology, 32(8), 5281-5292.

[5] Kong, Y., & Fu, Y. (2022). Human action recognition and prediction: A survey. International Journal of Computer Vision, 130(5), 1366-1401.

[6] Wu, H., Ma, X., & Li, Y. (2021). Spatiotemporal multimodal learning with 3D CNNs for video action recognition. IEEE Transactions on Circuits and Systems for Video Technology, 32(3), 1250-1261.

[7] Lu, J., Pan, L., Deng, J., Chai, H., Ren, Z., & Shi, Y. (2023). Deep learning for flight maneuver recognition: A survey. Electronic Research Archive, 31(1), 75-102.

[8] Islam, M. M., & Iqbal, T. (2021). Multi-gat: A graphical attention-based hierarchical multimodal representation learning approach for human activity recognition. IEEE Robotics and Automation Letters, 6(2), 1729-1736.

[9] Gupta, N., Gupta, S. K., Pathak, R. K., Jain, V., Rashidi, P., & Suri, J. S. (2022). Human activity recognition in artificial intelligence framework: A narrative review. Artificial intelligence review, 55(6), 4755-4808.

[10] Challa, S. K., Kumar, A., & Semwal, V. B. (2022). A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. The Visual Computer, 38(12), 4095-4109.

[11] Singh, R., Khurana, R., Kushwaha, A. K. S., & Srivastava, R. (2021). A dual stream model for activity recognition: exploiting residual-cnn with transfer learning. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 9(1), 28-38.

[12] Maqsood, R., Bajwa, U. I., Saleem, G., Raza, R. H., & Anwar, M. W. (2021). Anomaly recognition from surveillance videos using 3D convolution neural network. Multimedia Tools and Applications, 80(12), 18693-18716.

[13] Zhou, T., Qi, S., Wang, W., Shen, J., & Zhu, S. C. (2021). Cascaded parsing of human-object interaction recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6), 2827-2840.

[14] Savchenko, A. V., Savchenko, L. V., & Makarov, I. (2022). Classifying emotions and engagement in online learning based on a single facial expression recognition neural network. IEEE Transactions on Affective Computing, 13(4), 2132-2143.

[15] Guo, L., Lu, Z., & Yao, L. (2021). Human-machine interaction sensing technology based on hand gesture recognition: A review. IEEE Transactions on Human-Machine Systems, 51(4), 300-309.

[16] Xu, Y., & Qiu, T. T. (2021). Human activity recognition and embedded application based on convolutional neural network. Journal of Artificial Intelligence and Technology, 1(1), 51-60.

Edited by: Hailong Li
Special issue on: Deep Learning in Healthcare
Received: Feb 3, 2024
Accepted: Apr 1, 2024