Moving Athlete Detection for Multimedia Physical Education

Meiqin Chen[1,*], Jiaping Xu[2], Ru Jia[1] and Hengyong Xu[1]

[1]*Physical Eduation College, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, P.R. China*

[2]*Sport Department, Jiangsu University, Zhenjiang, Jiangsu, P.R. China*

Abstract: In the process of the multimedia physical education, the moving athlete detection is the key to tracking and recognizing the athlete. This task contributes to the gesture and action modeling and simulation. However, since the scene of sports events is complicated, the background is commonly affected by the noise and illumination changes, making the background difficult to remove. In order to solve this problem, a novel motion saliency calculation based method is proposed for detecting the moving athlete in sport videos. In our method, four motion features, as the optic flows, direction vectors and frame difference are fused by the Quaternion Fast Fourier Transform. The motion saliency map is generated to identify the area of the moving athlete. The experimental result demonstrated that our method has better ability to detect the moving athletes, contributing to the multimedia physical education..

Keywords: Athlete detection, motion saliency, multimedia technology, physical education.

1. INTRODUCTION

Multimedia education properly choice and apply the modern media technology, such as the CAD and computer vision, into the teaching process. By the multimedia technology, the traditional education process is changed into more vivid and understandable, and the teacher can make a proper structure for education process. Physical education is actually a bilateral behavior between the teacher and student. However, in the traditional physical education process, the teaching task is merely taken by using the boring speech or the simulation of the teachers. This traditional model would make the students feel tired, gradually losing the interests on the physical education. Moreover the traditional physical education which is a coercive process would promote the sluggishness of the students who even skip classes. As the result, the physical educations can neither enhancing the physique nor relaxing the mind.

In contrast to the traditional physical education methods, the multimedia education has the ability to inspire the interests of the students on the physical learning. For physical education, the samples which are visualized and infective have more opportunity to absorb more interests and be feasibly accepted. This requirement can be cater by the multimedia education technology. Vivid and animate video and sound soft wares change the traditional education mode which is based on the simulation and the explanation made by the teachers. The attention and the interest is raised and kept by the multimedia samples. By using the multimedia courseware, animation or video processing technology, the slow motion, the single frame and the reset frames are given to represent the educational content. This process makes the instant behavior and gesture clear to the students. As the result, the technological detail of the education samples can be seen and understood clearly, accelerating the learning process on the difficulties and the key points. At last, by combining the practical application, the understanding and experience are both enhanced.

For most multimedia education, the athletes in the video are the interest objects. The teacher should extract the area of the athletes to show the detail of the behavior and the technological key. In this case the computer vision method, especially the object detection and extraction method, is required. Many moving object detection method have been introduced into the multimedia education tasks, while specialized to the physical education, moving object detection methods have not been widely introduced.

In the field of the computer vision, classical moving object detection methods mainly include frame difference method, background subtraction and optical flow [1-6]. However, as aforementioned for physical education the context and content of the scene, such as the football game, is complicated. Moreover, since the powerful artificial lumination is widely used in the playground, the lamination change would make the intensity of the collected image changed in large extent. These two cases would significantly decrease to the accuracy of the object detection method. Due to the high computational efficiency and accuracy, motion saliency based moving object detection methods have been widely used in the computer vision task as the preprocessor for object detection, recognition and object tracking. Examples include the popular Itti model [7] and its updated editions [8]. However these excellent motion saliency extraction methods have not been introduced into the physical multimedia education process. This paper presents a novel method for athlete detection from the sport video. Moreover, in order to solve the problems in the moving object detection, the multiple feature fusion strategy is applied to produce com-

prehensive motion saliency maps. The athlete detection results with more accuracy are given by our method.

2. RELATED WORK

Many saliency detection methods have been proposed for object detection from the image or videos. Generally, these approaches can be divided into two subcategories. One employs low-level image features. Examples include the popular Itti model [7] and its updated editions [8]. As these methods are based on local features, local saliency is excessively emphasized, while the global saliency is lost. In order to solve this problem, other methods which can prevent too much local saliency operate in the frequency domain. They are commonly called global methods, such as the Spectral Residual (SR) method [9]. The theoretical basis for this category is that variations in the frequency spectrum reflect the saliency in the image. Due to the emphasis on global information, these methods have more opportunities to find the global saliency which is dominant across the whole scene. A typical example include the combination of the center-surround model and the dynamic features [10, 11].

By using the aforementioned saliency detection methods, many moving object detection methods have been proposed. These methods generally introduce and fuse the motion features into the saliency extraction models to transfer the static saliency model to the motion saliency edition. Mahadevan[9] introduced the spatial information and the moving information into the Itti model. Although this method caters for the motion saliency detection, the large computational complexity generally blocks the application. To solve this problem, Cui [12] proposed a time slice spectral analysis based saliency extraction. This method is computationally economic and can well meet the real-time requirements. However, this method is specialized to the video with long interval. For the short-term frame sequences, the motion saliency is weak. In order to enhance the motion features, Guo Chenlei [13] introduced the inter-frame difference feature into the PQFT model to detect moving targets. However, the inter-frame difference is only as the feature to compensate static characteristics such as color and gray-scale. Due to less moving information, test results include a large number of static salient features, such as texture background interferences.

Although these above methods obtained many good results in many tasks, none of them has the ability to process the video for multimedia physical education. To correctly detect the moving athletes in the sport video, we should increasingly enhance the motion feature extracted from the video. To enhance the motion feature, the multiple feature fusion strategy is applied in this paper.

3. MOTION SALIENCY

Traditional static saliency map is mainly obtained by calculating the static features, such as the intensity, color, orien-

tation, etc. However to extract the motion saliency, the motion feature is the key. In the computer vision researches, many motion features have been proposed. By each feature, the motion characteristic is depicted. Hence, in order to comprehensively model the motion state of the interest objects, the feature fusion strategy should be applied. For fusing multiple motion features, this paper establishes a novel motion saliency extraction model. This model contains 3 modules, as feature fusion calculation, phase spectrum calculation and saliency calculation. We select four motion features as the input of the model. The phase spectrum calculation is used to analyze the statistical regularities of these features in the frequency domain. The statistical regularities in the frequency domain are kept in line with the varies in the temporal domain. By analyzing the frequency spectrum, the temporal motion information is given. The motion saliency extraction process is shown in Fig. (**1**).

3.1. Multi-features Fusion

Since single motion feature can hardly describe the comprehensive motion information of the interest objects. We take the multi-features fusion strategy into account. For feature fusion, three motion features are selected, as the frame-difference, optical flow and direction vector. The frame difference is a common motion feature extraction method, its advantage lies in the high computational efficiency and the results can reflect the differences between frames. However, due to the corresponding pixel subtraction strategy, the background noise is left in the result if the background changes or shakes. To solve this problem, we take the optical flow as the second feature. The optical flow estimates the motion vectors of each pixel. By filtering the estimated motion vectors, it can suppress the background noise with small motion vectors in some extent. Finally, the direction vector is extracted to suppress the state variable noise in the background. This paper fuses above three motion features into a quaternion feature.

Assuming that the extracted optical flow, optical flow, direction vector and frame difference feature are $U(t), V(t), \theta(t), M(t)$, the quaternion feature set can be shown as:

$$q(t) = M(t) + U(t)\mu_1 + V(t)\mu_2 + \theta(t)\mu_3 \tag{1}$$

where $\mu_1^2 = -1$, $\mu_3 = \mu_1\mu_2$ and mutually orthogonal between μ_i。

3.2. Phase Spectrum Calculation

For a static image, the phase spectrum curve can reflect the changes which called salient areas in original input signal.

Assuming that the input signal is $f(x)$, the amplitude and phase information given by the Fourier transform is
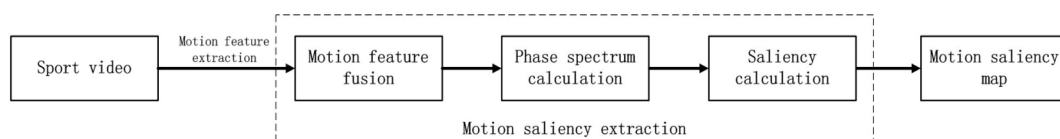

Fig. (1). Moving saliency extraction.

$|F(\omega)|$ and $\theta(\omega)$ respectively, then, phase spectrum can be represented as:

$$f_p(x) = IDFT(|F(\omega)|e^{j\theta(\omega)}) \tag{2}$$

If the input quaternion feature set is $q(t)$, then, formula (1) can be represented as:

$$q(n,m) = f_1(n,m) + f_2(n,m)\mu_2 \tag{3}$$

Quaternion Fast Fourier Transform can be calculated as follows：

$$F_i[u,v] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{1} \sum_{n=0}^{1} e^{-\mu_1 2\pi((mv/M)+(nu/N))} f_i(n,m) \tag{4}$$

Where n,m represent the time domain, u,v represent the frequency domain. Assuming that the input image after the QFFT can be expressed as $Q(t)$, the phase spectrum can be calculated in polar coordinates by the following formula:

$$Q(t) = \|Q(t)\|e^{\mu\Phi(t)} \tag{5}$$

When $\|Q(t)\| = 1$, $Q(t)$ only contains the phase information in the frequency domain.

3.3. Motion Saliency Extraction

By analyzing and calculating the spectrum information, we can extract the moving information in frequency domain. This motion information is in line with the motion state in the temporal domain. Then, this moving information in the frequency domain is required to be computationally projected into the time domain, generating the moving saliency map. This paper transforms the motion information from the frequency domain into temporal domain by Quaternion Fast Fourier Inverse Transform. Assuming the image given by the phase spectrum calculation is as $F_i[u,v]$, then, QFFIT can be represented as:

$$f_i(n,m) = \frac{1}{\sqrt{MN}} \sum_{v=0}^{M-1} \sum_{u=0}^{N-1} e^{\mu_1 2\pi((mv/M)+(nu/N))} F_i[u,v] \tag{6}$$

Assuming the phase spectrum after the QFFIT is $q'(t)$, the final fusion moving saliency map can be expressed as:

$$MSM(t) = g(t) * \|q'(t)\|^2 \tag{7}$$

Where $g(t)$ is the 2D Gaussian filter. $MSM(t)$ is the moving saliency map obtained by the phase spectrum method.

4. MOVING ATHLETE DETECTION PROCESS

Moving saliency calculation based moving athlete detection process is shown in Fig. (**2**). This process includes mo-

tion feature extraction, moving saliency calculation, additive correcting fusion.

4.1. MOTION FEATURE EXTRACTION

This paper selects the optical flow, direction vector and frame difference as the input of model. These three motion features are extracted as below:

(1). Optical Flow

The optical flow feature represents the instantaneous change in the intensity. Gradient-based optical flow algorithm is used to get optical flows in x and y directions.

Assuming the gray value is $I(x,y,t)$ at time t. At time $(t+\Delta t)$, the gray value is $I(x+\Delta x, y+\Delta y, t+\Delta t)$. Since the gray value remained unchanged at time Δt, the optical flow constraint equation is described as follows:

$$I(x,y,t) = I(x+\Delta x, y+\Delta y, t+\Delta t) \tag{8}$$

$$u^{(k+1)} = \overline{u}^{(k)} - I_x \frac{I_x\overline{u}^{(k)} + I_y\overline{v}^{(k)} + I_t}{\lambda^2 + I_x^2 + I_y^2} \tag{9}$$

$$v^{(k+1)} = \overline{v}^{(k)} - I_y \frac{I_x\overline{u}^{(k)} + I_y\overline{v}^{(k)} + I_t}{\lambda^2 + I_x^2 + I_y^2} \tag{10}$$

Where u,v represent optical flows in x and y directions.

(2). Direction Vector

Direction vector reflects the direction of the motion between consecutive frames, it is capable of describing the corresponding relationship between pixels of adjacent frames.

Assuming the motion vector component in the x axial is V_x^{motion}, y axis direction is V_y^{motion}, then the direction vectors can be obtained as follows:

$$\theta = arctg\left(\frac{V_y^{motion}}{V_x^{motion}}\right) \tag{11}$$

(3). Frame Difference

Difference algorithm is applied to two or more consecutive frames to obtain the frame diffidence vector as one of the input features.

$$M = |I(t) - I(t-\tau)| \tag{12}$$

Where $I(t)$ represents the gray value at time t, $I(t\text{-}\tau)$ represents the gray value after time τ. We can set $\tau = 1$。
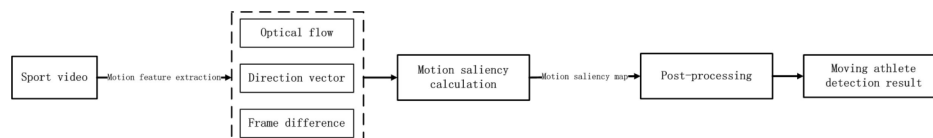

Fig. (2). Moving athlete detection process.

4.2. Post-processing

Based on moving saliency calculation model and multi-features fusion method, the moving saliency map is given. However, when moving targets maintain static at a certain frame or several consecutive frames, the target will be regarded as background noise to be suppressed. Therefore, the post processing is applied.

(1). False Object Removal

Assuming the moving saliency map at time $k-1$ is $MSM(k\text{-}1)$, moving saliency map at time k is $MSM(k)$, the false object can be recognized as follows:

$$M(k) = MSM(k) - MSM(k-1) \tag{13}$$

Where $M(k)$ is the subtractive result of moving saliency map. Assuming the threshold value is T, target can be recognized according to the following criteria:

Criterion 1 if $M(k)>T$ over 3 consecutive frames, then the target is disappeared, stopping testing.

Criterion 2 if $M(k)>T$ within 3 consecutive frames, then the target is ephemeral static target, correcting should be done to the result.

Criterion 3 if $M(k)\leq T$, then the ephemeral static target is not exist.

(2) Additive Correcting Fusion Method

Assuming the moving saliency map at time $k-1$ is $MSM(k-1)$, at time k, the moving saliency map is $MSM(k)$. Additive correcting fusion method is as follows:

$$MSM(k) = MSM(k) \oplus MSM(k-1) \tag{14}$$

Ephemeral static target can be detected by the fusion of moving saliency map in the time domain. It achieves the integration of static target detection and moving target detection and improves the accuracy of test result.

5. MOTION FEATURE EXTRACTION

In order to demonstrate the contribution of our method for the multimedia physical education, the correctness and the feasibility of the presented method is shown. For the physical education, in order clearly indentify the behavior of the athletes we should correctly identify the area of them. Hence, this paper takes the object detection accuracy to evaluate the performance of our method. This paper selected four typical scenarios to compare the performance of our method with traditional detection methods, as the spectral residual (SR), the Phase Quaternion Fourier Transform (PQFT). Experimental platform is PC with 2.5GHz of frequency, 4G of memory, the interval between frames is 0.3s.

5.1. Motion Saliency Extraction Results in Different Video

Fig. (**3**) shows the results obtained in a teaching video for the tennis sport. These results demonstrate that our method has better performance on motion saliency detection and the salient points focus on the moving objects which are the teacher and the tennis racket. However, other methods both mistake the static texture in the background as the saliency area and we can hardly identify the moving teacher.

Fig. (**4**) are test results obtained in a teaching video for the table tennis. In contrast to the static saliency given by the SR and the motion saliency given by the PQFT methods, the motion saliency extracted by our method completely focuses on the teacher, while the background is highly removed. With this saliency map, we can extract the teacher and his gesture or behavior can be clearly identified.

As we can see from the Fig. (**5**), only our method has the ability to accurately detect the moving ball in the snooker video, while the table in the background is removed. However, for the results given by other method the snooker table generates static saliency.

Fig. (**6**) shows the results obtained in a video collected in a real badminton game. In this scene, the background has abundant texture feature. In this case, there are many background noises in saliency maps given by the SR and PQFT. By these saliency maps, we can hardly indentify the moving athletes. By contrast, in the results given by our method, only the moving athletes are left, while the background and the static objects are removed largely.
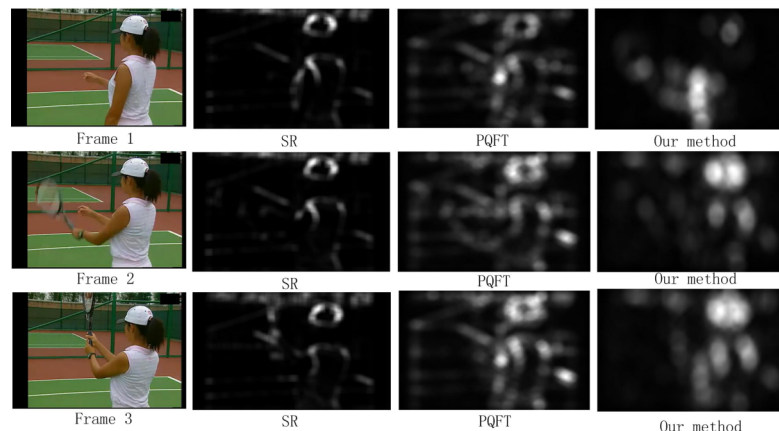

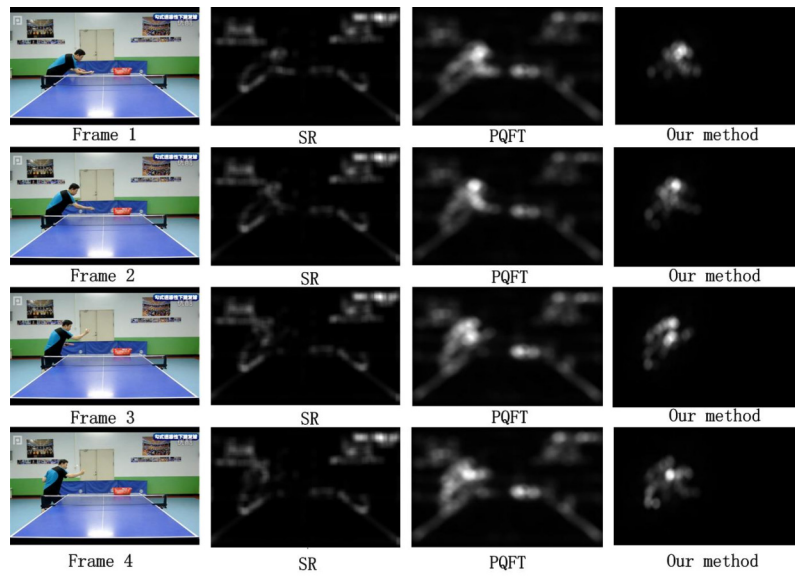**Fig. (3).** Motion saliency extraction in the tennis video.

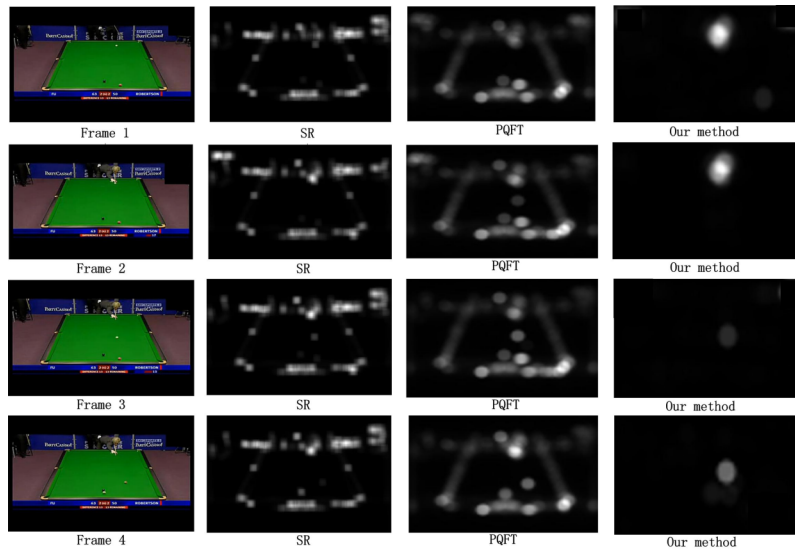**Fig. (4).** Motion saliency extraction in the table tennis video.

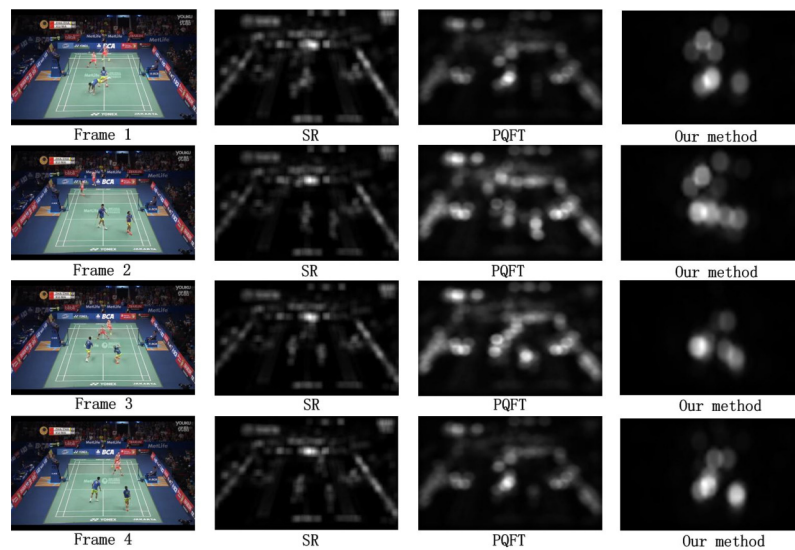**Fig. (5).** Motion saliency extraction in the snooker video.

Fig. (6). Motion saliency extraction in the badminton game.

5.2. Moving Athlete Detection Results and Evaluation

Fig. (**7**) shows the results for moving athlete detection based on our motion saliency extraction method. From the results shown in Fig. (**7**), the moving athletes or objects are correctly detected. Moreover, in order to evaluate the pros and cons of the proposed method, four factors are used for performance comparision, such as C_{good}, C_{false} and computational time t.

$$C_{good} = \frac{card\left\{\Omega_{in} \cap \Omega_o\right\}}{card\left\{\Omega_o\right\}} \tag{15}$$

$$C_{false} = \frac{card\left\{\Omega_{in} \cap \Omega_b\right\}}{card\left\{\Omega_b\right\}} \tag{16}$$

Where, Ω_{in} is the internal domain of extracted target, Ω_o is the real domain of the target and Ω_b is the background.

C_{good} is the ratio that correctly extracted target region accounted for the real target area, C_{false} is the ratio that incorrectly extracted target region accounted for the background area. Five methods are referred for performance comparison, as the frame difference, optical flow, mixture Gaussian, SR and PQFT methods. The quantitative evaluation is shown in Table **1**.

It can be seen from Table **1** that the correct ratio of our algorithm is the largest. Since the algorithm can suppress background noise and non-interested saliency static targets. Hence by our method, the athlete area can be correctly detected. From the results given by our method, teachers and students can clearly identify the gesture and the behavior of the teacher and the athletes. In addition, the algorithm has high computational efficiency. These benefits of our method contribute to the multimedia physical education.
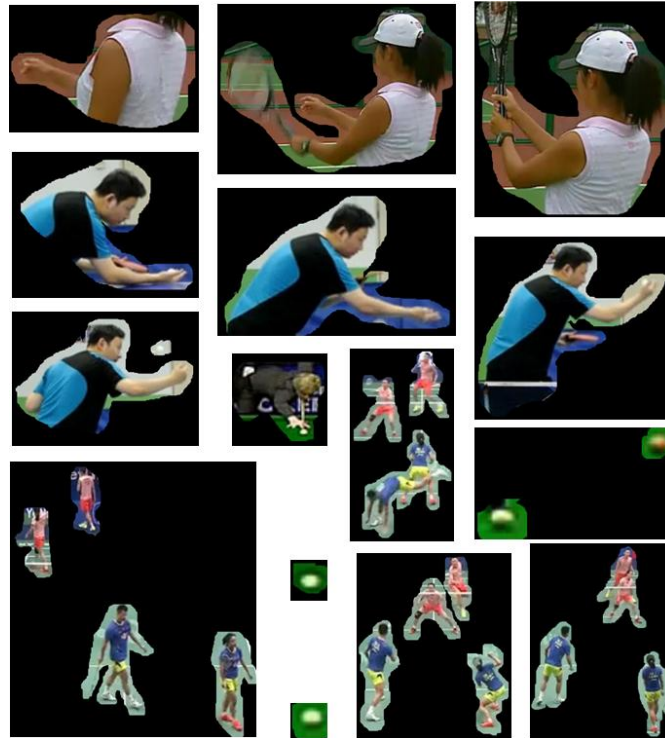
CONCLUSION

Fig. (7). Moving athlete detection.

Table 1.　　Performance evaluation.

Method	C_{good}	C_{false}	Time (s)
Frame difference	0.772	0.084	0.24
Optical flow	0.703	0.102	2.63
Mixture Gaussian	0.765	0.072	2.78
SR	0.261	0.626	3.04
PQFT	0.554	0.507	0.45
Our method	0.875	0.028	0.62

Multimedia teaching contributes to the physical education and training. In order to present the gesture and the behavior of the athlete in the teaching process, This paper proposes a novel moving athlete detection method based on motion saliency extraction. This method extracts three motion features between adjacent frames of video sequence as the quaternion input. The experiment results proved that our method can remove the background texture and static noises, accurately detecting the multiple moving athletes with high accuracy. The results given by our methods have the ability to clearly represent the behavior and the gesture of the moving athletes in the teaching video.

CONFLICT OF INTEREST

The authorsconfirm that this article content has no conflict of interest.

REFERENCES

[1]   S. T. Su, and Y. Y. Chen, "Moving object segmentation using improved running gaussian average background model, in Digital Image Computing," *Techniques and Applications (DICTA)*, pp.24-31, 2008.

[2]   O. Barnich, and M. Van Droogenbroeck, "ViBE: A powerful random technique to estimate the background in video sequences, in Acoustics," *Speech and Signal Processing*, ICASSP IEEE International Conference on,pp. 945-948, 2009.

[3]   W. Fengchao, L. Xingtang, and H. Shucai, "Target fusion detection with multi-feature based on fuzzy evidence theory," *Acta Optica Sinica*, vol. 30, no.3, pp. 713-719, 2010.

[4]   J. Yang, Y. D. Sun, M. J. Wu, *et al*. "Multi-class moving target detection with Gaussian mixture part based model//Consumer Electronics (ICCE)," *IEEE International Conference on*, pp. 386-387, 2014.

[5]   C. Yin, R. Kan, G. Guohua, *et al*. "Moving Object Detection Based on Improved Single Gaussian Background Model," *Chinese Journal of Lasers*, vol. 11, pp. 040-047, 2014.

[6]   Z. H. Q. H. L. Rui, and L. Shangqian, "Multiscale Truncation for Dim and Small Target Background Suppression," *Acta Optica Sinica*, vol. 10, pp. 011-015, 2010.

[7]   L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no.11, pp.1254-1259, 1998.

[8]   J. K. Tsotsos, and A. Rothenstein, "Computational models of visual attention Scholarpedia," vol. 6, no.1, pp. 6201-6205, 2011.

[9]   X. Hou, and L. Zhang, "Saliency detection: A spectral residual approach, Computer Vision and Pattern Recognition," *CVPR'07. IEEE Conference on*, pp. 1-8, 2007.

[10]   V. Mahadevan, and N. Vasconcelos, "Biologically inspired object tracking using center-surround saliency mechanisms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no.3, pp. 541-554, 2013.

[11]   V. Mahadevan, and N. Vasconcelos, "Spatiotemporal saliency in dynamic scenes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no.1, pp. 171-177, 2010.

[12]   X. Cui, Q. Liu, and D. Metaxas, "Temporal spectral residual: fast motion saliency detection," In: *Proceedings of the 17th ACM international conference on Multimedia*, ACM, pp. 617-620, 2009.

[13]   C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," *Computer vision and pattern recognition, cvpr conference on*. IEEE, pp. 1-8, 2008.