

Research on Automatic Tracking of Marking Points in Sports Image Sequences Based on Machine Learning

Peng Cheng, Xiaoni Man, Tianbo Yu*

Shenyang Jianzhu University, Shenyang, China

*Correspondence Author

Abstract. The article aims at dynamically fusing sports images to obtain the constraint items of the edge highlight model. At the same time, we combine the RGB feature decomposition technology to extract the feature mark point features. The article combines the steady-state matching technology to complete the automatic tracking of the landmark points of the sports image sequence. Preliminary application results show that this method has low requirements on the experimental environment, strong robustness, and high recognition rate. It can not only be used for human motion analysis, but also has great application potential for automatic tracking of moving targets in other free backgrounds.

Keywords: machine learning; sports images; image sequence; landmark tracking; shadow removal.

1. Introduction

In recent years, domestic researches on human motion image analysis systems based on ordinary cameras have been carried out. Although this system uses low-cost surveillance image systems to replace special equipment, which greatly reduces the cost, it increases the difficulty of automatic identification and tracking. Currently, there are mainly two types of recognition methods: model-based and landmark-based. The former does not need to limit the experimental background, but cannot perform precise joint tracking [1]. The latter's joint point positioning accuracy is high, but in general, automatic tracking can be achieved only when the experimental environment is strictly restricted, and even the subject is required to wear black tights. This not only restricts the application range of this type of system, but also interferes to a certain extent. The subject's exercise affected the test results.

Traditional methods for tracking sports image dynamic feature landmark points mainly include accelerated robust feature algorithm (SURF) and SIFT algorithm, etc. The Cyber City-Modeler system launched by the Swiss Federal University of Technology Zurich is used to track and recognize sports images, using templates. The matching technology performs 3D feature reconstruction of sports images and contour feature detection of sports image sequences. However, the real-time performance of this system in the tracking of characteristic landmarks is not good.

Aiming at the above problems, we propose an automatic tracking algorithm for dynamic feature landmarks of sports images based on block fusion feature matching. The dynamic feature mark points of sports images are extracted, and the block fusion technology is adopted to realize the automatic tracking of dynamic feature mark points.

2. Sports target detection

For a video system captured by a still camera, the background is relatively static as time goes by, and only the moving target has displacement, and the displacement causes a large change in the gray value of the moving area. By detecting this change in the gray value, the target can be detected and the movement of the target can be analyzed [2]. The inter-frame difference method is to detect and extract the target by subtracting two or several consecutive frames in the image sequence. The basic inter-frame difference method is to use pixel-based time difference between two consecutive frames of images and threshold the difference result to extract the moving area in the image. The basic inter-frame difference method is expressed by the formula as follows:

$$D_k(x, y) = |f_k(x, y) - f_{k-1}(x, y)|$$

$$R_k(x, y) = \begin{cases} 0, & D_k(x, y) < T \\ 1, & D_k(x, y) \geq T \end{cases} \quad (1)$$

Among them, D_k is the image after difference, R_k is the image after binarization, f_k and f_{k-1} are two consecutive frames of image, and T is the binarization threshold. For the differential image obtained by the above formula, a series of image post-processing can be added as needed to remove noise interference to obtain better detection results [3]. The principle of the two-frame difference method is shown in Figure 1 (the picture is quoted from Vision-based vehicle detection and counting system using deep learning in highway scenes):

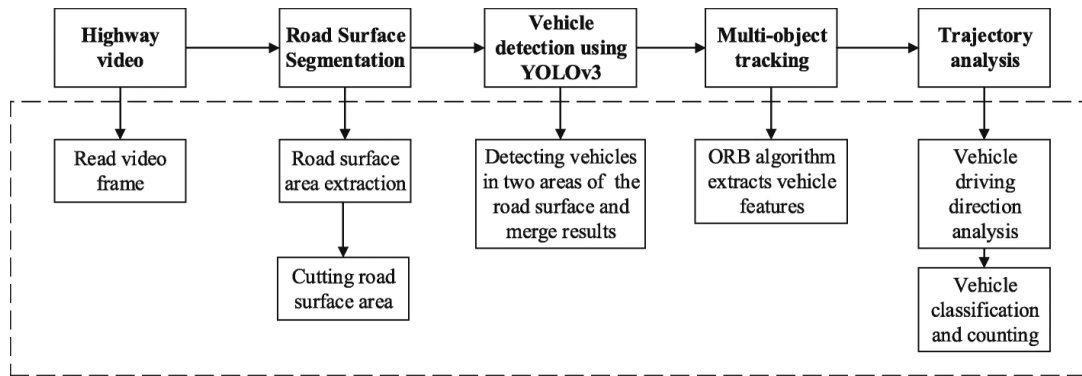


Figure 1. Schematic diagram of the two-frame difference method

Assuming that f_{k-1}, f_k, f_{k+1} is a continuous three-frame image, the pixels in the image satisfying formula (2) are judged as foreground pixels:

$$|f_k(x, y) - f_{k-1}(x, y)| \geq T_1 \quad \text{and} \quad |f_{k+1}(x, y) - f_k(x, y)| \geq T_2 \quad (2)$$

Among them, T_1, T_2 is the preset binarization threshold. First, the background image is acquired, and then the current frame image and the background image are used to make a difference, and then the difference is thresholder to detect the motion area. If there is a moving target in the current frame, the brightness of the moving target area will change [4]. By comparing the current image frame with the background image, it is determined that the area with greater brightness change is the foreground area. Similar to the inter-frame difference method, the background difference method can be simply expressed as the following formula:

$$D_k(x, y) = |f_k(x, y) - B_k(x, y)|$$

$$R_k(x, y) = \begin{cases} 0, & D_k(x, y) < T \\ 1, & D_k(x, y) \geq T \end{cases} \quad (3)$$

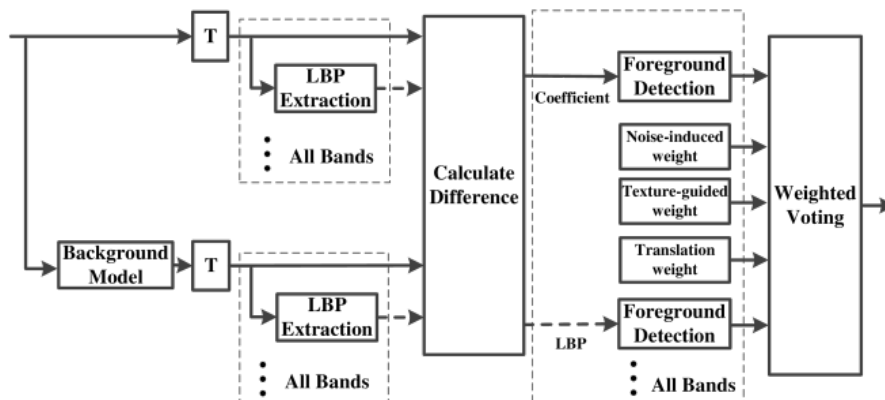


Figure 2. Schematic diagram of the background difference method

Among them, D_k is the differential image, R_k is the binarized image, f_k is the current frame image, B_k is the background image, and T is the binarized threshold. The principle of the background difference method is shown in Figure 2 (picture quoted from Foreground Detection in Camouflaged Scenes):

In order to make the background image adapt to the dynamic changes of the lighting in the scene, the background needs to be updated in real time. The simplest background update method is as follows:

$$B_{k+1} = \alpha f_k + (1 - \alpha)B_k \quad (4)$$

In the formula, f_k is the input image of the current frame, and B_{k+1} is the accumulated image data, which is used as the background. α is a predefined constant whose value is between [0,1] and is used to adjust the background update rate. The larger the α , the faster the update rate of the background image with the current input image, and the stronger the ability to adapt to the environment, but it is susceptible to noise interference [5]. Conversely, it takes longer for the background to adapt to environmental changes, and the background update rate is slower.

3. Sports video shadow removal

3.1 Optical properties of shadows

Under outdoor lighting conditions, the light in the non-shaded area consists of two parts: sunlight and scattered light from the sky. In the umbral area, because objects block the direct sunlight, only the scattered light from the sky. In the penumbra area, it is a combination of scattered light from space and sunlight. According to the chromaticity reflection model, the reflected light $L_r(\vec{q}, \lambda)$ at a given point $\vec{q}$ with respect to wavelength λ can be expressed as:

$$L_r(\vec{q}, \lambda) = L_a(\lambda) + L_b(\vec{q}, \lambda) + L_s(\vec{q}, \lambda) \quad (5)$$

Among them, $L_a(\lambda)$ represents the reflected light around wavelength λ ; $L_b(\vec{q}, \lambda)$ represents the reflected light of the body at wavelength λ ; $L_s(\vec{q}, \lambda)$ represents the reflected light on the surface of wavelength λ . When the changing point is a shadow point, the volume reflected light and the surface reflection light are both 0, and the reflected light of wavelength λ of the shadow point is:

$$L_r(\vec{q}, \lambda) = L_a(\lambda) \quad (6)$$

Therefore, the brightness of the shadow area is much lower than that of the corresponding background area. That is to say, the shadow area can be regarded as a semi-transparent area, and in the shadow area, the optical gain relative to the background image is a fixed value, and less than 1. The chromaticity of the pixels in the shadow area does not change significantly, that is, under certain brightness conditions, the chromaticity of the same object in the shadow area and not in the shadow area is close to the same.

3.2 Extraction of foreground binary image

After the background model is established by adopting the mixed Gaussian background modeling method based on the RGB color space, the original image and the background are compared, and the result of a moving object moving under this background can be reflected.

For each pixel value $X_{ij,t}$ of the current frame, and the corresponding background pixel value $B_{ij,t}$, the foreground binary image FG can be obtained by formula (7):

$$FG_{ij,t} = \begin{cases} 1 & \text{if } |X_{ij,t}^R - B_{ij,t}^R| > 2.5\sigma_{ij,k,t} \text{ or} \\ & |X_{ij,t}^G - B_{ij,t}^G| > 2.5\sigma_{ij,k,t} \\ & \text{or } |X_{ij,t}^B - B_{ij,t}^B| > 2.5\sigma_{ij,k,t} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Among them, 1 represents the foreground area, that is, the moving target; 0 represents the background area.

3.3 Shadow removal algorithm in RGB color space

Usually, the videos used in the experiment are based on the RGB color space, but in the literature, the HSV color space is used to remove the shadows [6]. From the characteristics of the shadows, the HSV model seems to be very suitable for shadow detection, and the color space It can separate the chromaticity and brightness components very well. Therefore, we must convert between the RGB color space and the HSV color space, as in the following formula:

$$\begin{aligned} V &= \text{Max}(R, G, B) \\ S &= \begin{cases} V - \text{Min}(R, G, B) \times 255 / V & \text{if } (V \neq 0) \\ 0 & \text{otherwise} \end{cases} \\ H &= \begin{cases} (G - B) \times 60 / S & \text{if } (V = R) \\ 180 - (G - B) \times 60 / S & \text{if } (V = G) \\ 240 + (G - B) \times 60 / S & \text{if } (V = B) \end{cases} \end{aligned} \quad (8)$$

In the use of color information, the advantage of HSV is that it can separate the brightness (V) and the two parameters that reflect the essential characteristics of the color, hue (H) and saturation (S) one by one, and the three components are independent of each other. Since the three components of HSV are independent, they can be considered separately. The correlation between the three variables of RGB is stronger than that of HSV. When each variable is processed independently in the RGB space, more information may be ignored; and the influence of the change of light intensity on the three components of R, G, B Both are relatively large, so building a model on the RGB space cannot well resist the impact of brightness changes [7]. The covariance operator can be obtained by improving the direct difference operator. We take the average m of the center pixel and the eight pixels in its neighborhood, namely

$$m = \frac{1}{9} \left(\sum_{k=-1}^1 \sum_{p=-1}^1 f(i+p, j+k) \right) \quad (9)$$

Therefore, the covariance about the center pixel is expressed as

$$COV = \sqrt{\sum_{p=-1}^1 \sum_{k=-1}^1 (f(i+p, j+k) - m)^2} \quad (10)$$

However, the detection effect obtained only by the brightness is not good, and the determination of the threshold value is also more difficult. Therefore, we use color invariance to detect and remove shadows. The candidate areas of moving objects and shadows can be obtained by the following formula

$$S_c = \begin{cases} R_r(x, y) - R(x, y) > b_R \\ G_r(x, y) - G(x, y) > b_G \\ B_r(x, y) - B(x, y) > b_B \end{cases} \quad (11)$$

Among them, S_c is a candidate area containing shadows and moving targets; $R_r(x, y)$, $G_r(x, y)$, $B_r(x, y)$ are the estimated color components of the background image; $R(x, y)$, $G(x, y)$, $B(x, y)$ are the color components of the current image; b_R , b_G , b_B are the set thresholds.

3.4 Shadow removal effect

In order to verify the feasibility of the shading removal algorithm based on the RGB color space, we conducted experiments on the image sequence taken. The resolution of the image sequence to be detected is 640×480 pixels, and the experimental platform is MATLAB software. The result of the detection is shown in Figure 3 (the picture is quoted from Shadow Detection: A Survey and Comparative Evaluation of Recent Methods).

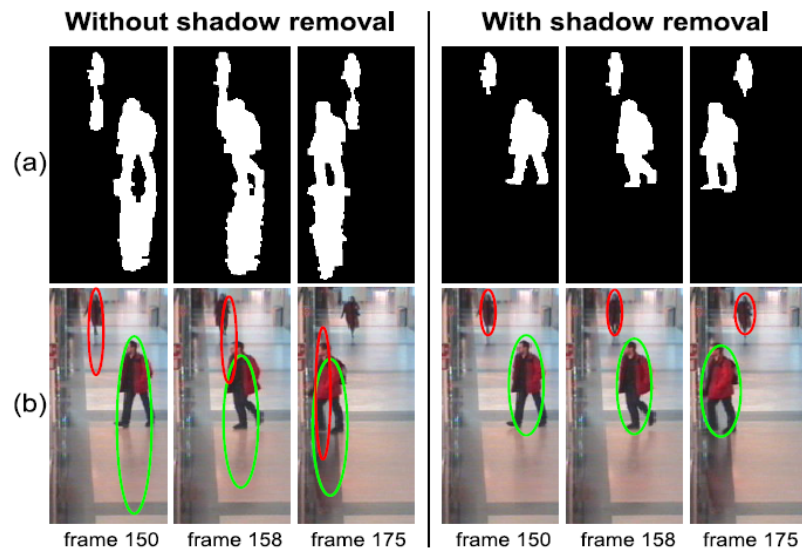


Figure 3. Schematic diagram of the results of sports video shadow removal

Through the formula, the foreground binary image with the shadow removed can be obtained. It can be seen from Figure 3 that the use of a shadow removal algorithm based on the RGB color space can effectively separate the shadow from the moving area, thereby obtaining a moving area that is more in line with the real target, which is convenient for subsequent tracking, recognition and other processing. Although the cast shadows are basically removed, there are also some target areas that are not part of the shadows but are very similar in color to the shadows, leaving holes in the target area.

4. Experimental Design

Table 1. Experimental environment for moving image key frame tracking

Environment type	Parameter	Value
Hardware environment	CPU	AMD6 core 3.0GHz
	RAM	32GB 512MB
	Graphics card	Win10 Professional Edition
	operating system	
Software Environment	development tools	Matlab 2017
Open-source library	OpenVC	

Table 2. Experimental subjects of moving image key frame tracking

Motion picture type number	Specific type	Quantity
1	pingpong	200
2	badminton	200
3	football	500
4	tennis	100
5	basketball	500

In order to test the tracking performance of the moving image key frame of the adaptive unsupervised clustering algorithm, a simulation test experiment was carried out on it. The experimental environment of the simulation test is shown in Table 1. In order to make the moving image key frame tracking results of the adaptive unsupervised clustering algorithm more convincing, five types of moving images are selected as experimental objects, and the number of each type of moving image is shown in Table 2.

In order to make the moving image key frame tracking results of the adaptive unsupervised clustering algorithm comparable, the moving image key frame tracking method of the semantic model and the moving image key frame tracking method of the Gaussian mixture model are selected for comparison experiments, and they are statistically compared to 5 categories. The tracking accuracy of moving image key frames is shown in Figure 4 (the picture is quoted from Shadow Detection: A Survey and Comparative Evaluation of Recent Methods).

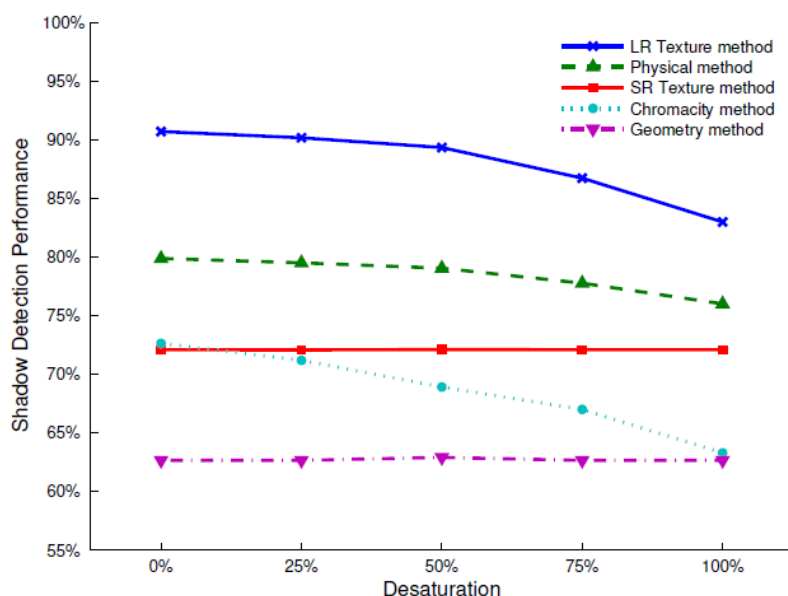


Figure 4. Comparison of tracking accuracy of moving image keyframes with different methods

A comparative analysis of the tracking accuracy of the moving image key frame in Figure 4 shows that the adaptive unsupervised clustering algorithm has the highest tracking accuracy of the moving image key frame, followed by the moving image key frame tracking method of the Gaussian mixture model, and the moving image key frame tracking The least accurate is the semantic model. The comparison results prove that the adaptive unsupervised clustering algorithm can accurately track the key frames of the moving image, and obtain the ideal moving image key frame tracking results. Although the color of human skin is very close to the color of the background, and there are some similar landmarks on the body that interfere with recognition, all 4 landmarks are accurately identified.

5. Conclusion

The motion tracking method in free background proposed in this paper can not only accurately identify the edge of the mark point of the human motion image, and accurately calculate the geometric center of the mark point, but also has strong robustness, is not sensitive to the experimental environment, and has low requirements on the experimental background. This method can not only be used for human motion analysis, but also has great application potential for automatic tracking of moving targets in other free backgrounds.

References

- [1] Yang Qiuju, Zhou Penghui. Automatic segmentation of auroral images based on machine learning. Chinese Journal of Geophysics, vol. 63, pp.15-20, January 2020.
- [2] Sun Dong, Song Yang, Cen Xuanzhen, et al. Research Progress of Unmarked Recognition of Sports Action Based on Computer Vision. Journal of Shanghai Institute of Physical Education, vol. 45, pp.16-22, September 2021.
- [3] Wei Tong, Zhou Yinhe. Blind track image positioning based on machine learning recognition and labeled watershed segmentation. Optics and Precision Engineering, vol. 15, pp.10-19, January 2019.
- [4] Huang Xiaolong. Research on Feature Detail Recognition of Astronomical Photoelectric Images Based on Machine Learning. Automation and Instrumentation, vol. 4, pp.4-11, June 2020.
- [5] Lu Laibing, Bruce Lee. The clustering and evolution of artificial intelligence technology in the field of international sports. Journal of Shandong Institute of Physical Education, vol. 36, pp. 12-19, March 2020.
- [6] Li Jing, Fan Wenliang, Lei Ziqiao, et al. Recognition of Parkinson's patients based on machine learning model of magnetic resonance diffusion tensor imaging. Chinese Medical Equipment, vol. 36, pp.45-47, October 2021.
- [7] Ye Chunming, Zhao Shengwen, Yang Xiuhong, et al. Analysis and prediction of young athletes' ability to respond to the new crown pneumonia epidemic based on machine learning. Journal of Physical Education, vol. 27, pp. 68-70, March 2020.