International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.2 (2016), pp. 23-32
http://dx.doi.org/10.14257/ijmue.2016.11.2.04

Research on Target Detection in Sports Video

Gang Liu[1], Yang Liu[2], Yang Wang[1,*] and Jingmeng Sun[1]

[1] Physical Education Department, Harbin Engineering University, Harbin 150000, China
[2]Harbin Sport University, Harbin 150001, China
Liugangt@hrbeu.edu.cn

Abstract

For moving objects detection, a background subtraction algorithm based on adaptive Gaussian mixture model is proposed in order to extract moving regions. The OTSU algorithm is researched in order to adapt to the changes in the background images; In order to accelerate model updating rate, a novel mechanism is the combination of expected sufficient statistics and L-recent window.

Keywords: *moving object detection, background subtraction, Gaussian mixture model*

1. Introduction

The paper introduces several traditional detection methods and makes comparative analyses [1-2]. On that basis, for the features of sport videos, it proposes motion object detection method based on background subtraction. The method initializes background images by Gaussian mixture modeling approach and seeks the threshold by OTSU; then, it restores detection area with connected area analysis method and improves the accuracy rate of moving object detection in volleyball videos; finally, it proves the effectiveness of the mentioned method with experimental results [3-4].

2 The Traditional Detection Method

The motion object detection is the foundation for all kinds of subsequent advanced treatment [5-6]. The objective is to extract from image sequences the foreground which moves against background and split two-dimensional image features like edge, texture and grey scale to independent objects. In practical applications, to avoid camera shake, illumination change and noise interference caused by other environmental factors, it's required to eliminate the interference of false targets. The object detection and extraction are carried out in steps as seen in Figure 1. Hereunder we introduce some common motion object detection methods.

2.1 Optical Flow

Optical flow was proposed by Gibson in 1950. It is the speed of image mode movements. Optical flow is a two-dimensional instantaneous velocity field. The fundamental principle of optical flow detecting motion objects is: assign one velocity vector to each pixel point on the image to form one image motion field; at given motion movement, points on the image corresponding to those on three-dimensional objects; according to motion vector features of each pixel point, make dynamic analysis of images; if no motion objects are found in images, the light flux vector changes continuously in the whole image area; if relative movements exist between objects and image background, the velocity vector of moving objects differs from neighborhood background velocity vector and thus the positions of moving objects are found out.

ISSN: 1975-0080 IJMUE

Optical flow method is able to detect single moving objects without any known information about scenarios and is applicable to detect camera motions. However, most optical flow approaches are of complicated calculation and bad real-time. It's used very few when cameras are static.

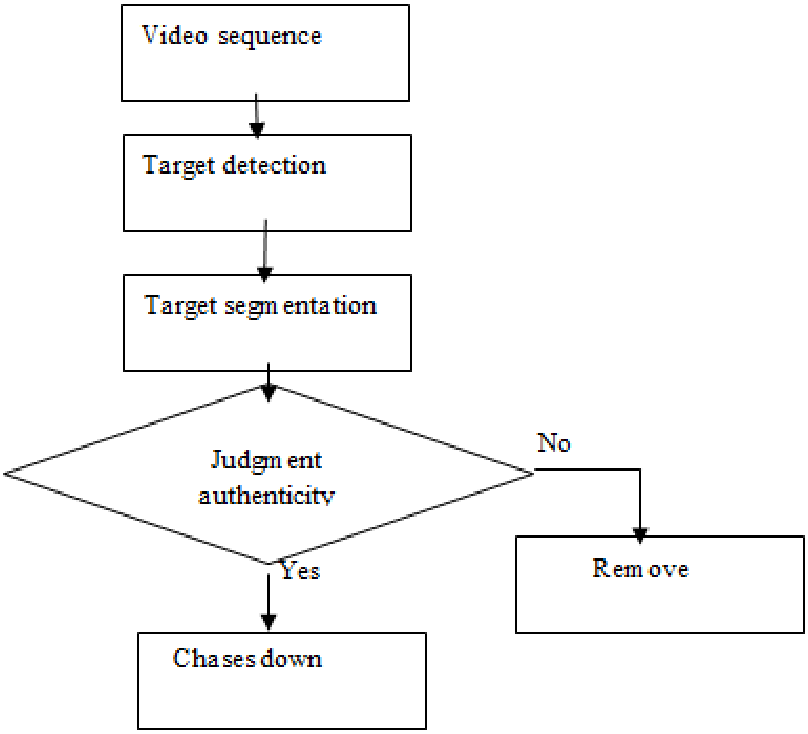

Figure 1. The Block Diagram of Target Detecting and Picking-Up

2.2 Frame Subtraction

Frame difference method has the basic ideas: use two or more consecutive frames in video sequence for subtraction; if the difference value is very small, it's believed that no motion objects pass over the point; otherwise, there are motion objects and calculate the difference between the kth and the (k-1)th image to get differential images D_k, which are then binarized D_k. The principle of this method is described as equation (1) and its work flow is shown in Figure 2.

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

$$R_k(x,y) = \begin{cases} 1 \ \ Foreground \ \ \text{if} \ \ \mathrm{D}_k(x,y) \geq \text{Threshold} \\ 0 \ \ \text{Background} \ \ \text{if} \ \ \mathrm{D}_k(x,y) < \text{Threshold} \end{cases}$$

(1)

Obviously from the principle procedure, frame difference method is easily implemented. Being of less complicated program design, it's easily implemented for computer programming. The instantaneity is easily sufficed. Also it has favorable adaption to the dynamic environment. Due to short time interval between adjacent frames, the method is not susceptible to the changing lights of scenarios and shadow. However, empty spaces appear easily in moving objects especially when the target moves too fast, the precision of extracting objects is affected.

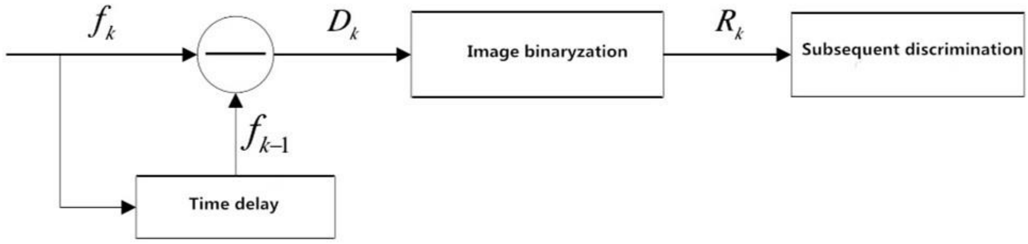

Figure 2. The Schematic Diagram of the Frame Subtraction Method

2.3 Background Subtraction

Background subtraction method's principle is simple. It uses the current frame image f_k to subtract stored or background image B_k obtained after real-time update and then binarizes D_k the difference image; if pixel difference is bigger than one threshold, it's considered that the pixel is motion object; otherwise, it's background. The principle is described as equation (2). Its work flow is shown in Figure 3.

$$D_k(x,y) = | f_k(x,y) - B_k(x,y) |$$
$$R_k(x,y) = \begin{cases} 1 \ \ Foreground \ \ \text{if} \ \ \mathrm{D}_k(x,y) \geq \text{Threshold} \\ 0 \ \ \text{Background} \ \ \text{if} \ \ \mathrm{D}_k(x,y) < \text{Threshold} \end{cases} \quad (2)$$

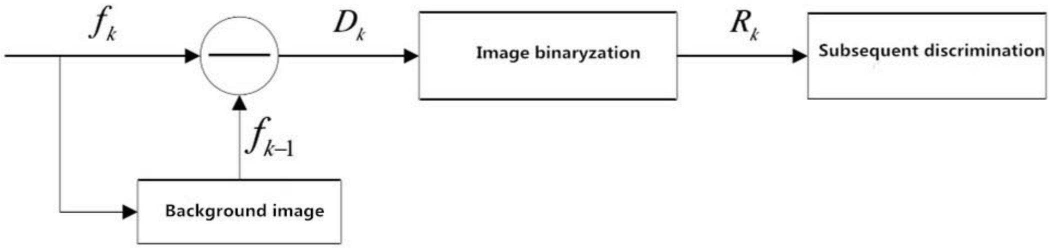

Figure 3. The Schematic Diagram of the Background Subtraction Method

When the camera is fixed, the method can get directly such information as the position, size and shape of motion objects. Its calculation complexity is moderate and it meets the requirement of instantaneity. However, it's sensitive to changes of dynamic scenarios, like camera movement and other moving objects. On that basis, we proposed adaptive background difference method, which allows background to renew adaptively with time and adapt to tiny variations of dynamic scenes.

3. The Proposed Motion Object Detection Method

After comparison of the above three methods and based on features of volleyball videos, we suggest using background subtraction method to detect objects. Firstly, get background model with Gaussian mixture modeling approach; next, eliminate noises by median filtering of each frame image, perform background subtraction operation to detect motion targets with the adaptive OTSU threshold segmentation method, and renew in real time background model to fit dynamic changes of scenes; lastly, perform mathematical morphological alterations of detection results to discern accurately targets.

3.1 Create Background Model with GMM

3.1.1 Gaussian Mixture Model (GMM)

In the case of fixed camera, if background is static, each pixel in background image can be described as a Gaussian distribution. To present dynamic changes in scenes, use K mixed Gaussian distributions to represent each pixel. Gaussian mixture model utilizes the basic ideas: for each pixel, define K statuses of it and each status is expressed by one Gaussian function; some of those statuses stand for background model and the others mean motion foreground model [7-8].

Set the value X_t of one pixel at the time point t, whose probability density function is put as equation (3):

$$p(X_t) = \sum_{i=1}^{K} \frac{\omega_{i,t}}{(2\pi)^{n/2} \left|Cov_{i,t}\right|^{1/2}} e^{-\frac{1}{2}(X_t-\mu_{i,t})^T \Sigma_{i,t}^{-1} (X_t-\mu_{i,t})} \tag{3}$$

To enhance real-time performance of background model, we rank K Gaussian distributions in descending order as per the size of ω_k / σ_k value. By referring to threshold T, we can decide the first b Gaussian distributions as background distribution, in the following:

$$B = \arg\min_b \left(\sum_{i=1}^{K} \omega_{i,t} > T\right) \tag{4}$$

3.1.2 Model Matching

In the method, we make the maximum quantity of Gaussian distribution of each pixel is $K_{\max} = 4$; the number of Gaussian model of each pixel is initialized $K = 1$, with the pixel value of each point in the first frame to initialize Gaussian distribution mean; the standard deviation is made bigger and mixed Gaussian weight is $1/K_{\max}$.

Match the current pixel value X_t respectively with the previous b distribution models. When matching, we use 2.5x standard deviations as the matching criterion. So the model which meets $|\mu_{i,t} - X_t| < 2.5\sigma_{i,t}$ is identified as matching model and others as non-matching models.

3.1.3 Parameter Updating

Looked back to the matching principle mentioned above, if in the current pixel, there is ith Gaussian distribution matching with it, the pixel is labeled as background pixel;

$$\begin{aligned} \mu_{i,t+1} &= (1-\beta_i)\mu_{i,t} + \beta_i X_t \\ \sigma_{i,t+1}^2 &= (1-\beta_i)\sigma_{i,t}^2 + \beta_i(\mu_{i,t} - X_t)^2 \end{aligned} \tag{5}$$

If successful matching, update parameters of the distribution model; if there's no any Gaussian distribution matching with it, the pixel is put as foreground pixel and update background model's parameters by the following rules:

(1) When $K < B$ add one Gaussian distribution whose mean is based on the current pixel value;

(2) When $K \geq B$, use Gaussian distribution whose mean bases on the current pixel value to replace the one with biggest covariance in the mixture model; the new added Gaussian distribution has smaller weight and bigger covariance before initialization.

After renewal of the model, the weights of K Gaussian distributions require update.

$$\omega_{i,t+1} = (1-\alpha_i)\omega_{i,t} + \alpha_i M_{i,t} \tag{6}$$

3.2 Background Updating Criteria

The traditional adaptive Gaussian mixture model can cope robustly with secular changes in dynamic scenes and interferences of various motion objects. However, the model learning speed is too slow at the initial stage, particularly when the environment is too complicated. To quicken its learning speed and maintain its robustness for dealing with dynamic scenes [9], we take advantage of the solution combining expected sufficient statistics and L-recent window to improve the model. The solution has two parts:

(1) For the first L frame images, it takes expected sufficient statistical updating approach to accelerate the model's convergence rate;

(2) For images after L frame, it uses L-recent window updating way.

The expected sufficient statistical updating way is described as follows:

$$\mu_{i,t+1} = \mu_{i,t} + \frac{M_{i,t}}{\sum_{i=1}^{t} M_i}(X_t - \mu_{i,t}) \tag{7}$$

$$\sigma^2_{i,t+1} = \sigma^2_{i,t} + \frac{M_{i,t}}{\sum_{i=1}^{t} M_i}\left[(\mu_{i,t} - X_t)^2 - \sigma^2_{i,t}\right] \tag{8}$$

$$\omega_{i,t+1} = \omega_{i,t} + \frac{1}{t}(M_{i,t} - \omega_{i,t}) \tag{9}$$

3.3 Selection of OTSU Threshold

The selection of threshold [10-11] is one key of background difference method. Selecting too high or low value of threshold will influence the extraction results of motion objects. Many algorithms can define the threshold. Here we introduce five common methods: p parametric method, grey histogram peak-valley method, dual fixed threshold method, differential histogram method and discriminative analysis technique.

3.3.1 P parametric Method

Make the area S_0 S of original binary image and the area S of split object. Give out one value of $\rho = S_0 / S$ and calculate image's histogram. We can largely get the threshold T.

3.3.2 Gray Histogram Peak-Valley Method

Choose gray value of valley position as threshold of binary image to achieve the segmentation of target and background. This method chooses threshold quite visually. If big differences exist between targets in the given image and background grey scale, the method is simple; but for complicated background images, the method is not workable.

3.3.3 Dual Fixed Threshold Method

Use two fixed thresholds T_1 and T_2. Compare grey level $g(x, y)$ separately with them:

(1) If $g(x, y) < T_1$, points in binary images marked with 0;

(2) If $g(x, y)$ is between [T_1, T_2], points in binary images marked with 1;

(3) If $g(x, y) > T_2$, points in binary images marked with 0.

This technique can be applied when several targets appear in the scenes. It is shown in equation (10-11).

$$g(x, y) = \begin{cases} 0 & g(x, y) < T_1 \\ 1 & T_1 \leq g(x, y) < T_2 \\ 0 & g(x, y) \geq T_2 \end{cases} \tag{10}$$

$$g(x, y) = \begin{cases} 1 & g(x, y) < T_1 \\ 0 & T_1 \leq g(x, y) < T_2 \\ 1 & g(x, y) \geq T_2 \end{cases} \tag{11}$$

3.3.4 Differential Histogram Technique

When boundaries between objects in the image and background are where gray scale changes rapidly, such as edges, it's impossible to use directly the grey value of images but the differential value to determine that value.

3.3.5 Discriminative Analysis Method

The method has basic ideas like: select one optimal separation threshold T; then use T to divide grey collection of gray histogram into two groups, of which one's gray value below T and the other's above T; when the ratio between inter-group variance and intra-group variance of two groups of gray mean reaches the maximum, the obtained gray separation threshold T is the best solution.

Here we use OTSU method to improve and complete threshold selection of binary images. OTSU, renamed maximum inter-class variance algorithm [12-13], was proposed by Japanese scholar dajin in 1979, so it's also called dajin method. It has been always considered as a kind of non-parameter non-supervision selective method of threshold. The algorithm analyzes histogram of input grey images, separating the histogram into background and foreground. The division point to maximize the variance between the two parts is the acquired threshold. Whether partial objects are falsely discerned background or partial background is mistaken as objects, it will result in smaller variances. Hence, the maximum segmentation of inter-class variance means the minimal possibility of wrong determination.

The theory of maximum inter-class variance method is: set image's gray range $\{0,1,2...l-1\}$; the number of pixel points with gray value i is n_i; image's total pixel points $N = \sum_{i=0}^{l-1} n_i$; the probability of pixels with gray value i is expressed like:

$$P_i = n_i / N , P_i > 0, \sum_{i=0}^{l-1} P_i = 1 \tag{12}$$

The volleyball videos used here are of resolution 720*480 at 25 fps. After running the program, we get selected threshold in Table1.

In the whole experiment, the video processing frame rate can reach up to 20fps, sufficing the requirement of real-time. From object detection results, it's noticed that complete motion objects are fetched with adaptive threshold. The method provides ideal segmentation threshold. To enhance instantaneity, we can change the algorithm, rather than computing threshold in each frame.

Table 1. The Threshold of OTSU Algorithm

The current frame	OTSU threshold	The current frame	OTSU threshold	The current frame	OTSU threshold
30	82	45	80	60	81
75	85	90	88	105	90
120	78	135	83	150	82
165	89	180	79	195	91
210	88	225	82	240	85
255	81	270	85	285	87
300	85	315	83	330	86

3.4 Analysis of Connected Areas

After morphological treatment of images, some small interfering areas were removed and tiny gaps and cavities were padded. However, there are still big white areas or black cavities. To fill up them, we should firstly estimate the surface of each connected black area. If one black space's area is smaller than one threshold, the space is changed to white region. After that, we can calculate the coverage of each connected white area. If white space's area is bigger than one threshold, it's judged that background has changes.

4. Experimental Analysis and Results

4.1 Detection of Motion Targets

We detect players in volleyball videos and use the results to prove effectiveness of the proposed method. We utilize it to process video images. The detection effect of players is shown in Figure 4.

(a) Thirtieth Frames (b) Background Map (c) Detection Map

Figure 4. The Result of Our Improve Detection Algorithm

4.2 Evaluation of Quality

To validate effectiveness of the method here, we analyze and evaluate its detection effects by the following rules, regarding the target detection of different video sequences:

$$\text{Precision} = \frac{\text{correct number of detected target area}}{\text{total number of detected target area}} \times 100\%$$

$$\text{Recall} = \frac{\text{correct number of detected target area}}{\text{total number of the target region of the frame image}} \times 100\%$$

As for the detection of players in video fragment 1, we make 100 consecutive frames as statistical unit and divide into 10 groups to assess the effect. Evaluation results are given in Figure 5, where axis X is group and Y is percentage of detection result. Table2 sums up moving detection results in different video scenarios, of which clip 1 is detection result of players and clip 2 is of volleyball.

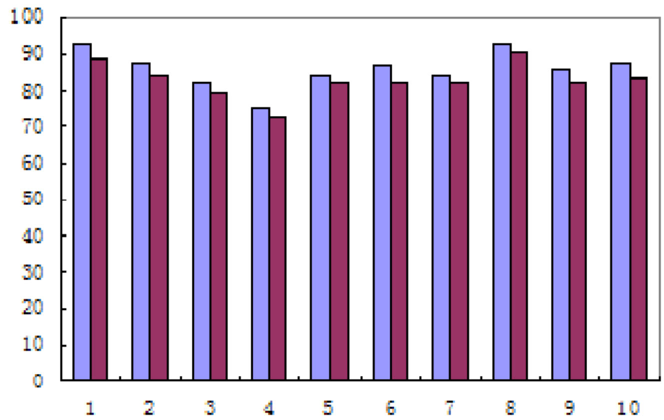

Figure 5. The Statistic of Player Detection

Table 2. The Result of Player Detection

Video sequence	Number of frames	Average recall	Average precision
Clip1	1200	92.3	90.8
Clip2	2300	94.7	91.5

We can see from Figure 5 the detection effect varies from different groups. Experiments indicate that when players are at near distance, the detection result is affected greatly. Table2 lists out detection results of players and volleyball in different video sequences. The discussed algorithm has good real-time performance and effectiveness. Recall and precision rates in clip 1 are relatively low because changing background is quite close to players who are trying to follow the ball.

5. Conclusion

We discussed the application of the motion object detection algorithm in detecting players and volleyball in such videos. It firstly introduced a few traditional similar techniques and compared their merits and shortcomings. Then, based on features of each and for the goal of this research, we proposed adaptive background subtraction method based on Gaussian mixture model to discern moving targets. By improving Gaussian mixture model and creating new model, more accurate and more real-time updating was realized. Experiments demonstrated the satisfactory detection effect and sound robustness, laying good foundation for the future research aim.

Acknowledgement

This work was supported by The Fundamental Research Funds for the Central Universities. No. HEUCF151601.

References

[1] L. Diping, "Detection and motion recognition method of human body posture based on video", Central South University, **(2012)**.
[2] Q. Zhen, "Research on the key technology of video multi modal information processing", Harbin Engineering University, **(2012)**.
[3] G. Huiwen, "Study on fast multi target detection and tracking in video surveillance", Hunan University, **(2013)**.
[4] L. Tao, "Research on detection and tracking of the ball in soccer video", Xi'an Electronic and Science University, **(2012)**.
[5] D. Zhichao, "Study on target detection and classification in video sequences", Xi'an Electronic and Science University, **(2012)**.
[6] S. Chen, "In video surveillance preprocessing, target detection and tracking method research", Nanjing University of Posts and Telecommunications, **(2014)**.
[7] Z. J. Chen and X. J. C. Ho, "Moving target detection based on improved mixture Gaussian model", Journal of image and graphics, vol. 12, no. 9, **(2007)**, pp. 1585-1589.
[8] W. Lijuan and O. C. V. Gauss, "Moving target detection based on hybrid modeling", Electronic test, vol. 9, no. 9, **(2009)**, pp. 86-89.
[9] W. Yong, T. Y. Hua and T. J. Wen, "Based on shadow elimination and mixed Gaussian model of video segmentation algorithm", Opto electronic engineering, vol. 35, no. 3, **(2008)**, pp. 21-25.
[10] F. Zhongliang, "Image threshold selection method China structure", Journal of image and graphics, vol. 5, no. 6, **(2000)**, pp. 466-469.
[11] W. Yong, H. Yuanjun and C. Hongming, "Optimal threshold selection algorithm in edge detection based on wavelet transform", Image and Vision Computer, vol. 23, no. 13, **(2005)**, pp. 1159-1169.
[12] Y. Bazi, L. Bruzzone and F. Melgani, "Image thresholding based on the EM algorithm and the generalized Gaussian distribution", Pattern Recognition, vol. 40, no. 2, **(2007)**, pp. 619-634.
[13] W. Xiangke and Z. Zhiqiang, "Otsu multi threshold fast segmentation algorithm and in the color image", Computer applications, vol. 26, **(2006)**, pp. 14-15.

Author

Liu Gang, He is an Associate Professor at Physical Education Department of Harbin Engineering University. He is in the research of physical education and training.