

Moving Target Detection and Tracking based on Camshift Algorithm and Kalman Filter in Sport Video

Baojun Zhang^{*}

Department of Physical Education, Harbin Institute of Technology, Harbin, 150001, China

Abstract

With the rapid growth of the video data's amount, how to efficiently retrieve useful information has become very urgent. As the base of video indexing and searching, video annotation has great significance for its application prospect and research value. In the semantic detection, moving object detection and tracing is the basis. In the paper, adaptive Gaussian Mixture Model is used to background model; Camshift and Kalman filter are used to trace the players and ball. The implement of the algorithms is all based on Visual C++ and Visual c#2008. OpenCV and Aforge.net class base are also used. Experimental result shows that the method annotates well.

Keywords: Sport Video; Kalman filter; Camshift algorithm; object detection and tracing

(Submitted on October 19, 2018; Revised on November 17, 2018; Accepted on December 23, 2018)

© 2019 Totem Publisher, Inc. All rights reserved.

1. Introduction

At present, the number of images and videos on the network is increasing by geometric series. The rapid expansion of the network visual information, the development of society and people's daily life is having a profound impact. How to organize and manage visual information effectively so that users can quickly check the vast amount of visual information is a problem. In order to solve this problem as well as the issue of being able to find the necessary information, visual information retrieval Image and video than text, voice and other media needs to carry more information. However, other media sources need to be more intuitive and vivid so that users will be more open to using these sources [1-2]. Although unstructured degree of image and video is relatively high, its expression also has ambiguity, which brings difficulties in image retrieval and video [3-4].

At present, the main methods used in video annotation can be divided into the following categories: hand annotation, video annotation from several rules and video annotation based on machine learning. The system is easy to implement, but it is time-consuming, laborious and easy to introduce human error. In addition, different people have different standards of judgment, and the differences of these criteria will also affect the consistency of annotation [5]. The rule based video annotation method first uses the expert knowledge in a certain field to establish the classification rules, and then the video annotation. A method for aircraft based tagging learning video is based on training samples or video data manually annotated by learning to build all the various semantic conceptual model or conceptual difference model, and then use the model to complete the label of other video data [6].

Sports video annotation has a wide range of applications, but also contains important academic value. It is regarded as a separate subject by many researchers. The core technology is the analysis of the relationship between events and events. The related events are annotated and organized, and the query needs of most users can be satisfied [7]. From the academic point of view, event detection and recognition is a typical problem of pattern recognition and computer vision. In sports videos, the event is an order structure target with high level abstract semantics, which is very challenging. The difficulty mainly lies in the fact that video is a multidimensional signal; it has both time and space information, so the data is large and rich in content.

^{*} Corresponding author.

E-mail address: baojunzhangbjz@sina.com

The semantic event in sports video is a dynamic process. It changes with time, so the static target has more complex patterns and variability. There are different definitions in different sports competitions, which are different from their internal logic or causal structure. If we can find a way to solve these problems, and construct a sports video content analysis system, we can provide reference for other similar semantic analysis problems, and then promote the development of the whole video annotation and retrieval. Since the league's push, Chinese players entering the receiving basketball have become more and more welcomed by the masses. It has also become an important part of people's lives. It is time-consuming and laborious work to find the video segments of interest in a large amount of basketball video. For the general audience, the point of interest lies in the highlights of the game, and for the field of basketball experts, their point of interest lies in the ability to reflect the players' skills and team tactics video segment [8-9].

Therefore, it is very important to study the video annotation system of basketball game, which will enable the user to get the video segment. Shot boundary detection is designed to complete shot segmentation. There are two types of boundary mutations and gradients between the shot and the shot. A shot change is the process of switching from one camera to the next. The shot has a process of switching. The key of the shot boundary detection is to distinguish between the camera and the camera and object motion caused by changes in the shot [10].

Motion target detection is foundation to target tracking. The existing methods for detecting moving target include optical flow approach, frame difference method and background subtraction method. Of them, the optical flow approach estimates motion field of image and merges similar motion vector makes a distinction between foreground and background. Frame difference method uses the difference of temporal serial image to realize detection of motion target. Background subtraction solution detects moving target with the combined use of image sequence and background model. Here, we introduce the background subtraction method based on adaptive Gaussian mixed model.

2. Moving Object Detection

2.1. Gaussian Mixed Model (GMM)

Assume λ is an M-Gauss Gaussian mixed model [11]. For a feature vector of $n \times 1$ dimension, its likelihood function $P(x | \lambda)$ is defined as Equation (1).

$$P(x | \lambda) = \sum_{i=1}^M w_i p(x | \lambda_i) \quad (1)$$

The function is linear aggregation of M single modal Gauss $p(x)$, where W is weighting coefficient. Each $p(x)$ includes one $n \times 1$ dimensional mean vector μ_i and $n \times 1$ covariance matrix Σ_i . Equation (2) can be expressed as follows:

$$P(x | \lambda_i) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu_i)^T (\Sigma_i)^{-1} (x - \mu_i)\right) \quad (2)$$

From this, Gaussian mixture model can be described with each Gaussian weight, mean vector and covariance matrix, i.e. λ can express $\lambda = \{w_i, \mu_i, \Sigma_i\}$ by a triple.

2.2. EM (Expectation Maximization) Algorithm

Set λ is initial value of GMM parameter; $\bar{\lambda}$ is parameter after optimization of steps. Function Q is constructed as Equation (3).

$$Q(\bar{\lambda} | \lambda) = \sum_{t=1}^T Q_t(\bar{\lambda} | \lambda) = \sum_{t=1}^T \sum_{\theta_t=1}^M P(\theta_t | x_t, \lambda) \quad (3)$$

EM algorithm includes two steps:

(1) In the case of an original given model, estimate function Q to get statistical value; this is called E (Expectation) step.

(2) Optimize function Q to get estimation of new model; this is called M (Maximization) step. $L(\bar{\lambda} | X) \geq L(\lambda | X)$ is satisfied after each iteration. After numerous iterations, model parameter will converge. If iteration repeats too many times, model parameter will approximate too much the space distribution of training data, causing that its ability to describe testing data space distribution becomes bad, i.e. over-training. Generally after 4-5 iterations, the ideal result will be achieved.

2.3. Moving Target Detection Method based on Adaptive Gaussian Mixture Model

We utilize adaptive Gaussian mixture model to do background modeling and target detection [12]. First, determine approximate distribution of playground's main color; then, fit it with Gaussian mixture model, update model parameter with EM algorithm. The steps are described below.

2.3.1. Creation of Background Model

Set one main color pixel P of playground has value $(x_1, \dots, x_i, \dots, x_l)$ at point $(1-t)$, in which x_i is Hue value; used K Gaussian distributions are used to depict those pixel values. The probability density function of x_i is described in Equation (4).

$$p(x_i) = \sum_{i=1}^k w_{i,i} \cdot G(x_i, \mu_{i,i}, \sigma_{i,i}^2) \quad (4)$$

2.3.2. Moving Target Detection

For observational value x_{t+1} of pixel p at point $(t+1)$, we can decide its background point or target point in following way:

(1) Calculate in order the distance between observational value and each Gaussian distribution until the first Gaussian distribution is found which suffices $d_i < T_2$; T_2 empirical value is 2.5. It is shown in Equation (5).

$$d_i = \sqrt{(x_{t+1} - \mu_{i,i})^T \Sigma^{-1} (x_{t+1} - \mu_{i,i})} \quad (5)$$

If no Gaussian distribution is found to meet the condition, use a new Gaussian distribution to replace the one with minimum confidence. x_{t+1} is used as mean value of the new distribution, which has bigger covariance and smaller weight. If Gaussian distribution j meeting condition is found, judge whether j belongs to one of g background distributions. If yes, the pixel point is background point; otherwise, it's target point.

(2) Update parameter of each Gaussian distribution; weight of distribution is updated in the following manner. It is shown in Equation (6).

$$w_{t,i} = (1-a) \cdot w_{t-1,i} + a \cdot M \quad (6)$$

(3) Each updated Gaussian distribution is placed in descending order per \tilde{w}/σ value as to get new reliable background distribution waiting for the detection at next time point.

The detected targets include player, basketball and non-dominant color area. The kick-off circle and foul lane can be recognized and removed because their connecting regional area feature is very noticeable. To complete detection of moving targets like player and basketball, other players and basketball can be detected with target recognition algorithm based on Hu invariant moment.

3. Tracking of Motion Target

In basketball match videos, there are players, basketball and referee. They're moving in different ways. If only one tracking method is employed to trail all objects, it's not possible to make better effect. So for different target, we use different ways to follow up [13-14].

3.1. Moving Target Tracking based on Camshift

Tracking the moving target, as one branch of computer visual research, has become an important part in video processing research in recent years. The research results are widely used in man-machine interaction video monitoring sites. Classical tracking methods include template matching, optical flow, filtering method, Mean-shift algorithm based on color histogram feature distribution, Camshift algorithm and particle filter tracking algorithm.

In basketball match video, players have characteristics like:

- (1) Big target, obvious color feature;
- (2) Moving not faster than basketball;
- (3) Occlusion among them. Based on those features, we choose Camshift algorithm to implement tracking of players.

Camshift algorithm (Continuously AptiveMeanShift) is an improved Mean-shift strategy. It regards the chromatic information of tracked target as feature. Firstly, it calculates and processes chromatic information; then, it projects to the next frame image, seeks out the target, updates source image, and analyzes next frame image. Repeat those steps to realize continuous tracking of target. Before each search, it's required to make initial value of searching window, which is current location and size of moving target.

The searching window of Camshift algorithm lies in nearby area where motion target may appear, reducing the calculation amount of search and thus saving searching time. So, it has good instantaneity. Camshift algorithm searches moving target through color feature matching. Since color information of moving target does not obviously change, it has favorable robustness. Camshift algorithm is an improvement of Mean-shift algorithm. We make a brief introduction of Mean-shift algorithm.

3.1.1. Introduction of Mean-Shift Algorithm

Mean-shift algorithm is currently a popular parameter-free fast matching algorithm. It has merits like small calculation amount, good real-time, modeling with kernel function histogram, insensitive to edge occlusion and uneven background movement. Mean-shift was firstly proposed by Fukun et al. in 1975. It originally means off-set mean vector. With theoretical evolution, its concept changed as well. Mean-shift algorithm is an iterative step. First, calculate shifted mean value of current point and move it to the shift mean point; then, use it as a new starting point and continue moving until a certain condition is met and the iteration completes.

Mean-shift algorithm is a non-parameter kernel density gradient estimation algorithm. It searches targets by obtaining the extreme value of probability distribution through iterations. This algorithm has steps as follows:

- (1) Select searching window in the image and window size is s .
- (2) Calculate centroid of searching window in the image.

Calculate the zero order matrix. It is shown in Equation (7).

$$M_{00} = \sum_x \sum_y I_{(x,y)} \quad (7)$$

Matrix of x and y is shown in Equation (8).

$$\begin{aligned} M_{10} &= \sum_x \sum_y x I_{(x,y)} \\ M_{01} &= \sum_x \sum_y y I_{(x,y)} \end{aligned} \quad (8)$$

Search window centroid location. It is shown in Equation (9).

$$x_c = \frac{M_{10}}{M_{00}}, y_c = \frac{M_{01}}{M_{00}} \quad (9)$$

3.1.2. Implementation of Camshift Algorithm

Camshift algorithm is an improved method based on Mean-Shift algorithm. It's implemented in these ways:

Step 1 Calculate hue histogram of searching area in the current frame image.

Step 2 Calculation of backward projection.

Step 3 Mean-Shift projection.

Step 4 CamShift projection.

(1) CamShift Feature extraction

Camshift algorithm uses day-mark hue histogram in back projection calculation to get the back projection image. Back projection image refers to calculating the hue histogram of source image and then getting hue probability distribution projection image through convertible calculation. The process is described below.

Step 1 Convert source image to HSL color space and get its Hue value.

Step 2 Calculate histogram of Hue value. Suppose the used histogram is N-level, the position of n pixels is defined as $\{x_i\}$. Histogram is defined as $\{\bar{q}_u\}$; define Hue value of pixel whose location is x_i as $h(x_i)$. We obtain the histogram distribution, as shown in Equation (10).

$$\bar{q}_u = \sum_{i=1}^n \delta[h(x_i) - u] \quad (10)$$

Step 3 The scale of acquired straight grid box is converted from $[0, \max(q)]$ to $[0, 255]$ by the Equation (11).

$$\{p_u = \min(\frac{255}{\max(q)} \bar{q}_u, 255)\} \quad (11)$$

Step 4 Substitute the value u of pixel x_i in source image with the resultant p_u ; deal with image in the same way. The new projection image is obtained.

(2) Target searching of the algorithm

In the above Mean-Shift algorithm, we mentioned how to determine the centroid location of searching window. After the centroid location is found, use it to update new central location. Then, use the updated location for calculation to get a new centroid position; repeat until centroid converges. Until now, the centroid location is position of target center in the image frame; that is, target searching process of Meanshift algorithm. Camshift algorithm's target searching invokes Meanshift algorithm's target searching process.

Camshift algorithm employs elliptic region to track target. The long axis, short axis and direction angle of ellipse are calculated through Equation (12)-(14).

Step 1 Compute the two matrix of the target. It is shown in Equation (12).

$$\begin{aligned}
M_{20} &= \sum_x \sum_y x^2 I_{(x,y)} \\
M_{02} &= \sum_x \sum_y y^2 I_{(x,y)} \\
M_{11} &= \sum_x \sum_y xy I_{(x,y)}
\end{aligned} \tag{12}$$

Step 2 Calculate of process parameters. It is shown in Equation (13).

$$\begin{aligned}
a &= \frac{M_{20}}{M_{00}} - x_c^2 \\
b &= 2(\frac{M_{11}}{M_{00}} - x_c y_c) \\
c &= \frac{M_{02}}{M_{00}} - y_c^2
\end{aligned} \tag{13}$$

Step 3 Calculate the long axis of the ellipse, the short axis and the direction angle. It is shown in Equation (14).

$$\begin{aligned}
l &= \frac{\sqrt{(a+c)} + \sqrt{b^2 - (a-c)^2}}{2} \\
w &= \frac{\sqrt{(a+c)} - \sqrt{b^2 - (a-c)^2}}{2} \\
\theta &= \frac{1}{2} \tan^{-1}(\frac{b}{a-c})
\end{aligned} \tag{14}$$

(3) Player tracking based on Camshift algorithm

The paper adopts Camshift algorithm to trail players. Player region is obtained through moving target detection and recognition; it's considered as initial searching window of Camshift algorithm. When choosing color feature, we select only the most noticeable upper part of players (not including head). Then, the improved Camshift algorithm based on background weighted histogram is utilized to track players. It proves that the method has a better tracking effect.

3.2. Motion Target Tracking based on Kalman Filter

3.2.1. Principle of Kalman Filter

Kalman filter is a linear recursive filter. It can make optimal estimation of next system state by relying on system's previous state sequence. Kalman algorithm mainly includes state equation and observation equation.

State equation. It is shown in Equation (15).

$$x_k = \phi_{k,k-1} x_{k-1} + W_{k-1} \tag{15}$$

Observation equation. It is shown in Equation (16).

$$Y_k = H_k x_k + V_k \tag{16}$$

In the paper, we suppose system is a determined one. $\phi_{k,k-1}$ and H_k are known, and W_{k-1} and V_k meet assumption terms and are known. Set P_k is posterior misestimate covariance matrix of x_k and \hat{x}_k ; P'_k is prior error covariance x_k and \hat{x}_k matrix. Kalman filter calculates in following steps:

Step 1 t_0 time, \bar{x}_k and P_0 .

Step 2 t_k time, the system state prediction Equation (17) is:

$$\bar{x}_k = \phi_{k,k-1} \bar{x}_{k-1} \quad (17)$$

Step 3 t_k time, the system state update Equation (18) is:

$$\bar{x}_k = \phi_{k,k-1} \bar{x}_{k-1} + K_k (Y_k - H_k \phi_{k,k-1} \bar{x}_{k-1}) \quad (18)$$

3.2.2. Basketball Tracking based on Kalman Filter

In basketball match videos, changes of basketball movement and camera angle lead to divergent movement in basketball area in the image. Generally, since time interval between adjacent image frames is very short, only speed is enough to indicate the moving tendency of basketball. We use basketball position and speed as moving parameter of basketball; define system state of Kalman filter as $(xs_k, ys_k, vx_k, vy_k)^T$. In the image, only the target's position is observed; hence, the observation vector is defined as $(xw_k, yw_k)^T$.

Assume basketball moves at constant speed in unit time interval. Define state shift matrix $\phi_{k,k-1}$ as Equation (19).

$$\phi_{k,k-1} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (19)$$

The basketball motion estimation algorithm based on Kalman filter is described like:

(1) With basketball position and speed acquired above, we assign initial value to x_0 and assign initial value to P_0 , Q_0 and R_0 . When the initial speed is unknown, it's replaced by 0.

(2) Before processing next frame image, calculate the time interval Δt with former frame image and put it into Equation (17). Equation (18) is used to predict current moving state \hat{x}'_k . In \hat{x}'_k , detection in certain area $(\hat{x}'_k s_k, \hat{y}'_k s_k)$ is the center and search the optimal matching of basketball mode. If no result is available for full image matching, substitute (xw_k, yw_k) into state correction equation to get $(\hat{x}_k s_k, \hat{y}_k s_k)$.

(3) Update state of Kalman filter with $y_k = (xw_k, yw_k)^T$ and update speed with Equation (20) and Equation (21).

$$vx_k = \frac{xs_k - xs_{k-1}}{\Delta t} \quad (20)$$

$$vy_k = \frac{ys_k - ys_{k-1}}{\Delta t} \quad (21)$$

After Kalman prediction, it's required to find true position of basketball around the predicted point. We use histogram matching method to guarantee tracking timeliness. We use diamond searching method for searching; the predicted point is the center and the region, which is five times bigger than the basketball template as searching area. This replaces the original full image searching. Make template size $M \times N$, the template histogram after normalization is q_u , size of matching image is $M' \times N'$, and the normalized histogram of candidate matching part of size $M \times N$ is $p_u(x, y)$. Similarity Equation (22) is defined as:

$$\rho(x, y) = \sum_{u=0}^{255} \sqrt{p_u(x, y)q_u} \quad (22)$$

In basketball match videos, when the ball is sheltered by player and flies out of playground while being held by player, it's difficult to accurately detect the ball. At the moment, accurate observation value can't be acquired and Kalman filter can't update; thus, Kalman filter can't track precisely. In this case, we can decide if basketball flies off the court by virtue of information like detected basketball position and speed. If the ball doesn't fly away, we can tell if it's occluded or carried as per undetected basketball video frame. According to features of basketball match, if the frame is few, it's thought that the ball is occluded; if frame is big and uninterrupted, it's thought that the ball is carried. If the ball is occluded by player, we can do curve fitting with least square method to predict the position of basketball in next frame. This can be used as observation value to update Kalman filter until the ball is detected. If the ball is held by a player, we can track player who is closest to predicted point of basketball trajectory function.

4. Experimental Analysis and Results

4.1. Environment of Simulation Experiment

Hardware environment is Inter(R) Pentium®3.0GHZCPU, 2GM memory. Software environment is WINXP OS, Visual c++6.0, Visual c#2008, OPenCV, Aforge.net.

4.2. Experimental Results and Analysis of Moving Target Detection

In this part, we utilize adaptive Gaussian mixed model for the court pixel modeling. Through the distance between observed value and playground pixel, we can tell observation point belongs to background point or target point. The data for the experiment is chosen from NBA basketball match videos.

4.2.1. Experimental Results and Analysis of Player Tracking

The experimental video is selected from match between Lakers and Celtics. For testing, we choose upper body of one player in Lakers (white) as tracking objective. It is shown in Figure 1.

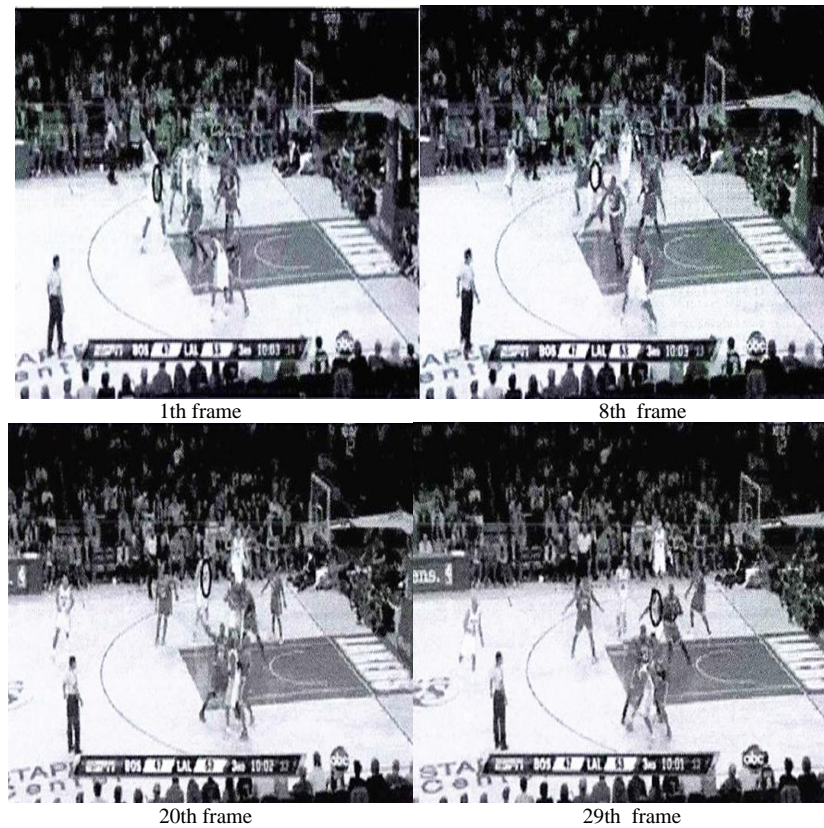


Figure 1. The tracing results of Camshift

From Figure 1, we note that in 20th frame, due to occlusion, Camshift tracked upper part of another player. In the 29th frame when there's no occlusion, the target has certain variations; but, Camshift tracks correct target. It shows that Camshift has good effect in player tracking.

4.2.2. Experimental Results and Analysis of Basketball Tracking

The experimental video is chosen from a match between the Lakers and Celtics. In this paper, we use Kalman filter to track basketball. It is shown in Figure 2. We find in the 50th frame, the basketball changed direction and speed as a result of external force. Therefore, big gap exists between observation value and true value. However, with basketball moving steadily towards new direction at new velocity, Kalman filter tracks it very well. This is seen in the 51th frame. It suggests that Kalman filter performs better in tracking basketball.

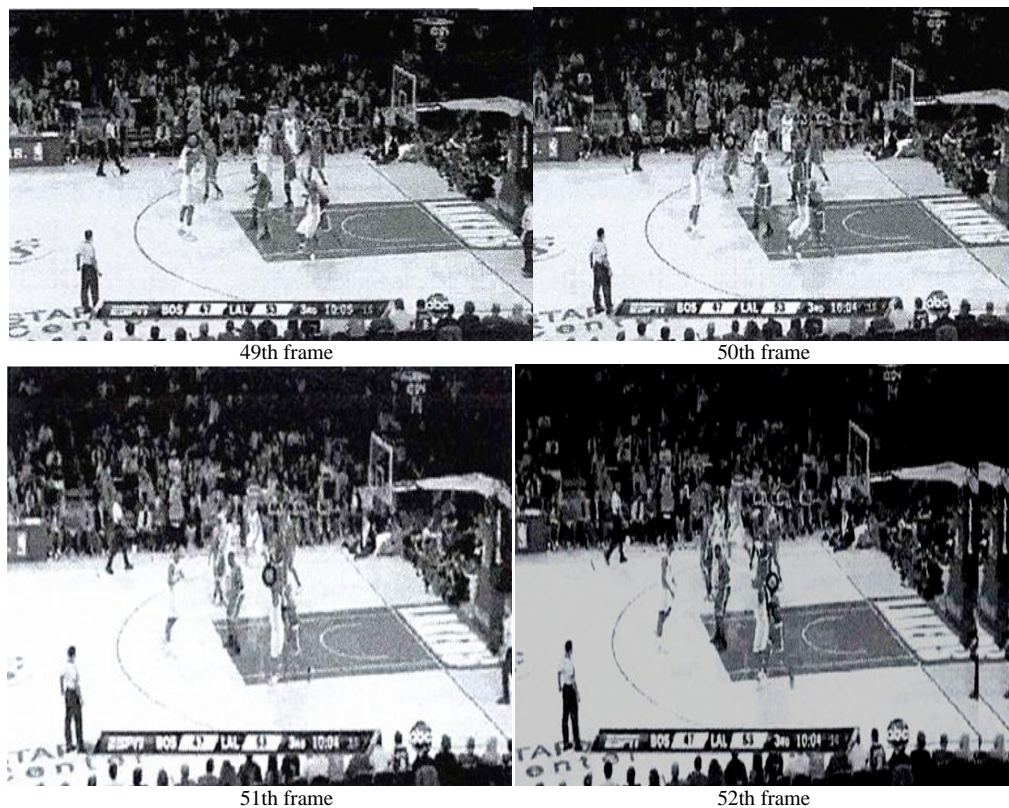


Figure 2. The tracing results based on Kalman filter

5. Conclusion

In this paper, we used adaptive Gaussian mixed model to model playground pixel, dividing foreground point and background point. Next, with target recognition method based on H invariant moment, we completed recognition of players and basketball, as well as detection of moving targets like players and basketball. For target tracking, considering features of basketball match videos, we applied Camshift algorithm to track players and the one based on Kalman filter to track basketball. Using adaptive Gauss mixture model to model the stadium pixels, in the video will be the foreground and background points to distinguish. In the target tracking, combined with the characteristics of basketball video, using the algorithm to track the players, the use of the filter to track the basketball, and achieved good result. Experiments reveal that the proposed method makes better effect in both tracking and detection.

References

1. Y. F. Peng, "Research on Moving Target Detection and Tracking Algorithms," Wuhan University of Technology, 2010
2. B. Sun, "Mobile Background under the Research on Moving Object Detection and Tracking Technology," *Electronic Measurement and Instrument*, Vol. 3, pp. 206-210, 2014
3. P. H. Tian, "Research on Moving Object Detection and Tracking in Video Image," Chang'an University, 2013
4. Y. H. Wang, "Research on Moving Target Detection and Tracking Algorithm based on OpenCV," Hangzhou Dianzi University,

2016

5. X. H. Lu and Z. K. Shi, "Based on Dynamic Template Matching of Air Moving Target Detection and Tracking Control," *Electronic Measurement and Instrument*, Vol. 10, pp. 935-941, 2014
6. B. H. Yuan and D. X. Zhang, "Detection and Tracking of Video Moving Objects based on OpenCV," *Application of Computer System*, Vol. 5, pp. 90-93, 2014
7. W. Yang and X. F. Bai, "An Improved Moving Object Detection and Tracking Method," *TV Technology*, Vol. 1, pp. 180-182, 2016
8. M. Y. Liu, "Research of Moving Object Detection and Tracking Algorithm," *Journal of Hebei United University (Natural Science Edition)*, Vol. 1, pp. 65-70, 2015
9. X. F. Tong and Q. S. Liu, "Analysis of Sports Video," *Journal of Computer Science*, Vol. 7, pp. 1242-1251, 2008
10. S. W. Hai and G. J. Li, "Player Detection and Tracking in Sports Video," *Computer Engineering*, Vol. 19, pp. 230-232, 2008
11. H. I. Xu, "Sports Video Sequence of Moving Object based on IMM Tracking Algorithm," *Journal of Image and Graphics*, Vol. 5, pp. 920-924, 2009
12. Y. Jiang, "Research on the Motion Video Tracking Technology based on Meanshift Algorithm and Color Histogram Algorithm," *Journal of Soochow University (Engineering Science Edition)*, Vol. 2, pp. 33-36, 2012
13. P. Qu and T. Kang, "Based on Underlying Visual Information of Sports Video Intelligent Analysis," *Journal of Sports Adult Education*, Vol. 3, pp. 49-51, 2012
14. Y. Xiu, "Research on Moving Target Tracking and Detection Algorithm in Volleyball Video," *Science and Technology Bulletin*, Vol. 7, pp. 160-162, 2015

Baojun Zhang received his M.S degree from school of physical education, Northeast Normal University. He is an associate professor in Harbin Institute of Technology. His research interests include sports teaching and training.