Accepted Manuscript

Robust player detection and tracking in broadcast soccer video based on enhanced particle filter

Ying Yang, Danyang Li

PII: S1047-3203(17)30064-0
DOI: http://dx.doi.org/10.1016/j.jvcir.2017.03.008
Reference: YJVCI 1978

To appear in: *J. Vis. Commun. Image R.*

Received Date: 26 September 2016
Revised Date: 20 February 2017
Accepted Date: 3 March 2017

Please cite this article as: Y. Yang, D. Li, Robust player detection and tracking in broadcast soccer video based on enhanced particle filter, *J. Vis. Commun. Image R.* (2017), doi: http://dx.doi.org/10.1016/j.jvcir.2017.03.008

Robust player detection and tracking in broadcast soccer video based on enhanced particle filter

Ying Yang[a], Danyang Li[b]

[a] College of Information and Electrical Engineering, China Agricultural University, Beijing, 100083, China (Corresponding Author, E-mail: hbxtyy@126.com; Address: No. 27 Qinghua East Road, Haidian District Beijing, China; Postcode: 100083; Tel: (8610)62600655)

[b] College of Information and Electrical Engineering, China Agricultural University, Beijing, 100083, China (E-mail: 276437131@qq.com; Address: No. 27 Qinghua East Road, Haidian District Beijing, China; Postcode: 100083; Tel: 18811703255)

ABSTRACT: It is significant to detect and track soccer players in broadcast sports video, which is helpful to analysis player activity and team tactics. However, it is challenging to efficiently detect and track soccer players with shots switched and noise caused by auditorium and billboards. And for multi-player tracking how to treat the increase or decrease of player are also difficult. In this paper, a robust player detection algorithm based on salient region detection and tracking based on enhanced particle filtering are proposed. Salient region detection is used to segment sports fields, and then soccer players are detected by edge detection combined with Otsu algorithm. For soccer players tracking, we use an enhanced particle filter which we improve the algorithm in sample and the likelihood function combing the color feature and edge feature. Experimental results show the proposed algorithm can quickly and accurately detect and track soccer players in broadcast video.

Keywords: object tracking, particle filter, salient region detection, Otsu algorithm

1 INTRODUCTION

Player detection and tracking in broadcast soccer video plays an important role in multimedia analysis [39, 40, 42, 44, 45], which contributes to players observation and analysis, and player tracking can help coach observes each player in frames [22, 27, 28, 29, 31]. And a variety of applications are based on the algorithm such as player running diagram, give a good account of match containing possession, shots, pass combinations and so on. Object detection and tracking is a hot research topic in computer vision research field, which is widely applied in broadcast sports video field [25, 30, 32, 36, 38].

In soccer player tracking, tracking resorts to results of player detection, and it is important to determine the position of each player. In related work, some researchers detect players in frames of broadcast sports video [3, 21, 23, 24, 26, 43]. Heydari et al. proposed the player detection algorithm that uses k-means clustering by clustering the players in similar color [1]. Liu et al. detected players by a boosted algorithm, and trained classifier by cascade of haar features for different training datasets [5].The detection algorithms based on cascade classifier segment players directly in frames. In the method, it is easy to be affected by complex background that noise from auditorium and shot switching. Some researches detect players by feature attributes such as SIFT, color, and edges in sports field. Min et al. proposed the football players detection and tracking algorithm based on SIFT [2], and the players position based on the scale-invariant feature detection, which is accurate and efficient. Zhang et al. use canny operator and Morphology detect the field in soccer game [4], and segment players by the threshold. However, soccer player detection algorithm may result in error when there are many lines in the frames. Although there are many approaches for player detection, in the broadcast sports video, it is so much noise that influences player detection such as lines in sports field, Goalmouth with many grids, audience on the grandstand.

Many tracking algorithms based on particle filter have been presented [9-13, 33, 34, 35], which are adaptive to nonlinear and non-Gauss dynamic systems. Yu et al. proposed an improved particle filter algorithm, which proposed the hybrid proposal distribution of adaptive optimization, considering current observed information to optimize sample distribution in the particle filter. The algorithm adaptively generates annealing parameters in hybrid proposal [6]. However, with particles propagation and sample in particle filter, a few particles weight increase and the weight of most particles is close to 0, the weight of particles become concentrated, the effectiveness and diversity of particles become degradation. Based on the shortcoming, many researchers enhanced particle filter algorithm to improve particle degradation. Li et al. devoted to optimizing diversity of particles in resample, proposed the minimum sampling variance [7]. During update weight of particles, Zhang et al. proposed average likelihood functions by diverse proportion, using multi-observations to calculate the similarity [8, 41]. Besides, researchers have been searching methods in object tracking algorithm, in particle filter framework based on sparse representation [14-16], target template is represented as candidate particles with a linear combination, and candidate particles' weights are computed basis of the similarity. The method can effectively decrease noise and improve tracking accuracy. These tracking algorithms based on particle filter are lack of analysis of particle numbers. Due to different targets occlusion in tracking, it is inappropriate to track players with fixed number particles. In addition, likelihood functions depend only upon single feature, and the weight computed by likelihood functions presents error, which may result in loss in tracking.

In this paper, we propose a detection and tracking algorithm based on enhanced particle filter in broadcast sports video. As we can see from frames in broadcast soccer video, there are auditorium and many advertisement boards, which influence player detection and tracking. In figure 1 the sports field segmentation module, we shield off the noise by segmenting sports fields, and salient region detection is used to segment sports fields in sports video frames. Another noise is the lines on the sports field, the top-hat algorithm is used to reduce the noise. In broadcast sports video, it is critical for players detection to determine player positions, and sampling particles from the latest detection information. In figure 1 soccer player detection module, the detection approach combines Otsu Algorithm with edge detection method (EOD), which makes good use of advantages of the two methods. Player detection in sports field improves the efficiency of tracking in target position determining and particle sampling. It is robust to track targets with dynamic number in broadcast sports video. In figure 1 module soccer players tracking, we proposed a tracking algorithm based on enhanced particle filter, we sample particles by detection and particles movement (DPF). Therefore, increasing particle diversity may lead to track accurately. We improve the likelihood functions combining the color feature and edge feature, which compute particle weight in player tracking.

The contributions of this work are three aspects. First, we proposed a segmentation algorithm based on salient region detection. Segmenting sports field in the frame is efficient to avoid noise off sports field. Second, a soccer player detection algorithm based on fusion of Otsu Algorithm and edge detection method is proposed, which contributes to tracking in sample particles and improve the diversity of particles. Third, a soccer player tracking algorithm based on enhanced particle filter is proposed. Sampling particles added latest observation improves the tracking efficiency and accuracy. Besides, we use the likelihood functions by combining the color feature and edge feature, which is robust to the illumination change.

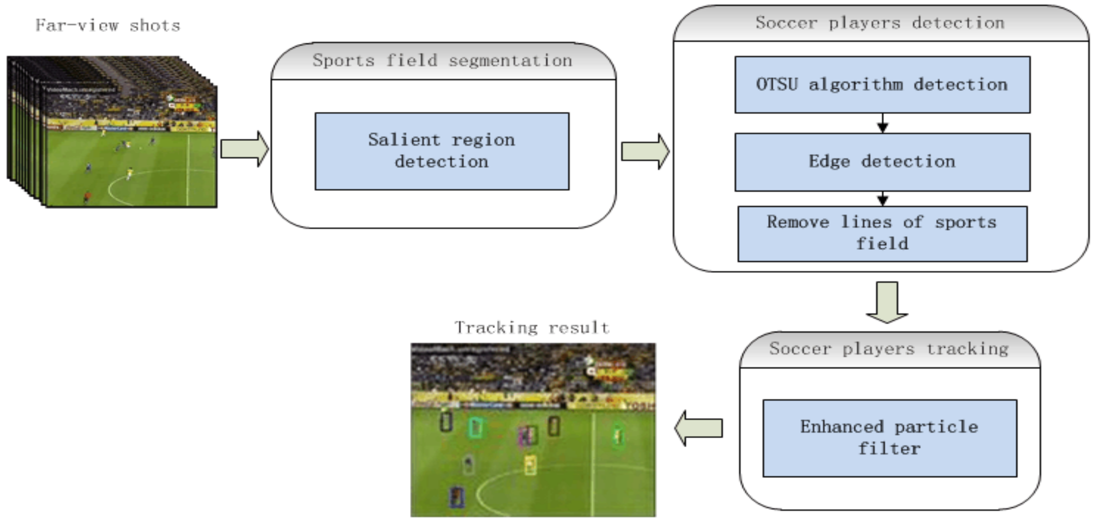


Figure 1 the whole framework

2 THE SPORT FIELDS SEGMENTATION BASED ON SALIENT REGION DETECTION

In broadcast sports video, there are many shots only containing auditorium or noise, which doesn't contain players we tracking. As we can see, the frames containing tracking targets mainly consist of sports field, in which the main hue is green. So we select the frames from videos by the color statistics. That is to say, in the frame containing sports field, the pixels number in green are large than other color. As we can see from the figure 2(a), the frame doesn't contain sports field. As a result, we delete the frames because the frame not exist the targets to track. And in figure 2(b), we can track soccer players in the sports field.

(a) The Audience off the field **(b) the players on the sports field**

Figure 2 the frames of the sports field and off the sports field

In the frames containing targets, there are noises coming from auditorium, many advertisement boards and auditorium, which may cause problem for player detection. Soccer players are in sports field, separating the sports field from the audience's stand is critical to detect and track soccer players. In previous segmentation algorithm [19], they segment sports field based on dominant color [2, 4].However, when the hue value in the auditorium is near to the hue value in the fields, it is difficult for the Dominant Hue Method to segment the sports fields from the auditorium.

Motivated by a global contrast based salient region detection proposed by Cheng [17], we proposed an innovative approach to separate sports field from its surroundings based on the histogram contrast method. First, we compute saliency value of each pixel in the frames, and segment the saliency map (HC map, figure 3 (b)) with the pixels in different saliency values. Due to the saliency map, we segment it from its surroundings.

In the frame, we compute the pixel saliency value by the color in pixel. By contrasting other color in the L*a*b space, we can compute the color saliency value.

$$\mathrm{D} = \sqrt{\Delta L^2 + \Delta a^2 + \Delta b^2} \qquad (1)$$

The two color distance D is computed by the components in L, a, b.ΔL、Δa、Δb is the distance of the two color in the L、A、B components。And color c in the image are denoted as $c_1, c_2 \cdots c_k, \cdots c_N$ and $\mathrm{s}(c_k)$is the salient value of color c_k. Thus

$$\mathrm{S}(\mathrm{c_k}) = \mathrm{D}(c_k, c_1) + \mathrm{D}(c_k, c_2) + \cdots + \mathrm{D}(c_k, c_N) \qquad (2)$$

Where $\mathrm{S}(\mathrm{c_k})$ is saliency value of the color k, n is the number of colors in the image, and $\mathrm{D}(\mathrm{c_l}, \mathrm{c_i})$ is the color distance between color $\mathrm{c_l}$ and $\mathrm{c_i}$ in the L*a*b space. Besides color distance influences the saliency value, the pixel number is an influence factor. To compute the probability of pixel color, we use pixels statistics of the input image.

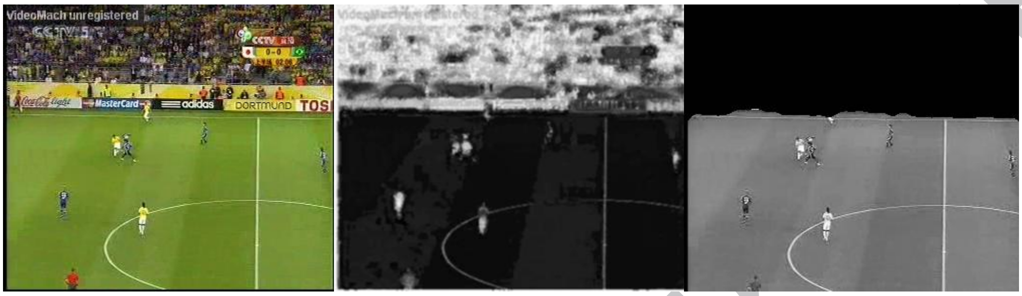
$$f_i = \frac{n_i}{N} \qquad (3)$$

Where N is the number of pixels in the image, $n_i$is pixels number in color i. Based on the analysis above, we compute the saliency value of each pixel based on the color.

$$\mathrm{S}(\mathrm{p_l}) = \mathrm{S}(\mathrm{c_k}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{f_i} \mathrm{D}(\mathrm{c_k}, \mathrm{c_i}) \qquad (4)$$

Where c_k is the color of pixel p_l, $S(c_k)$ is the saliency value of the color k, $S(p_l)$ is the saliency value of the pixel in the color, n is the number of colors in the input image, f_i is the probability of pixel color, and $D(c_l, c_i)$ is the color distance between color c_l and c_i in the L*a*b space.

As we can see from the figure 3, we first get salient maps by computing the saliency value in pixel across the images, such as figure 3(b). According to the salient map, we segment salient region from its surrounding, as show in figure 3(c). That is to say, we segment sports fields according to saliency value for detecting players region. In the section 5.1, we contrast segmentation result of the Dominant Hue Method with this algorithm.



a) original image **b)HC map** **c)segmentation result**

Figure 3 Segmentation result by salient region detection

3 Player Detection

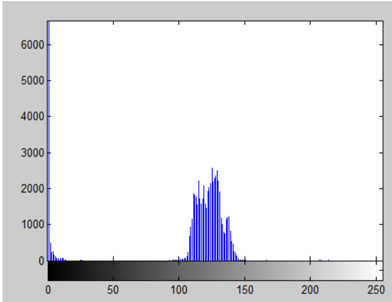
In our previous work, we get sports field segmentation according to the salient region detection, and next we need to detect players in sports field. This part is aimed at automatic soccer player detection in sports field and recording soccer player position with rectangular frame. Due to the detection result, it is important to improve the tracking accuracy by improving the sampling particles. In Otsu algorithm, it can separate players from background by threshold value. However, it may lose players in detection. Besides, soccer player detection by edge detection also exits error. In this paper, we propose a detection method fused Otsu algorithm and edge detection method, which can accurately detect soccer players, and decrease object detection error rate.

3.1 Player Detection Based on Otsu Algorithm

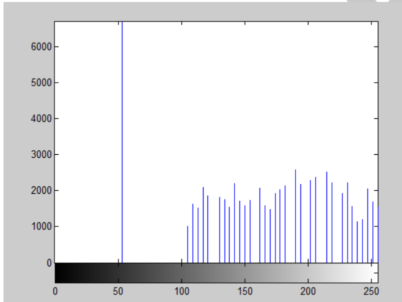
In this part, we devote to detecting players in sports fields. We adopt Otsu algorithm to detect players by different gray value between the sports fields and soccer players. In some frames, soccer player gray value is similar to value in sports fields in segmentation result. To detect players in sports fields, we adopt gray scale transform to enhance the contrast of image. Obviously, soccer players are divided into two teams in a soccer match, so they are distinguished by their clothing. To improve the contrast of the players and sports field background, we transform the gray scale with a linear function, as shown in figure 4. Firstly, from the figure 4(a), map the image of the sports field with 255 pixels into [0 1], and transform the gray value range of the image, as shown in figure 4(c).

$$x_f = \begin{cases} 255 \times x_1 & x < x_1 \\ \frac{y_2 - y_1}{x_2 - x_1} \times (x - x_1) \times 255 + 255 \times y_1 & x_1 < x < x_2 \\ 255 \times x_2 & x > x_2 \end{cases} \quad (5)$$

X is a value gray of the pixel in the image of the broadcast sports video between [0 1], x_f is the output of the gray scale transform, x_1 and x_2 are the input value to be linear change, and y_1 and y_2 are mapping interval of the gray scale transform.



(a) segmentation result (b) the segmentation histogram



(c)gray change of the segmentation result (d)the histogram of the gray change

Figure 4 Gray change of the image

And we transform the image into two categories, stretch the gray scale change range [0, 0.5] into [0 1], in which the gray value of the image is in low gray value (low gray value is the gray value lower than 125), and stretch the gray scale change range [0.6, 0.9] into [0 1], in which the gray value of the image is in high gray value (high gray value is the gray value higher than 125). According to the transform, we get two categories gray images, which enhance the contrast of the background and the objects.

(a) segmentation result **(b)the gray scale in[0,0.5]** **(c)the gray scale in[0.5,1]**

Figure 5 Gray scale change of the image

We enhance the soccer player and its surrounding contrast by the gray scale transform, and divide the image into object and background based on the Otsu algorithm [18]. Otsu algorithm computes the threshold value of two valued image segmentation, proposed by Japanese scholars Otsu in 1979. Due to this method, an adaptive threshold is computed and the biggest variance is between the foreground and the background image.

Assume that the number of gray level in the image is m, the number of pixels is n_i, where gray value is i, the number of all pixels are N, so the mean u in global gray value:

$$u = \sum_{i=1}^{m} \frac{n_i}{N} * i \tag{6}$$

The threshold value k make the image into binary image with background and players, $C_0 = [1 \cdots k]$and $C_1 = [k+1 \cdots m]$, the probability of the two parts respectively are $\omega_0$and ω_1, the average of gray value of the two parts respectively are $\mu_0$and μ_1,so gray value average of the two parts:

$$\mu = \omega_0\mu_0 + \omega_1\mu_1 \quad (7)$$

Variance between the two parts:

$$d(k) = \omega_0(\mu_0 - u)^2 + \omega_1(\mu - u)^2 \quad (8)$$

Compute the threshold k, which make the variance between the two parts is the largest. Based on the Otsu algorithm, we divide the image into pixels in black and white. Thus the players in sports field can be detected, which the gray value equal 1, the gray value opposite to the background is 0.

Soccer player map in high gray value contain some lines of the sports field, which cause some interference for player detection. To decrease the lines noise, top-hat method is used to decrease the effects of noise and uneven illumination [37]. Top-hat transform is applied to the detection of the white edge, and line in sports field is just white, so top-hat transform is easy to extract lines. It is often used to detect the gray level of the image signal that is smaller than the structural element, and low contrast regions can be extracted from the less obvious object. From figure 6(c) we can see, it is efficient to detect the lines in the frames. On the basis of the top-hat algorithm, it can quickly and precisely detect the lines on the frames. Compared with the top-hat detection, the Hough Transform limits the standard geometric lines, such as ellipse and lines. With the irregular lines in the frames, it is efficient for the top-hat algorithm to detect the lines in the frames.

$$H = A - (A°B) \quad (9)$$

A is the sports fields of segmentation, structural element B is a small sliding window, (In order to efficiently detect the pixels of lines, the size of the structural element B is 3×3). Make open operation to A, and dilation and erosion operation are used to B in the algorithm. And the result of the top-hat algorithm is the subtraction between original image A and open operation result of A and B.

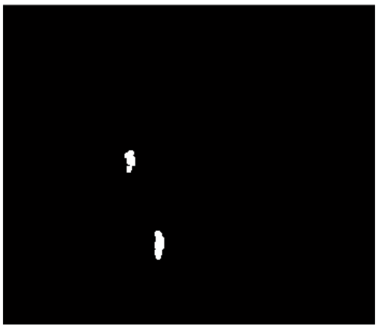
a) segmentation result **b) Open operation diagram** **c) the lines of the sports field**

Figure 6 remove the lines with the top-hat algorithm

As we can see in figure 7(a), it is the original image we truncate from the video clips, and in the frames, we detect the players to determine the position of soccer players. First, based on the transform of gray-level, the gray image separately stretches the image in gray value in low and high, which contributes to distinguishing the background and objects. After stretching the gray image, we separately detect the players with the OTSU algorithm. And as shown in figure 7(b), figure 7(c), soccer players in different gray value are detected in different change of gray-level. Finally, soccer players are the superposition of the detection results, shown as figure 7(d).

(a) original image

(b)the detection result of the players in low gray value



(c) the detection result of the gray value in high value

(d) the detection result of the image

Figure 7 soccer players detection result based on OTSU algorithm

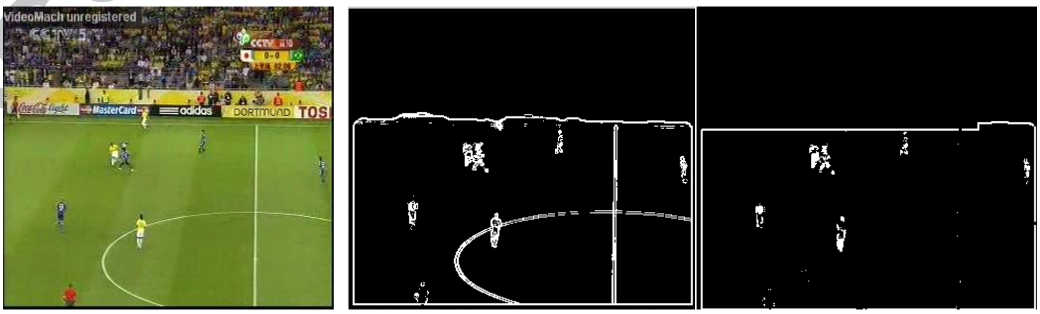
3.2 SOCCER PLAYER DETECTION BASED ON EDGE DETECTION

In broadcast sports video, when soccer players gray value is similar to the sports field, the Otsu algorithm detection may lose the effectiveness. To increase the accuracy rate of soccer players detection, we use the edge detection detect the players once again, as shown in figure 8(b).

We define a first order differential operator for the edge detection, and each of the pixels in the image is separately convolution operation with the operator.

$$\begin{bmatrix} -1 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

And convolution operation compare with the threshold value, if the result of convolution operation is bigger than threshold value than mark the pixels as edge points.



(a) original image (b)the result of the edge detection (c) edge detection result remove the lines

Figure8 soccer player detection with edge detection algorithm

After the edge detection, as shown in figure 8(b), there are some lines in the image. So we continue to use the top-hat algorithm to eliminate lines noise, figure 8(c). From the previous work, we get the detection result of edge detection and Otsu algorithm detection. And the detection result is the superposition result of the two detection result. Detection result with the two methods is coincident, otherwise, the detection result is loss error in one method. Integrated edge detection and the OTSU algorithm, we can detect the players on the sports fields.

According to the method of Otsu algorithm, soccer player detection based simply on gray value separate the soccer players and sports fields. While the gray value of soccer players and sports fields are closed, detection in this way fails to separate them. In edge detection method, edge detection template is used to detect the region in which luminance value are sudden change across sports field, this may cause amplify noise such as the lines on sports fields. And combine the two method, we can more accurately detect the players.

4 PLAYER TRACKING

As mentioned above, soccer player positions are detected in each frame, which is helpful to soccer player tracking. In this part, we track soccer player in sports field based on the detection. With soccer players' position, we can initialize the target for tracking by detection results. Determining the target for tracking avoids the manual labeling, and when the target numbers changed such as disappear or appear, it is efficiency to dynamic tracking the new targets. Besides, we add the latest observation information to the sample particles, which makes use of detection result to improve the efficiency of tracking.

In particle filter, predicted particles sample from prior distribution and detection. We sample particles in the new frame by the targets in current frame. According to brown movement, the state space model of dynamic system can be described as

$$x_{t+1} = 2 \times x_t - x_{t-1} + N_t \tag{10}$$

$$\mathrm{z_t} = \mathrm{h(x_t)} + \mathrm{v_t} \tag{11}$$

Where $\mathrm{z_t}$ stands for the observation value at time t, $\mathrm{h(x_t)}$is the observation function. Where $\mathrm{x_t}$ describes object state estimation at time t, $\mathrm{z_t}$ is the observation at time t. $\mathrm{N_t}$ is the noise.

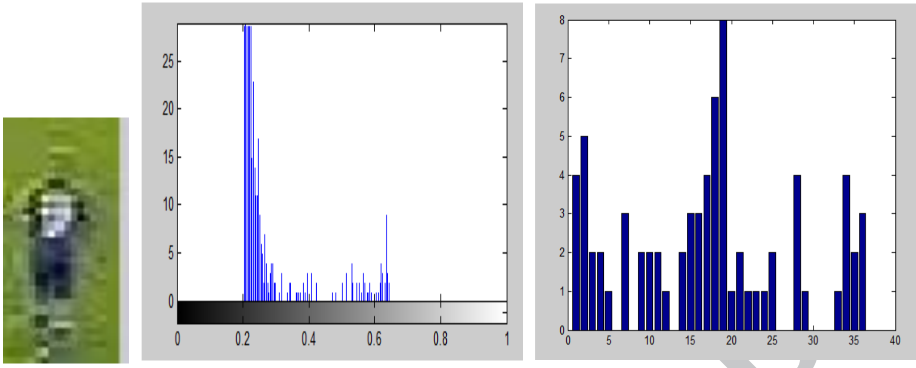
(a)frame in broadcast sports video **(b) detection result of soccer players** **(c)label players position**

Figure 9 soccer players detection

To be better estimate targets state, and sample particles are more diverse, we use the mixture distribution composed of observation information and prior distribution. Therefore, we sample from the mixed distribution composed by transition distribution and observations. From figure 9(b), we can see the results of detection, and soccer players are labeled with rectangle shown as figure 9(c). From figure 9(c) we can see, players close to the sports field can be detected well, but outside the stadium area players such as billboards can't be detected clear.

After sampling particles from frames, we predict each target by some predicted particles. However, the target numbers in continuous frames are uncertain. By computing the distance and overlap between particles and detections, we estimate whether new targets appear. If no overlap with all particles of all targets, then create new target. As mentioned above, it is robust to track in the continuous frames with uncertain numbers, which make full use of detection and sampling particles. According to Monte Carlo method, we predict targets by the weighted particles. And the weight of each particle is computed by contrast the similarity between the particles and targets.

To compute the similarity between targets and particles, we extracted the color feature and edge feature in the image. As we know, color feature is invariant to rotation and deformation and insensitive to partial occlusion, but it is susceptible to illumination and color change. The edge features have low dependence on illumination and color change, which make up the deficiency of color features, but are sensitive to the rotation and denaturation of target. Therefore, comprehensive considering color and lighting effects, we choose the likelihood composed by the edge feature and the color feature in Hue value, in HSV space, the hue value changes obviously.



a)players b) hue histogram c)orientation histogram

Figure 10 the histogram of players

For example, we random choose a player in a frame of broadcast sports video, and figure 10(b) is the hue histogram of the image, and figure 10(c) is the orientation histogram of the image. The hue histogram is the statistical hue value information about the player. Horizontal coordinate stands for bins, we divided hue value of the image into 256 bins from [0, 1], and vertical coordinate is the number pixels in the bins. And the orientation histogram describes edge feature of image. The edge points were computed by edge detection using canny operator. And gradient magnitude and orientation of edge points computed by:

$$E(x_i) = \sqrt{E_x^2(x_i) + E_y^2(x_i)} \quad (12)$$

$$\theta(x_i) = \arctan\left(\frac{E_y(x_i)}{E_x(x_i)}\right) \quad (13)$$

$\theta \in [0, 360^0]$;x_i is the coordinate vector of the pixel. $E_x(x_i)$ and $E_y(x_i)$ are the sobel operator convolved with the image to obtain the horizontal and vertical differential images. According to the paper [20], quantify the orientation of the edge points, we adopt the number of the quantization level is 36, and the quantization interval is 10.

And we compute the similarity between particles and targets by Bhattacharyya distance. The similarity is denoted as S, which compute by the edge feature and the hue value.

$$p_1(y) = C\sum_{i=1}^{n} k(\left||x_i - y|\right|^2)\delta(b(x_i) - u) \quad (14)$$

$$B(p_1, p_2) = \sum_{i=1}^{n} \sqrt{p_1^i p_2^i} \quad (15)$$

$$S = B_h * \alpha + B_e(1 - \alpha) \quad (16)$$

Where $B(p_1, p_2)$is the Bhattacharyya coefficient of discrete probability distribution, where p_1^i and p_2^i are the features of the particle and observation, and n is the number of the quantization level as 36, y is the central location of the target, C is normalization constant, δ is the Kronecker delta function, k is the kernel function, $b(x_i)$ is the feature value of the pixel x_i, and u is the value of the feature index. α is the coefficient of the two features, and experiment shows when the α between [0.3,0.4], the tracking result is better.

Therefore, targets estimated by the weighted sample particles. The particles denotes as $\{x_{t1}, x_{t2}, x_{t3} \dots x_{tn}\}$, where x_t stands for the target state at time t, and x_{t1} stands for the predict particle around the target x_t. And to estimate accuracy, we select particles with high weight, which is more similar to the target, and delete particles with low weight. This may result in diversity loss in particle, all sample particles originate from the same particle. We judge the degeneracy by the variance of particles V_t, if V_t higher than the threshold, the particles is degeneration, we resample particles from the detection and particles. To increase diversity of particles, we sample from detection results and targets. With the mixed distribution, it is efficient to increase the diversity of the sample particles. From figure 11 we can see, the variance of the sample particles. And the experiment shows that the particles in each frame are diversity. And horizontal coordinate stands for each frame in broadcast sports video. Vertical coordinate stands for the variance between sample particles in each target. From the figure 11, we can see the variance is acceptable for the particles, and the particles are diversity.

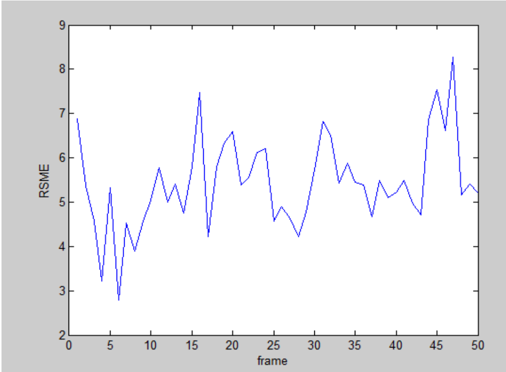


Figure 11 the variance of the sample particles

Based on the algorithm computing sample particles number [34, 36], we propose the adaptively adjusting sample algorithm to improve the sample efficiency. According to some methods in computing the number of particles, we compute the least number to track soccer players. The number was computed by:

$$\mathrm{N} = \min\left(\frac{a}{1+\exp\,(p-0.6)\times b}, N_{min}\right) \qquad (17)$$

Where p is the similarity between detection and target, N_{min} is the least particle number (in this paper, we set the N_{min} 50), a, b are parameters to adjust the number N in reasonable interval. In this paper, through experiment comparison, we select b=3, a=60.

As mentioned above, we enhanced particle filter algorithm in our soccer player tracking algorithm. Based on the latest observation, this algorithm is robust to detect players with increase or decrease. It is efficient to estimate target with dynamic, which can fast track the soccer players. Besides, the likelihood function computed by edge feature and color feature.

5 EXPERIMENTS AND ANALYSIS

In this part, we evaluate our soccer player detection and tracking in broadcast sports video algorithm in MATLAB 2012B on a PC. For multiple targets tracking, there is no public standard dataset for us to track in the experiment. So the data we used is European Cup video clips in 2006, and the European Cup video is the video we often watch. We randomly select 200 frames from the frames in the video. In different soccer videos, original frame sizes are different. To precisely detect and track players in the same standard, in this paper, we resize the size of the image 240 x 320 x 3.

According to the framework of our algorithm, the experiment consists of three parts. In sports field segmentation part, we experiment on the frame which the hue value in the auditorium is near to the hue value in the fields. Experiment shows salient region detection is robust to segment compared to the Dominant Hue Method. In soccer player detection part, we select different teams from soccer video clips. In different frames, experiment shows the detection algorithm is adaptive to detect different team players. In soccer player tracking, experiments show the tracking is robust when the shots switch or player numbers change. Besides, we contrast the particle filter to our algorithm in different video clips. Experiment shows our algorithm improve the accurate rate.

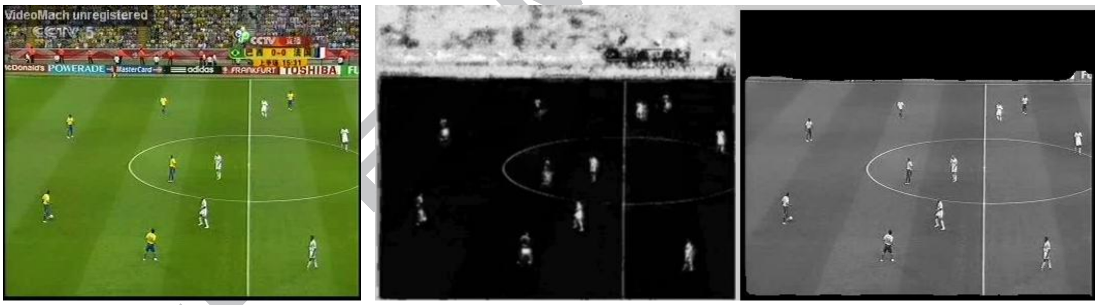
5.1 THE SPORTS FIELD SEGMENTATION

To test the validity of salient region detection, we experiment our algorithm on the frame contrasting to the Dominant Hue Method (DH) proposed Yang et al. DH method segment sports field according to the dominant hue value. Dominant hue value is pixel value in majority of pixels in frame. As we can see from figure 12(b), the DH method fails to separate the sports field in the image, as the hue value in the auditorium is near to the hue value in the fields. Therefore, when the hue value is similar, it is difficult to segment the sports fields from the auditorium in the Dominant Hue Method.

a)original image **b)DH**

Figure 12 Dominant Hue Method segment result

Our algorithm segment the sports field by salient region detection method, according to salient region detection proposed by Cheng [17]. Experiment shows salient region detection method can segment the sports field well. Figure 13 shows the experiment result.



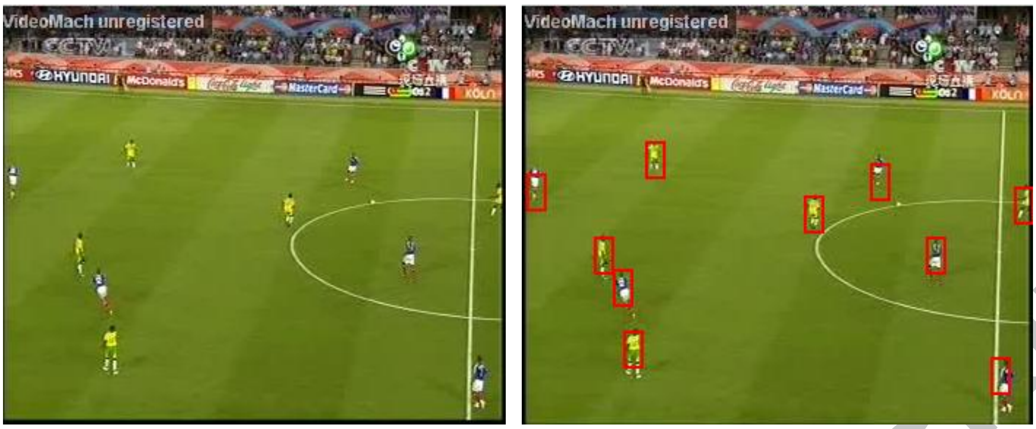
a) original image **b)salient detection** **3)segmentation**

Figure 13 segment result based on the salient region detection

5.2 THE SOCCER PLAYER DETECTION

We evaluate our detection algorithm on different teams in different video clips, Brazil VS France, Japan VS Brazil, Togo VS France, Spain VS France. We mark soccer players with a red rectangular box. Experiment shows our detection algorithm is adaptive to different video clips, and the experimental result shows our algorithm is efficient on different players, especially the players whose color is similar to the field. In figure 14, the detection method is robust to detect players in the frame, when sports field color is similar to player color.

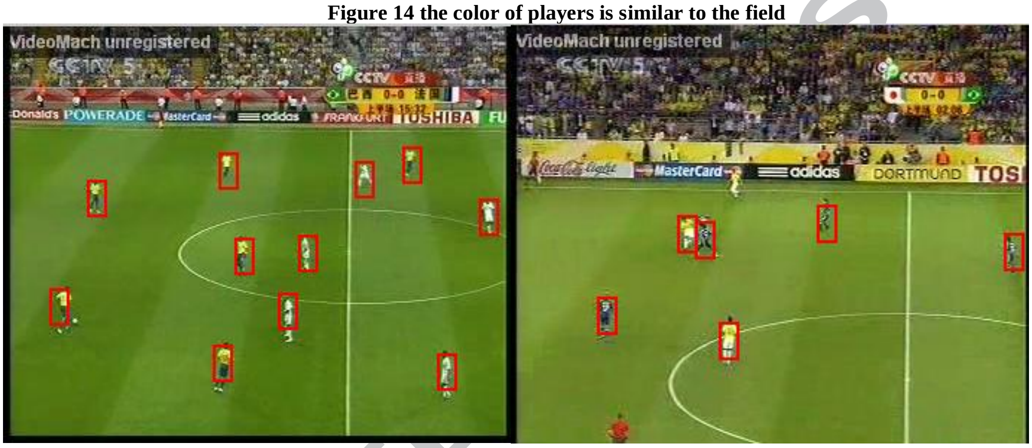
However, players near to the auditorium fail to be detected, as shown in figure 15(b). Due to players near to the auditorium or background noise around the auditorium, it is easy to regard the players as background, when segment the auditorium from sports field. So when players near to the auditorium, the detection method may fail to detect the players. Another disadvantage of the detection method shows as figure 15(d), soccer in the image is detected, the detection method regard soccer as players. The method can’t distinguish the players and soccer.



a) original image

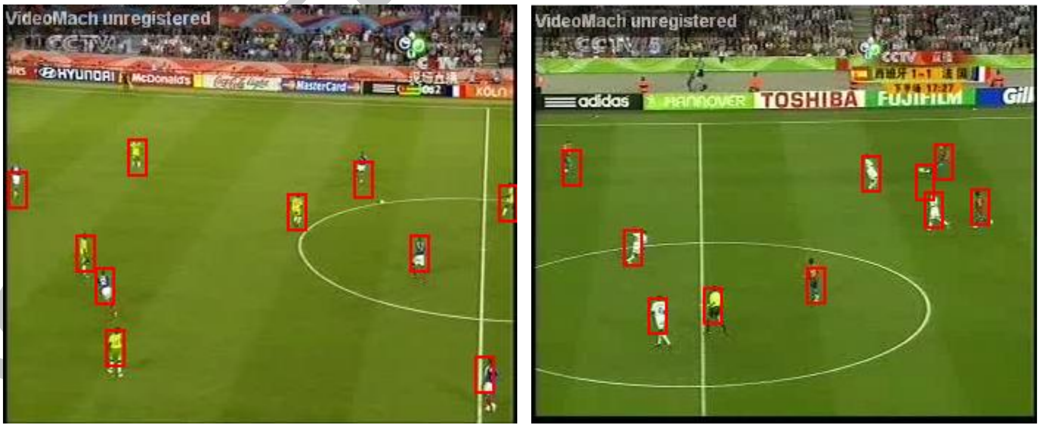
b) the detection result of the image

Figure 14 the color of players is similar to the field



a)Brazil VS France

b)Japan VS Brazil



c)Togo VS France

d) Spain VS France

Figure 15 Player detection of different game

5.3 THE SOCCER PLAYER TRACKING

In broadcast sports video, when the shots turn to the auditorium, soccer players in the frame out of sight. In figure 16 frame 74, it is difficult for particle filter to track with the targets losing in the frames. In our tracking method, we detect the frame until targets appear. When targets disappear (in the frame 74), there is no particle in the frames. In frame 78, targets appear, predicted particles on the basis of the particles in the frame 73. In the frame 78, tracking is detection result in the frame 78 combine with the tracking result in the frame 73. That is to say, the targets tracking make use of detection results and predicted particles computed by the targets before shots turn.

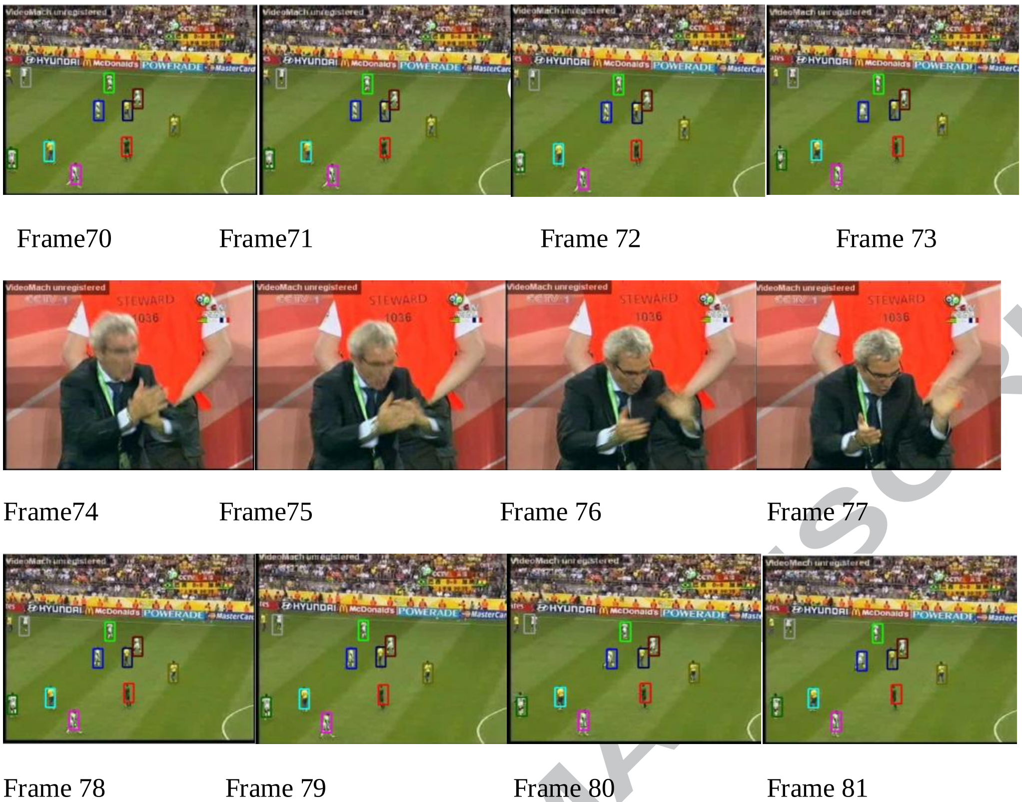


Figure 16 tracking results with the shots switch in Brazil VS Garner

In broadcast sports video, it is common to see soccer player numbers are not fixed. When a new player added, it is challenging to track the new targets. From frame 26 to frame 27 in figure 17 we can see, in the right of frame, a new player emerges. In our algorithm, new target is found by comparing the particles to detection result. When the target detection has no overlap with any particle, it means it is a new target. The experiment shows that when the frame adds new targets, the algorithm is robust to dynamically track targets.

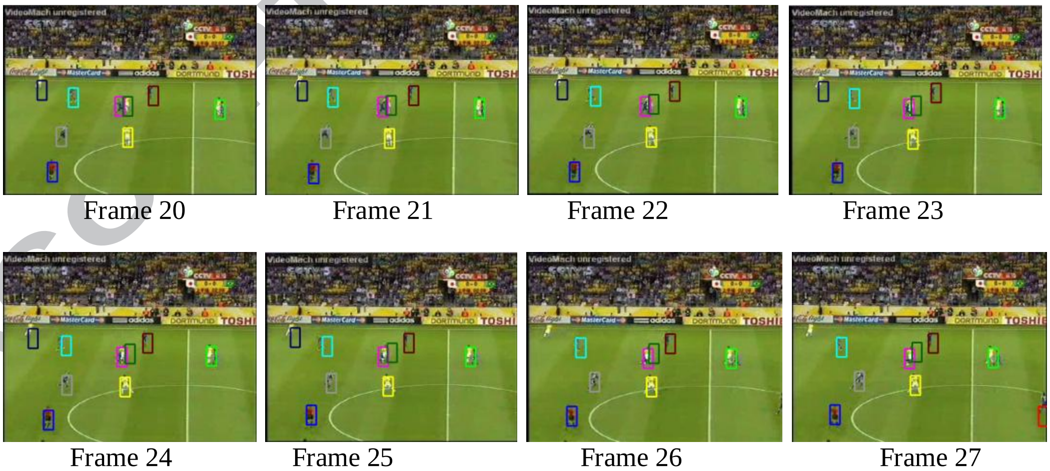
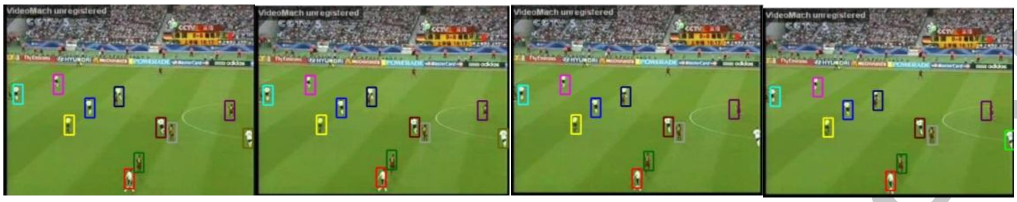


Figure 17 add new targets

From figure 18 frame 36 to 37, when players fail to track, we consider the player as a new target, and track the new target. As we can see in the frame 37, track the new targets with a green rectangular box. With detection result, the algorithm is robust to soccer players tracking.

Frame 30 Frame 31 Frame 32 Frame 33

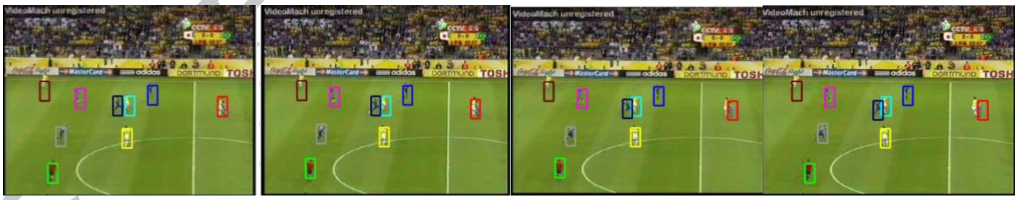


Frame 34 Frame 35 Frame 36 Frame 37

Figure 18Germany VS Portugal

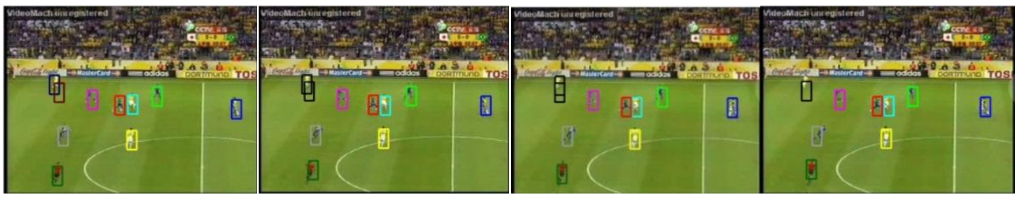
Compared to particle filter, the algorithm we proposed improves accurate and robust. We select two groups to evaluate the algorithm. The videos are in Japan VS Brazil and Brazil VS France. Contrast to figure 19 and figure 20, we can see player tracking in particle filter is not perfect when players near to the auditorium. In figure 20 frame 26 tracking result lose the target, in frame 20, there is also a tracking error. In our algorithm as shown in figure 19, the targets in the frame are on the basis of detection and predicted results. Therefore, the tracking algorithm we proposed is robust.

Frame 20 Frame 21 Frame 22 Frame 23

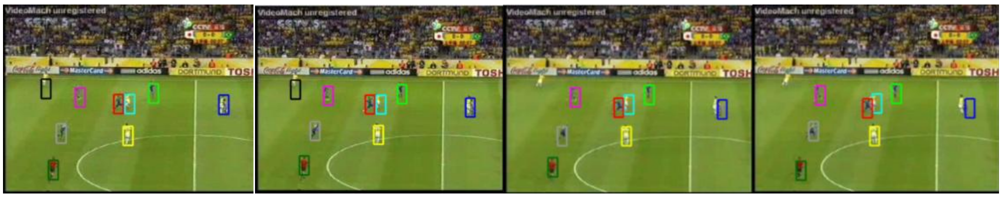


Frame 24 Frame 25 Frame 26 Frame 27

Figure 19 Japan VS Brazil (The algorithm of this paper)



Frame 20 Frame 21 Frame 22 Frame 23



Frame 24 Frame 25 Frame 26 Frame 27

Figure 20 Japan VS Brazil（particle filter）

Contrast to the figure 21 and figure 22, in particle filter, players in the frames can't track each player. That is to say, the particle filter is easy to lose tracking targets. With the detection result, the targets tracking increase the accuracy rate.

Frame 31 Frame 32 Frame 33 Frame 34

Frame 35 Frame 36 Frame 37 Frame 38

Figure 21 Brazil VS France (The algorithm of this paper)

Frame 31 Frame 32 Frame 33 Frame 34

Frame 35 Frame 36 Frame 37 Frame 38

Figure 22 Brazil VS France （particle filter）

For tracking result of the two algorithms, we compare the tracking accuracy by error number in table. First, we select 200 continuous frames in a video clips, and statistic the number of players in tracking result. Particle number is the number of particles around each target. Detection number is the number of players in detection result. Error number is the number of players in loss in tracking. in this table, we use some convenient symbols to express some words. EHF stands for the likelihood function composed by the edge feature and hue value in color feature, and HF stands for the function composed by hue value in color feature.

Table 1 tracking result comparison of two algorithms against player detection

influence factor / videos	Number of video frames	Particle number	Likelihood function	Detection number	Tracking number	Error number
Japan VS Brazil	200	30	EHF	9	9	0
Japan VS Brazil	200	30	EHF	-	8	1
Brazil VS France	200	30	EHF	9	9	1
Brazil VS France	200	30	EHF	-	8	2

Comparing the two methods, it is obvious that detection increase tracking accurate. From table 1 we can see, in broad sports video of Japan VS Brazil, Brazil VS France, with the same particle number and likelihood function, the tracking algorithm with detection decreases the rate of error.

Table 2 tracking result comparison of two algorithms against likelihood function

influence factor / videos	Number of video frames	Particle number	Likelihood function	Detection number	Tracking number	Error number
Japan VS Brazil	200	30	EHF	12	12	0
Japan VS Brazil	200	30	HF	12	12	0
Brazil VS France	200	30	EHF	12	12	1
Brazil VS France	200	30	HF	12	11	2

With the improvement in likelihood function, particles can accurate description of particles. From the table 2, contrasting to Brazil VS France, we can infer the likelihood function can improve the accuracy of tracking. Due to increase soccer player information, it is more accurate for soccer players tracking.

Table 3 tracking result comparison of two algorithms against particle number

influence factor / videos	Number of video frames	Particle number	Likelihood function	Detection number	Tracking number	Error number
Japan VS Brazil	200	30	EHF	9	9	0
Japan VS Brazil	200	25	EHF	9	9	0
Brazil VS France	200	30	EHF	9	9	1
Brazil VS France	200	25	EHF	9	9	1

In different soccer videos, player tracking result is different. In the simple background, it is efficient to track targets with few particles. From the table 3, in Japan VS Brazil, experiments show the tracking rate did not change with the less number of particles.

6 CONCLUSIONS

In this paper, we proposed a robust player detection and tracking in broadcast soccer video based on enhanced particle filter. We divide the player tracking algorithm into three parts, object segmentation, object detection and object tracking. Saliency region detection algorithm is used to sports field segment, which is efficient to avoid the noise off the field. Soccer players are detected based on edge detection algorithm combined with Otsu algorithm, fusing advantages of the two methods. Finally, enhanced particle filter algorithm is used to track soccer players, improving particle sampling in proposal distribution and improving the tracking efficiency with dynamic particle number. The shortcoming of this paper is inadequacy of identifying soccer players. When a lot of soccer players gathered in sports video, the soccer players tracking is easy to confuse. In the next step, we will at the point of view of deep learning to detect and track soccer players.

7 ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (Grant No.61202330) and the National Science and Technology Support Program (2015BAK04B01).

REFERENCES

[1] Heydari M, Moghadam A M E. An MLP-based player detection and tracking in broadcast soccer video[C]// International Conference on Robotics and Artificial Intelligence. 2012:195-199.
[2] Min, J and Z.H Meng, Football player detection and tracking based on SIFT. Information technology, 2016(05):P.195-198
[3] Yu X, Xu C, Leong H W, et al, Trajectory-Based Ball Detection and Tracking in Broadcast soccer video. Eleventh ACM International Conference on Multimedia, Berkeley, Ca, Usa, November. 2003:11--20
[4] Zhang L.F., Zhang M.J, field and athletes detection and segmentation in soccer video, National Conference on information retrieval and content security, 2004
[5] Liu J, Tong X, Li W, et al. Automatic Player Detection, Labeling and Tracking in Broadcast Soccer Video[C]// British Machine Vision Conference 2007, University of Warwick, Uk, September. DBLP, 2007.
[6] Yu J, Tang Y, Xu J, et al. An Adaptive Optimized Strategy for Particle Filter[C]// Control and Decision Conference. IEEE, 2012:3936-3940.
[7] Li T, Sun S, Sattar T P, et al. Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches [J]. Expert Systems with Applications, 2014, 41(8):3944–3954.
[8] Zhang, S. and L. Zhu, Improved Particle Filter Algorithm Based on Averaging Likelihood Functions with Diverse Proportion. Journal of Detection & Control, 2011, 33(1): p. 46-608.
[9] Yin, M., et al., Multi-cue-based CamShift guided particle filter tracking. Expert Systems with Applications, 2011, 38(5): p. 6313-6318.
[10] Huan,E.Y and L.R,Particle Filter object tracking based on adaptive feature fusion. Computer Science, 2015(02): P.316-319.
[11] Jin.T,and zhou F.G. simple and effective object tracking based on data fusion particle filter. in Infrared and Laser Engineering. 2007.
[12] Wu, S. et al. Particle filter tracking based on visual saliency feature. Journal of University of Science & Technology of China, 2015, 45(11): p. 934-942.
[13] Qi, S. and S.O. Information, Multiple target tracking using shape context features and particle filter, Application of Electronic Technique, 2015.
[14] Yang, D.W., Y. Cong and Y.D. Tang, Object Tracking Method Based on Particle Filter and Sparse Representation. Moshi Shibie Yu Rengong Zhineng, Pattern Recognition & Artificial Intelligence, 2013, 26(7): p. 680-687.
[15] Li, X., et al. Saliency Detection via Dense and Sparse Reconstruction, IEEE International Conference on Computer Vision, 2013.
[16] Lu, H., X. Jia and M.H. Yang, Visual tracking via adaptive structural local sparse appearance model, in IEEE Conference on Computer Vision and Pattern Recognition, 2012.
[17] Cheng, M., et al. Global contrast based salient region detection. in Computer Vision and Pattern Recognition. 2011.
[18] Otsu, N., A Threshold Selection Method from Gray-Level Histograms, IEEE Transactions on Systems Man & Cybernetics, 1979. 9(1): p. 62-66.
[19] Yang N C, Chang W H, Kuo C M, et al, A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval [J]. Journal of Visual Communication & Image Representation, 2008, 19(2):92-105.
[20] Zheng Y F, Ma X R, Zhao X L, et al. Mean shift target tracking algorithm based on color and edge features[J]. Guangdianzi Jiguang/journal of Optoelectronics Laser, 2011, 22(8):1231-1235.
[21] Okuma K, Taleghani A, Freitas N D, et al. A Boosted Particle Filter: Multitarget Detection and Tracking[C]// European Conference on Computer Vision. 2004:28-39.
[22] Najafzadeh N, Fotouhi M, Kasaei S. Multiple soccer players tracking[C]// International Symposium on Artificial Intelligence and Signal Processing. IEEE, 2015:310-315.

[23] Theodorou A, Panoutsakopoulos V, Paradisis G, et al. Differences in repetitive vertical jump performance parameters between professional soccer players and track and field athletes[C]// Congress of the Hellenic Society of Exercise Biochemistry and Physiology. 2016.
[24] Liu J, Tong X, Li W, et al. Automatic player detection, labeling and tracking in broadcast soccer video[J]. Pattern Recognition Letters, 2009, 30(2):103-113.
[25] Perera A G A, Srinivas C, Hoogs A, et al. Multi-Object Tracking Through Simultaneous Long Occlusions and Split-Merge Conditions[J]. IEEE CVPR, 2006, 1:666-673.
[26] Mazzeo P L, Spagnolo P, Leo M, et al. Visual Players Detection and Tracking in Soccer Matches[C]// IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance. IEEE, 2008:326-333.
[27] Tsai T Y, Lin Y Y, Liao H Y M, et al. Precise player segmentation in team sports videos using contrast-aware co-segmentation[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2016:1826-1830.
[28] Xing J, Ai H, Liu L, et al. Multiple player tracking in sports video: a dual-mode two-way Bayesian inference approach with progressive observation modeling.[J]. IEEE Transactions on Image Processing, 2011, 20(6):1652-67.
[29] Lu W L, Okuma K, Little J J. Tracking and recognizing actions of multiple hockey players using the boosted particle filter [J]. Image & Vision Computing, 2009, 27(1):189-205.
[30] De Padua P H C, Padua F L C, Sousa M T D, et al. Particle Filter-Based Predictive Tracking of Futsal Players from a Single Stationary Camera[C]// Conference on Graphics, Patterns and Images. 2015:134-141.
[31] Cheng X, Shiina Y, Zhuang X, et al. Player tracking using prediction after intersection based particle filter for volleyball match video[C]// Asia-Pacific Signal and Information Processing Association, 2014 Summit and Conference. IEEE, 2015: 1 - 4.
[32]Chai Y J, Park J Y, Yoon K J, et al. Multi target tracking using multiple independent particle filters for video surveillance[C]// IEEE International Conference on Consumer Electronics. IEEE, 2011:735-736.
[33] Xu T, Li D. NOVEL PARTICLE FILTER ALGORITHM BY DYNAMICALLY ADJUSTING PARTICLE NUMBERS [J]. Computer Applications & Software, 2013, 30(9):287-282.
[34] Jinxia Y U, Jingmin X U. Adaptive particle filter tracking algorithm by fusing multi-features [J], Computer Engineering & Applications, 2014.
[35] Fox D. KLD-Sampling: Adaptive Particle Filters.[J]. Advances in Neural Information Processing Systems, 2001:713-720.
[36] Koller D, Fratkina R. Using learning for approximation in stochastic processes[C]// 1998.
[37] Ying Yang, Shouxun Lin, Yongdong Zhang, Sheng Tang, “Highlights extraction in soccer videos based on goal-mouth detection,” in Proc. Int. Symposium on Signal Processing and Its Applications, pp.1-4, 2007.
[38] Ke Gao, Yongdong Zhang, Shouxun Lin, et al, “Visual stem mapping and geometric tense coding for augmented visual vocabulary,” IEEE Conference on Computer Vision and Pattern Recognition. 20 12.
[39] Liqiang Nie, Meng Wang, Yue Gao, Zheng-Jun Zha, Tat-Seng Chua:
Beyond Text QA: Multimedia Answer Generation by Harvesting Web Information. IEEE Trans. Multi media15 (2): 426-441 (2013).
[40] Liqiang Nie, Luming Zhang, Yi Yang, Meng Wang, Richang Hong, Tat-Seng Chua:Beyond Docto rs: Future Health Prediction from Multimedia and Multimodal Observations. ACM Multimedia 2015: 5 91-600
[41] Liu A A, Su Y T, Nie W Z, et al. Hierarchical Clustering Multi-task Learning for Joint Human Action Grouping and Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016:1-1.
[42]A.-A. Liu, W.-Z. Nie, Y. Gao, and Y.-T. Su, "Multi-modal clique-graph matching for view-based 3 D model retrieval," IEEE Transactions on Image Processing, vol. 25, no. 5, pp. 2103–2116, May 2016.
[43] Tong Hao, Ai-Ling Yu, Wei Peng, Bin Wang, Jin-Sheng Sun. Cross Domain Mitotic Cell Recognition, Neurocomputing, 2016, 195(C):6-12
[44]Nie, Wei-Zhi, An-An 3 Liu, Zan Gao, and Yu-Ting Su. "Clique-graph matching by preserving glo bal & local structure." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recogn ition, pp. 4503-4510. 2015.
[45]Hao T, Peng W, Wang Q, et al. Reconstruction and Application of Protein–Protein Interaction Network[J]. International Journal of Molecular Sciences, 2016, 17(6):907.

Highlights

1、 the sports field segmentation is based on salient region detection

2、 a method of soccer player detection was proposed, a fusion of the Otsu Algorithm and edge detection method

3、 a soccer player tracking algorithm based on enhanced particle filter, and mixture distribution of combination of prior distribution and observation distribution, and the likelihood functions by combining the color feature and edge feature.