Tracking and Recognition in Sports Videos

Mustafa Teke [a], Masoud Sattari [b]
[a] Graduate School of Informatics, Middle East Technical University, Ankara, Turkey
mustafa.teke@gmail.com
[b] Department of Computer Engineering, Middle East Technical University, Ankara, Turkey
masoud.sattari@gmail.com

ABSTRACT

Sport is one of the most growing industries in the world. As a result, media, teams and corporations are interested collecting statistics about the events to evaluate individual players and teams. These statistics might be average distance players run; number of passes completed and ball loses between teams. In this work, a method developed which is a composition of tracking and classification. Classification is achieved using background subtraction. Teams Referee and ball are trained by using objects which are extracted from tracking. Classification is performed using K-NN method.

Keywords: Tracking, Classification, Background Subtraction

INTRODUCTION

Current topics in sports video processing is described by [1]. Tracking is one of the issues in processing of sports videos. Tracking is a generally solved problem. Kalman filters are used to track object, [2] suggest a framework based on Kalman filters to track ball in broadcast videos. Another Kalman filter based method [3] detects and tracks ball uses ball trajectory. Background subtraction is a technique for segmenting foreground objects and it is widely used in video surveillance applications. It helps to detect changes in the environment and to track the moving objects in the environment. Background subtraction aims to detect stationary objects by subtracting the current image from a reference background image. Utilization of Gaussian Mixture Models (GMMs) with background subtraction has become popular[4]. Gaussian mixture models based background subtraction methods are very powerful background subtraction methods if background changes very fast and more than one background distribution exist. An Improved Adaptive Gaussian Mixture model which is based on mixture of Gaussians is proposed in [5].

Team possession is determined by tracking ball on SVM based method [6]. In the [7] the playfield color model is constructed using Gaussian Mixture models. For recognizing player region a classifier (SVM) is trained in advance to distinguish between the interested and uninterested object. Particle filter is used to track player. It uses sequential Monte Carlo method for online inference within Bayesian framework.

Content insertion is one of the applications of sports video analysis. Most sports are closely linked with commercial advertisement. There is a need of customizing the video for local audience, which includes replacing some commercial banners in the field a typical example can be found in Wan et al. [8]. Another task is to track the referee in sport field. Ahmet Ekin et. al has done an extensive work including goal detection, referee detection, and penalty-box detection in [9].

METHOD DETAILS AND EXPERIENCES WITH TRACKERS

Our tracking method is based on [5]. In this method background is modeled by using adaptive Gaussian mixture models. This method is also a pixel-based background subtraction technique like mixture of Gaussians method and a pixel could belong either to background or foreground. This method is powerful to detect background and foreground pixels. Also, since this method does not use a fixed number of components, it is proposed as more adaptive and robust when compared to mixture of Gaussians method. In other words, Improved, Adaptive Gaussian Mixture model [5] could automatically select the proper number of components per pixel and updates parameters (mean and standard deviation).

In order to make use of black and white image commands in Matlab, first we will convert each frame to black and white image. One of the useful functions that is used in this area is *regionprops*. It calculates some properties of image regions. We have used 3 of this proprties in this work as follows: *Area*, *MajorAxisLength and MinorAxisLength*. Area returns area of each region, MajorAxisLength specifies the length of the major axis of the ellipse (in pixels) that has the same normalized second central moments as the region, MinorAxisLength specifies the length of the minor axis of the ellipse (in pixels) that has the same normalized second central moments as the region. Also we apply two known function *Imopen* and *Imclose* to remove some noise in each frame.

Our goal is to track the player and the ball inside the field. In this step for each image we have moving regions and we will classify each region as four classes: Red player, White player, Referee and ball.First we use our information that we got from last step, if we calculate the ratio of *MajorAxis* and *MinorAxis* and assigning some treshold we can distinguiesh between human and ball. For ball we assume that this ratio changes around the 1 but, it will be vary for human at least near 3.We also use *Area* of each region to make an estimation on region size that may help us remove some redundent region.

Now we will classify each region as a specific class that mentioned before. K-Nearest Neighbors (K-NN) is one of the simplest and most useful classifiers. In K-NN all sample data is distributed n-dimensional Euclidian space. Mahalanobis distance can also be used instead of Euclidian distance. Classification is delayed until a new instance is acquired. K-NN aims to collects k (n) nearest neighbor. It starts with window V_n. Window size is increased until it covers k samples.

$$p_n(x) = \frac{k_n/n}{V_n}$$

In this method classification is done based on majority voting. Fig. 1 shows a window selection centered on black point, unknown class. Green and Red dots belongs to different classes, say G and R. According to majority voting Black dot belong to Red class. Distance weighted k-Nearest Neighbor algorithm assigns weights to neighbors based on their distance. Inverse square of distance could be used as a weight.

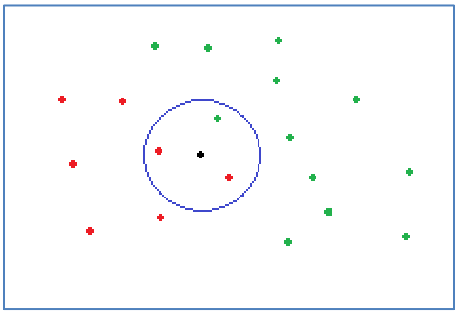

Fig. 1 K-NN classification

K-NN is a very effective against noisy data. Learning is very simple. On the other hand classification is costly. For each region we get RGB color and calculate the Mean value of them also, we calculate Standard Deviation. There is a pattern in different regions for these features so; they will be used to classify new regions. From a training video we have selected 30 frames and among these frames we have selected 70 regions to train to classifier. Each region will be assign to one of these classes: Red team player, White team player, Referee, Ball. Fig.2 shows a workflow that illustrates each step in algorithm.

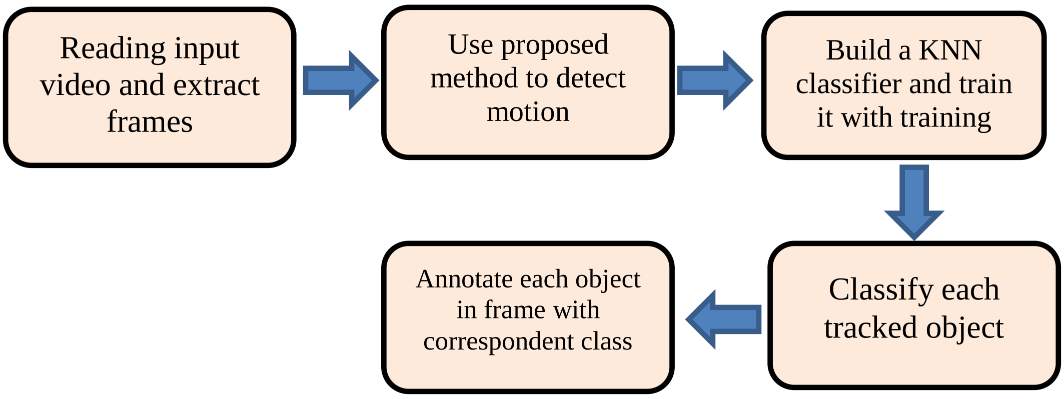

Fig. 2. Algorithm Workflow

EXPERIMENTAL RESULTS

In the tests, VS-PETS'2003 [8] data set used for tracking, training and classification. PETS data sets contains videos from outdoor events and sports. Evaluation is done for 2 tasks: Tracking, Classification. As we know *Precision* and *Recall* are the common criteria in this area. For simplicity we will consider only inside the field for example the coach or guards in margin of field are not counted.

$$\text{Precision} = \frac{tp}{tp + fp} \qquad \text{Recall} = \frac{tp}{tp + fn}$$

In case of Tracking, TP is number of objects that are correctly tracked, FP is number of objects that are tracked but they are not our interest, FN is number of our interested objects that are not tracked. Result for tracking evaluation for 18 frames is shown in Table 1.

TP	*FP*	*FN*	*Precision*	*Recall*
16	2	4	0.888889	0.8
16	1	3	0.941176	0.842105
16	3	3	0.842105	0.842105
16	5	3	0.761905	0.842105
16	5	3	0.761905	0.842105
15	5	4	0.75	0.789474
16	2	3	0.888889	0.842105
15	5	4	0.75	0.789474
17	5	2	0.772727	0.894737
17	8	2	0.68	0.894737
17	3	2	0.85	0.894737
16	0	3	1	0.842105
16	3	3	0.842105	0.842105
17	1	2	0.944444	0.894737
17	2	2	0.894737	0.894737
17	1	2	0.944444	0.894737
17	0	2	1	0.894737
17	0	2	1	0.894737

Table 1. Tracking Results for Several Frames

In case of classification, as an example for White team classification, TP is number of objects that are correctly classified as white player, FP is number of objects that are classified as white player but actually they are not white player, FN is number of white players that are not classified as white player. Result for classification for White team, Red team and Referee for 7 frames are show in Table2, Table3 and Table4.

TP	*FP*	*FN*	*Precision*	*Recall*
6	1	1	0.857142857	0.857142857
7	2	1	0.777777778	0.875
8	0	0	1	1
8	1	1	0.888888889	0.888888889
8	3	0	0.727272727	1
7	1	1	0.875	0.875
7	2	1	0.777777778	0.875

Table 2. Classification Results for Team A (White Team)

TP	*FP*	*FN*	*Precision*	*Recall*
6	2	3	0.75	0.666667
6	0	1	1	0.857143
6	1	2	0.857143	0.75
6	1	1	0.857143	0.857143

6	3	1	0.666667	0.857143
7	2	0	0.777778	1
7	1	0	0.875	1

Table 3. Classification Results for Team B (Red Team)

TP	*FP*	*FN*	*Precision*	*Recall*
1	2	0	0.333333	1
1	1	0	0.5	1
2	2	0	0.5	1
2	1	0	0.666667	1
2	0	0	1	1
1	2	1	0.333333	0.5
2	2	0	0.5	1

Table 4. Classification Results for Referee

COMPARING WITH OTHER METHODS

When we read literature can find that there are several methods for object tracking in video that have good results for different datasets. For example: Meanshift, Lucas-Kanade Tracker, Codebook, Motion histogram and Kalman filter. From available methods we have selected Codebook and Motion histogram and tested them on our dataset. Results are presented in Table 6 and Table 7.

TP	*FP*	*FN*	*Precision*	*Recall*
14	0	8	1	0.636364
15	2	2	0.882353	0.882353
15	5	4	0.75	0.789474
14	5	4	0.736842	0.777778
15	2	4	0.882353	0.789474
15	4	2	0.789474	0.882353
15	3	4	0.833333	0.789474
15	4	4	0.789474	0.789474
15	5	4	0.75	0.789474
14	3	4	0.823529	0.777778
15	5	4	0.75	0.789474
15	3	4	0.833333	0.789474
15	5	4	0.75	0.789474
16	4	4	0.8	0.8
15	3	3	0.833333	0.833333
15	5	4	0.75	0.789474
15	5	4	0.75	0.789474
14.88235	3.705882	3.941176	0.806119	0.793217

Table 5. Tracking result for Motion History method

TP	*FP*	*FN*	*precision*	*recall*
16	3	3	0.842105	0.842105
17	4	2	0.809524	0.894737
18	5	3	0.782609	0.857143
18	6	3	0.75	0.857143
19	5	2	0.791667	0.904762
19	3	2	0.863636	0.904762
18	3	3	0.857143	0.857143
18	4	3	0.818182	0.857143
18	4	3	0.818182	0.857143
18	5	3	0.782609	0.857143
17.9	4.2	2.7	0.811566	0.868922

Table 6. Tracking result for Codebook method

	TP	*FP*	*FN*	*Precision*	*Recall*
Proposed Method	16.33	2.83	2.72	0.86	0.86
Motion Histogram	14.88	3.71	3.94	0.81	0.79
Codebook	17.9	4.2	2.7	0.81	0.87

Table 7. Comparing our method with Codebook and Motion histogram

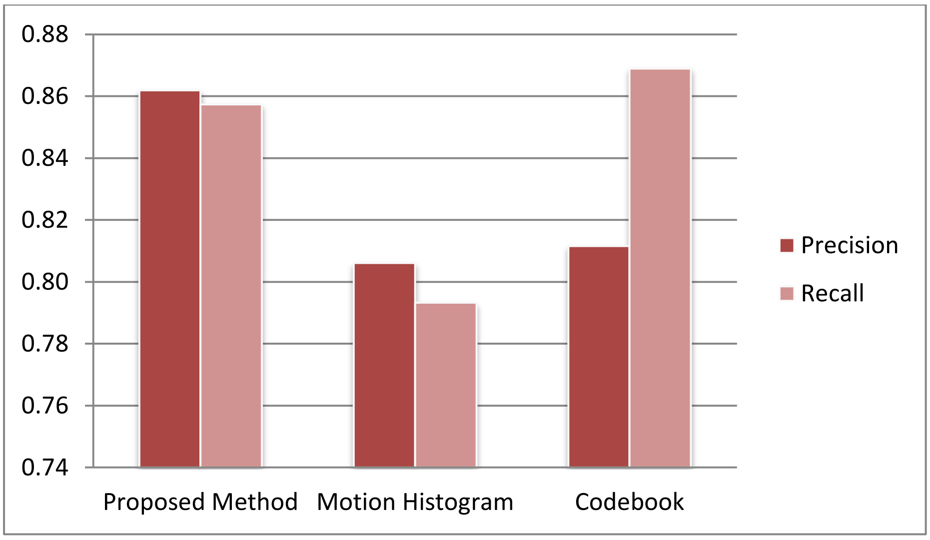

Figure 3. Diagram presentation of table 7

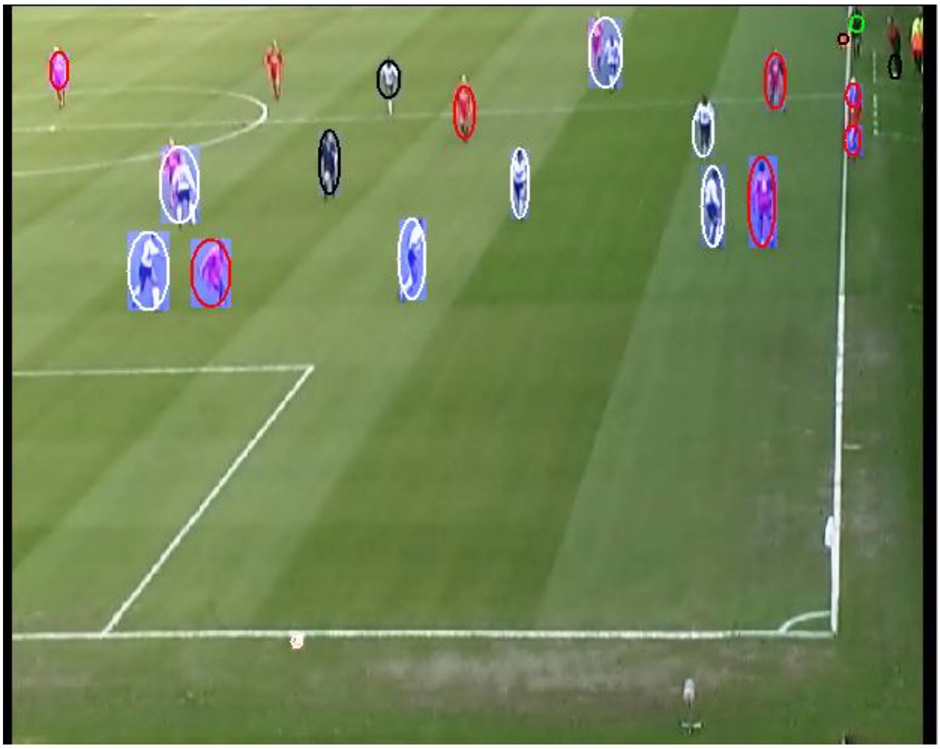

Figure 4. Visual results of tracking and recognition of the proposed method

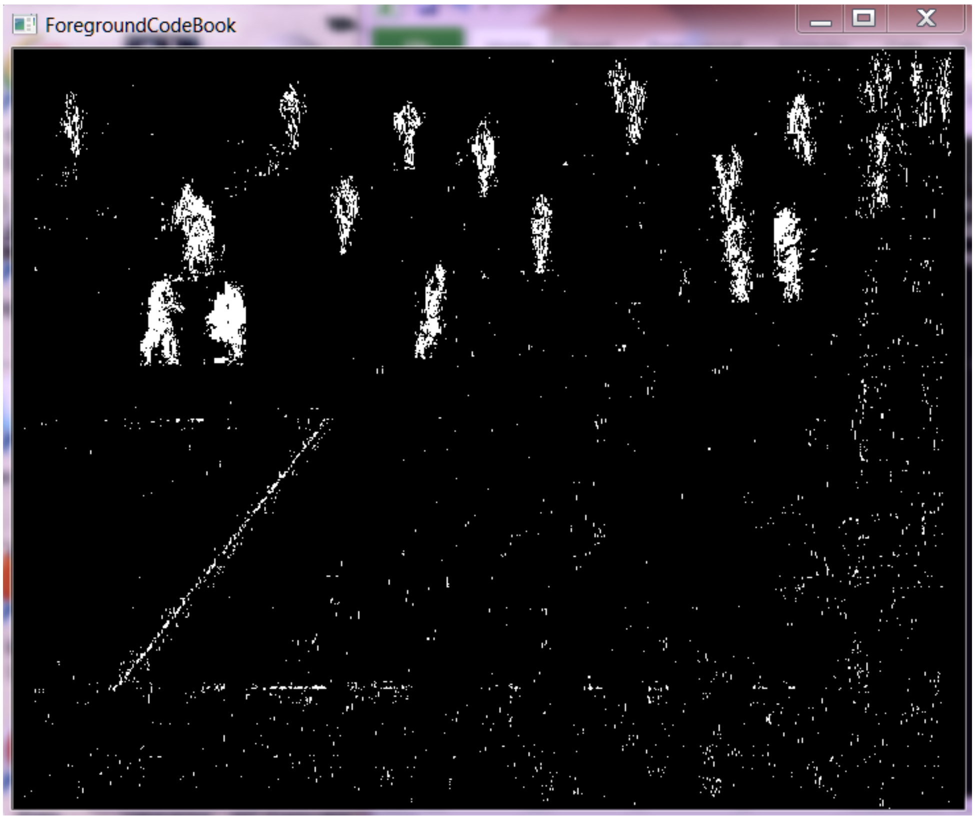

Figure 5. Visual results of tracking of Codebook method

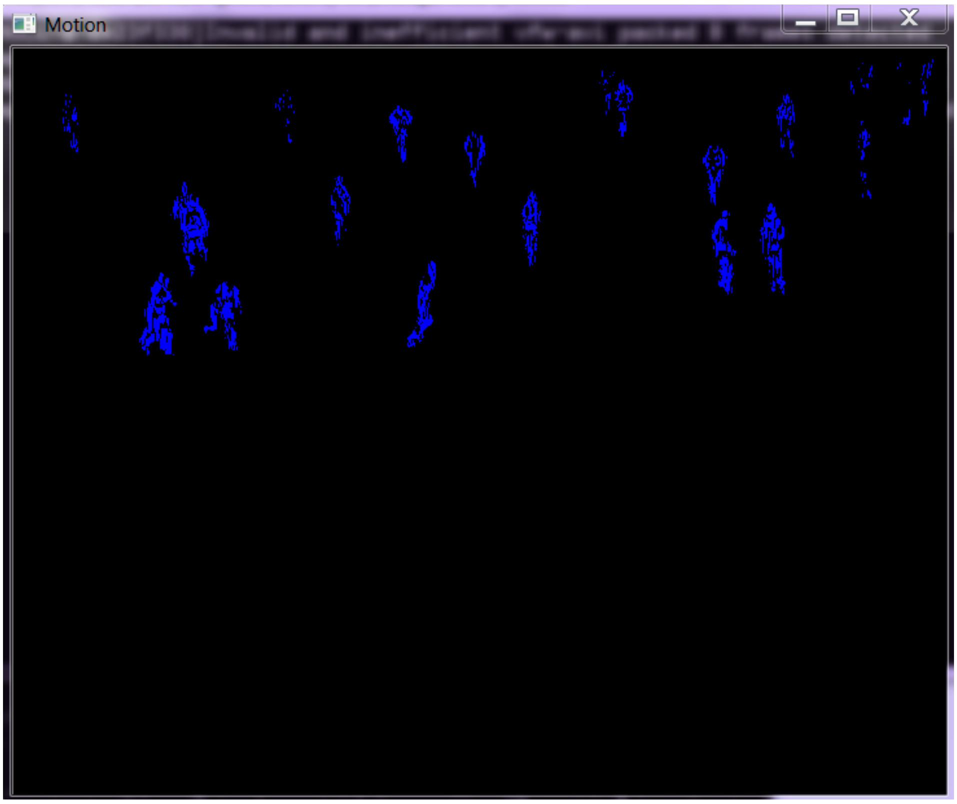

Figure 6. Visual results of tracking of Motion histogram method

Conclusions

Several works has been done in this field and various methods have been proposed. I this work we combine object targeting with classification technique I we hope to get good result. As figure 3. Shows comparing other two methods our algorithm has a high precision but for recall the proposed method is less than Codebook recall. Increasing the training data may affect the efficiency of our method. This method is not suitable for real-time videos because the classification is inherently slow.

Source Code is available in http://mustafateke.com/projeler/sports-videos-tracking/ under GPL.

References

[1] Yu, X. and Farin, D., "Current and emerging topics in sports video processing," IEEE International Conference on Multimedia and Expo, 2005. ICME 2005.
[2] Yu, X.; Xu, C.; Tian, Q., "A ball tracking framework for broadcast soccer video", in Proc. of ICME 2003
[3] Yu, X. and Xu, C. and Leong, H.W. and Tian, Q. and Tang, Q. and Wan, K.W., "Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video," in Proceedings of the eleventh ACM international conference on Multimedia, 2003
[4] Stauffer, C. and Grimson, W.E.L. (1999), Adaptive background mixture models for real-time tracking, Proceedings of IEEE Computer Vision Pattern Recognition.
[5] Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction, Proceedings of the 17th International Conference on Pattern Recognition, 28-31.
[6] Yu, X. and Leong, H.W. and Lim, J.H. and Tian, Q. and Jiang, Z., "Team possession analysis for broadcast soccer video based on ball trajectory," International Conference on Information, Communications and Signal Processing, 2003
[7] Automatic Multi-Player Detection And Tracking In Broadcast Sports Video Using Support Vector Machine And Particle Filter , Guangyu Zhu1, Changsheng Xu2, Qingming Huang3, Wen Gao1,3
[8] Wan, K; Yan, X; Yu, X & Xu, C S (2003b). Robust goalmouth detection for virtual content insertion, Proc. of ACM MM'03, Berkeley, pp.468-469.

[9] Automatic Soccer Video Analysis and Summarization , Ahmet Ekin, A. Murat Tekalp, Fellow, IEEE, and Rajiv Mehrotra, IEEE Transactions On Image Processing, Vol. 12, No. 7, July 2003
[10] http://www.cvg.rdg.ac.uk/slides/pets.html