LETTER

A Fully Automatic Player Detection Method Based on One-Class SVM

Xuefeng BAI[†], **Tiejun ZHANG**[††], **Chuanjun WANG**[†], **Ahmed A. ABD EL-LATIF**[††,†††], ***Nonmembers***, ***and*** **Xiamu NIU**[†,††a)], ***Member***

SUMMARY Player detection is an important part in sports video analysis. Over the past few years, several learning based detection methods using various supervised two-class techniques have been presented. Although satisfactory results can be obtained, a lot of manual labor is needed to construct the training set. To overcome this drawback, this letter proposes a player detection method based on one-class SVM (OCSVM) using automatically generated training data. The proposed method is evaluated using several video clips captured from World Cup 2010, and experimental results show that our approach achieves a high detection rate while keeping the training set construction's cost low.

key words: *player detection, one-class SVM, decision-making, broadcast sports video analysis*

1. Introduction

Due to its popularity and potential commercial value, sports video analysis has attracted a great deal of attention. As an important part in sports video analysis, many methods have been proposed for player detection.

The difficulty of player detection lies in appearance changes of various players. In [1] and [2], rule-based detection methods are proposed. Some rules are predefined on certain properties such as color, local edge, shape and area to filter non-player regions. These rules are effective for non-overlapping players. But when there is occlusion with other objects, players cannot be detected.

To address the limitation of the rule-based detection methods, several learning based player detection methods are proposed. Zhu et al. [3] proposed a two-class SVM (TCSVM) based player detection method, which trains a TCSVM detector using HSV histogram. Histogram of Gradient (HOG) is also used to train TCSVM based detector in [4]. Another category of training based player detection methods follows the Viola and Jones' object detection framework [5]. In [6] and [7], Haar-like features are extracted and boosted to detect players. Recently, Xing et al. [8] have added the gradient information into the weak feature to train the Adaboost detector.

Manuscript received May 24, 2012.

Manuscript revised September 28, 2012.

[†]The authors are with the Shenzhen Graduate School, Harbin Institute of Technology, 518055 Shenzhen, China.

[††]The authors are with the School of Computer Science and Technology, Harbin Institute of Technology, No.92, West Da-Zhi Street, Harbin, China.

[†††]The author is with the Mathematics Department, Faculty of Science, Menoufia University, Shebin El-Koom 32511, Egypt.

a) E-mail: xm.niu@hit.edu.cn

DOI: 10.1587/transinf.E96.D.387

Benefiting from the better description ability of learning algorithm, detection results of above methods are more robust. Fewer missed and false detected players appear in detection results of these methods. However, there are still two drawbacks of these learning-based detection methods:

1. To obtain a desired detection rate, training set of these methods should be labeled and selected carefully. Such manual data collection process requires a lot of labor, and thus automating this process is highly desired.
2. For player detection methods using two-class classifiers, negative samples are selected randomly from the background. Thus, it is hard to estimate pattern of negative samples, which will lead to some false detections. Therefore, classifiers learning based on just positive samples should be introduced.

To handle these problems, we propose a fully automatic player detection method based on one-class classifier. The novelty of our approach lies in its ability to achieve considerable detection rate using the automatically generated training data. Details of this method are described in the following sections.

2. Proposed Automatic Player Detection Method

An overview of the proposed automatic detection method is shown in Fig. 1. In the proposed method, there are mainly three key issues that should be addressed, including: automatic training data generation, classifier selection and decision-making.

2.1 Automatic Training Data Generation

In order to perform fully automatic player detection, training data of players should be collected automatically. In this step, considerable number of player samples should be collected while maintaining minimal manual intervention. Taking into account this requirement, we exploit rule-based method to collect the training data automatically.

Since it is inevitable to collect some false detected players using rule-based detection strategy, further filtration should be done to purify the training data. As both gradient and color are important for player detection, filtration using both of these two features are performed. Specially, we build a median model that can distinguish between positive and negative samples for gradient-based filtration. Then, the rest samples are grouped on color features using an unsupervised cluster method.

2.2 Classifier Selection

As there is no clear criterion for negative samples selection in player detection, they are collected from the background randomly. It is difficult to estimate such non-uniform pattern using classical two-class classifier. Thus, the classifier used for training players detector should focus on the uniform player samples.

Herein, one-class classifier is introduced. The term one-class originated from [9]. Objective of one-class classification is trying to find a decision hyperplane around the target pattern (i.e., the players in player detection). Patterns that lie inside this decision hyperplane are classified as targets. So, modeling player detection as one-class problem efficiently overcomes the problem of representing non-player pattern.

2.3 Decision-Making

Since the training samples are collected automatically, it is inevitable that some samples are very close to decision boundary or a little different from training samples. Thus, the decision boundary will deviate from the standard decision-making boundary [10]. To handle this problem, a new decision-making function should be designed instead of using sign function.

Essence of decision-making is identifying the pixels belonging to players. Considering the decision values will be uncertain about which class they should exactly belong to, we introduce the Fuzzy C-Means (FCM) [11] algorithm to perform the decision-making operation. Advantage of the FCM based decision-making method lies in its ability of being able to describe and process such uncertainty.

2.4 Scope of Work

The proposed method has several merits for player detection. First, manual labor is greatly reduced, as the training set is obtained automatically, rather than labeled and selected manually. Second, one-class classification technology is introduced to perform the detection, which avoid representing the clutter pattern of non-player samples. Third, a new decision-making process is designed to handle the deviation of decision boundary.

3. Implementation Details

In this section, we describe the implementation details of the proposed automatic player detection method.

3.1 Rules for Automatic Training Data Collection

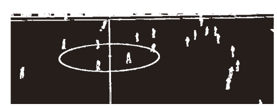
We first remove non-playfiled regions as in [12]. Next, we calculate the local range image using grayscale of the input image, where each pixel contains the range value (maximum value - minimum value) of the 3×3 neighborhood around the corresponding pixel in the input image. Then, a binary image is obtained by thresholding the LRV image using half value of the Otsu's threshold. Finally, we define a set of rules as in Eq. (1) to collect player samples automatically.

$$\begin{cases} 3 < w, 3 < h \\ w < h, \frac{w}{h} < \theta_{asp} \\ \theta_{min} < area < \theta_{max} \end{cases} \tag{1}$$

where h and w represent the height and width of the candidate region, $area$ is the number of pixels in the region, θ_{asp} is aspect ratio threshold, θ_{min} and θ_{max} are the minimum and maximum threshold of the $area$. In our implementation, for the video with resolution of 720×404, $\theta_{asp} = 0.8$, $\theta_{min} = 9$ and $\theta_{max} = 400$. An example of the binary image and player detection result is shown in Fig. 2.

3.2 Distance Based Samples Filtration Using Gradient

To filter false player samples, a gradient model using HOG is built as Eq. (2):



(a) Binary result of LRV image

(b) Rule-based detection result

Fig. 2 An example of rule-based player detection.

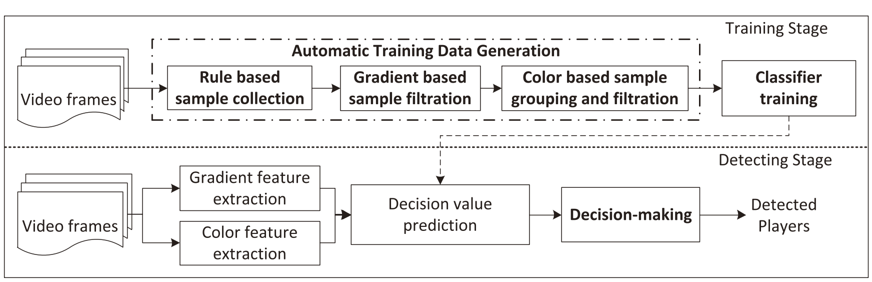


Fig. 1 Overview of the proposed method.

$$M_{hog} = mean(h_i), i = 1, 2, \ldots, N \tag{2}$$

where N is the number of samples detected, h_i is the HOG feature of a single sample. Then, the algorithm as described in Proc. 1 is used to filter samples by gradient, where $dist(\cdot, \cdot)$ is the Euclidean distance.

Proc. 1: Gradient-Based Filter Procedure

Input: Γ, the set of all candidate samples
Output: Ψ, filtered samples
Set $\Psi = \{\}$;
for *each element* $h_i \in \Gamma$ **do**
  $D_i = dist(M_{hog}, h_i)$
end
$\Lambda = \{x | x \in TopN(D_i)\}$, where $TopN$ obtains the largest 20% in D_i;
$\Psi = \Gamma - \Lambda$;
return Ψ;

3.3 GMM Based Samples Grouping by Color Feature

To group player samples unsupervisedly, we exploit the Gaussian Mixture Model (GMM) cluster algorithm. We use as the color feature the histogram of each component in HSV color space. Each of the HSV components is quantized into 10 bins uniformly. Then, Principal Component Analysis (PCA) is used to reduce the dimension of color feature. The first two dimensions of PCA in feature space are used in this letter.

We use algorithm as shown in Proc. 2 to group and filter samples by color feature, where T is set as 1000. GP_1, GP_2 will be identified as team 1 and team 2, and GP_3 will be dropped as the outlier.

Proc. 2: Color-Based Grouping and Filter Procedure

Input: X, candidate samples; T, number of iterations
Output: $GP_i, i = 1, 2, 3$, grouped samples
j=1;
Extract feature X_c from X as described in Sec.3.3;
while $j \leq T$ **do**
  Initialize the GMM parameters using random strategy;
  Use Expectation Maximization (EM) algorithm to estimate parameters in GMM model G_j for X_c;
  Record center C_j with respect to G_j;
  j++;
end
$\overline{C} = mean(C_1, C_2, ..., C_T)$;
$C^* = \min\limits_{1 \leq j \leq T} \left(dist\left(\overline{C},\ C_j\right)\right)$, $dist(\cdot, \cdot)$ is the Euclidean distance;
Record the model G^* with respect to C^*;
Assign the cluster label L to elements of X using G^*;
Set $GP_1 = max(count(L))$, $GP_2 = mid(count(L))$, $GP_3 = min(count(L))$, where $count(\cdot)$ counts the number of each cluster;
return GP_i;

3.4 Training Detectors Using One-Class SVM

Based on the above analysis, training process should focus on the target class. Due to its excellent generalization performance, one-class SVM (OCSVM) [13] is used in this letter. We train three OCSVM detectors for each video clip in the implementation. They are D_{hog}, D_{hsv_1} and D_{hsv_2}, where D_{hog} is trained using HOG feature on both two team players, D_{hsv_1} and D_{hsv_2} are trained using HSV feature on team 1 and team 2 players separately. The final detection result D is given by Eq. (3).

$$D = \{D_{hog} \cap D_{hsv_1}\} \cup \{D_{hog} \cap D_{hsv_2}\} \tag{3}$$

3.5 Decision-Making

As mentioned above, we choose FCM to perform decision-making in this letter. We assume that there are three types of decision value: background, players and gray zone between the background and players. Consequently, a three components FCM cluster method is used.

In the implementation, we normalize the decision value into $[0, 1]$ firstly. And then, considering the distribution of the decision value, we assume the center value of the background, gray zone and players are 0.2, 0.5, 0.8 separately. Thus, we initialize the membership matrix $V_{(i,k)}$ as Eq. (4).

$$V_{(i,k)} = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}} \tag{4}$$

where x_i is decision value of the ith pixels, $k = 1, 2, 3$, $\mu_k = \{0.2, 0.5, 0.8\}$ and $\sigma_k = \{0.05, 0.05, 0.05\}$ separately, and $V_{(i,k)}$ is normalized in accordance with the rows. Let $U_{i,k}$ be the membership value of pixel x_i to the kth cluster estimated by FCM, we distinguish the player pixels from the background according to Eq. (5):

$$D(x_i) \in Player,\ if\ U_{i,3} > U_{i,2} + U_{i,1} \tag{5}$$

4. Experiments and Analysis

4.1 Data Description and Experimental Setup

For evaluation, we use an experimental dataset comprised of three video sequences with the resolution of 720×404. They are captured from different matches of the World Cup 2010. The number of long shot frames that selected from these matches is 287, 147 and 200 respectively. To show detection ability with little human intervention, positive samples used in experiments are all collected automatically. The negative samples for training two-class classifier are obtained by randomly selecting from background as in [8].

4.2 Comparison with Player Detection Using OCSVM and TCSVM

Detection results of classical TCSVM are shown in Fig. 3 (a). It can be seen that there are many false detected players. It is mainly caused by the non-uniform distribution of negative samples. Figure 3 (b) shows the detection results of classical OCSVM. As can be seen, some players are missing detected. The reason is that the training set

(a) (b) (c) (d) (e)

Fig. 3 Examples of detection results and visual comparison. (a): The classical TCSVM. (b): The classical OCSVM. (c): TCSVM with FCM decision-making. (d): OCSVM with FCM decision-making. (e): Adaboost detector using Haar-like features.

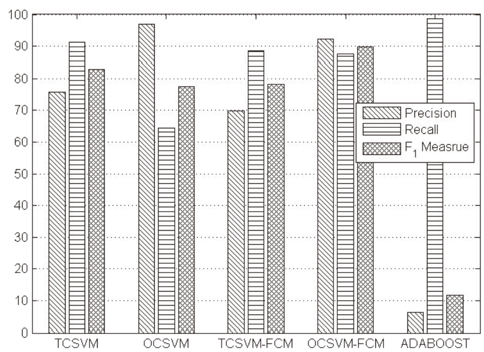


Fig. 4 Quantitative comparison of the detection result.

is imperfect due to the strategy of automatic generation of training data.

Detection results of TCSVM with FCM decision-making and OCSVM with FCM decision-making are shown in Fig. 3 (c) and Fig. 3 (d). It is clear from these two figures that there are even more false detected players for TCSVM with improved decision-making.The detection rate of the OCSVM with FCM based decision-making is greatly enhanced.

4.3 Comparison with Player Detection Using Adaboost

According to the Fig. 3 (e), there are a lot of false detected players in detection results of Adaboost detector. The reason for this is that the quality of training data is very important for train the Adaboost detector. The number of false detected players will increase if the training set was not collected carefully.

Quantitative comparison of the average detection rate are shown in Fig. 4. It can be seen that the proposed OCSVM with FCM decision-making detection method outperforms the competitive methods using the F_1 measure.

5. Conclusion and Future Work

This letter has presented a novel method to detect the players in soccer video automatically. Automatic training data generation strategy is introduced to player detection. Experimental results show that the proposed method significantly enhanced the usability of player detection. In the future, we will try to detect players in other types of sports video.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (60832010, 61100187), the Fundamental Research Funds for the Central Universities (Grant No. HIT. NSRIF. 2010046), the China Postdoctoral Science Foundation (2011M500666) and the Heilongjiang Province Educational Department Funds of China (12521107).

References

[1] O. Utsumi, K. Miura, I. Ide, S. Sakai, and H. Tanaka, “An object detection method for describing soccer games from video,” IEEE International Conference on Multimedia and Expo, pp.45–48, 2002.
[2] Y. Huang, J. Llach, and S. Bhagavathy, “Players and ball detection in soccer videos based on color segmentation and shape analysis,” Proc. 2007 International Conference on Multimedia Content Analysis and Mining, Weihai, China, pp.416–425, 2007.
[3] G. Zhu, C. Xu, Q. Huang, and W. Gao, “Automatic multi-player detection and tracking in broadcast sports video using support vector machine and particle filter,” IEEE International Conference on Multimedia and Expo, pp.1629–1632, Toronto, ON, Canada, July 2006.
[4] S. Mackowiak, M. Kurc, J. Konieczny, and P. Mackowiak, “A complex system for football player detection in broadcasted video,” International Conference on Signals and Electronic Systems, pp.119–122, Gliwice, Poland, Sept. 2010.
[5] P. Viola and M.J. Jones, “Robust real-time face detection,” Int. J. Comput. Vis., vol.57, no.2, pp.137–154, 2004.
[6] J. Liu, X. Tong, W. Li, T. Wang, Y. Zhang, and H. Wang, “Automatic player detection, labeling and tracking in broadcast soccer video,” Pattern Recognit. Lett., vol.30, no.2, pp.103–113, 2009.
[7] X. Tong, J. Liu, T. Wang, and Y. Zhang, “Automatic player labeling, tracking and field registration and trajectory mapping in broadcast soccer video,” ACM Trans. Intelligent Systems and Technology, vol.2, no.2, pp.15:1–15:32, 2011.
[8] J. Xing, H. Ai, L. Liu, and S. Lao, “Multiple player tracking in sports video: A dual-mode two-way bayesian inference approach with progressive observation modeling,” IEEE Trans. Image Process., vol.20, no.6, pp.1652–1667, 2011.
[9] M.M. Moya and D.R. Hush, “Network constraints and multi-objective optimization for one-class classification,” Neural Netw., vol.9, pp.463–474, April 1996.
[10] C. Voichita, P. Khatri, and S. Draghici, “Identifying uncertainty

regions in support vector machines using geometric margin and convex hulls," International Joint Conference on Neural Networks, pp.3319–3324, Hong Kong, China, June 2008.

[11] J.C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm.," Computers and Geosciences, vol.10, no.2-3, pp.191–203, 1984.

[12] X. Bai, T. Zhang, X. Song, and X. Niu, "Playfield detection using color ratio and local entropy," International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp.356–359, Dalian, China, Oct. 2011.

[13] B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, and R.C. Williamson, "Estimating the support of a high-dimensional distribution," Neural Comput., vol.13, no.7, pp.1443–1471, 2001.