

Player Detection based on Support Vector Machine in Football Videos

Chengjun Cui

Department of Physical Education, Nanjing University of Science and Technology, Nanjing, 210094, China

Abstract

An automatic player detection method based on fuzzy decision making one-class SVM is proposed. Detection results of statistical classifier player detection methods are better than rule based player detection methods. However, manually labelled training samples are used in these statistical classifiers based player detection methods. Thus, cost is very important. To resolve this problem, we propose an instinctive player detection method using fuzzy decision making one-class SVM and automatically collected player samples. In this method, one-class SVM (OCSVM) is introduced to train the player detector by drawing lessons from the human object category classification mechanism. Additionally, decision function of OCSVM is improved by dividing the decision value dynamically using the fuzzy decision method, which is able to reduce the detection error caused by the insufficient representativeness of the automatically collected training samples. Finally, a set of criteria is introduced to obtain the training samples automatically, and player detection experiments are performed on these training samples using FD-OCSVM. Experiments show that better detection results are obtained using the proposed method in the scenario of using automatically collected training samples, which improves the automatic degree of player detection.

Keywords: soccer videos; broadcast soccer video; object detection; support vector machine

(Submitted on November 16, 2017; Revised on December 29, 2017; Accepted on January 23, 2018)

© 2018 Totem Publisher, Inc. All rights reserved.

1. Introduction

The player is one of the main semantic objects in football videos and has an inseparable relationship with content analysis tasks regarding football tracking. Thus, the player detection is an indispensable partition of football video content analysis. By referring to research achievements by object detection method, scholars have proposed player detection methods based on statistical classifier [4]. In order to achieve the ideal detection result, such methods need to mark manually training sample sets, which will cause big expenses. Therefore, how to choose classifier and optimize/improve the automation degree of player detection has become an urgent problem in the field.

Player detection belongs to the issue of object type recognition, which is one of basic functions of the human visual cognitive system. By means of technical manners like neuroimaging, Kanwisher et al. revealed that the human high level visual cortex contains a limited number of class information processing zones, which are respectively responsible for the recognition of specific objects. Further studies indicated that those information processing zones oriented to a specific object class have certain closure property; that is, class information processing specific regions only deal with preferred stimuli with specified characterization, almost indifferent to non-preferred stimuli [9,10]. That is the specificity principle of class information processing zones in the visual system. It is the mechanism that guarantees that human beings recognize rapidly and precisely object classes such as human faces and human bodies [7].

Inspired by the specificity principle of human object class identification, this paper chooses one class support vector machine (OCSVM) to target player physical characteristic modeling and train the player detector, reducing the influence of complicated and changing non-player patterns on the training procedure. To increase detection performance, the paper analyzes the offset feature of OCSVM decision hyper-plane in the case of automatic acquisition of training samples; with the use of the fuzzy partition rule to optimize the decision function of original one class support vector machine, this paper

* Corresponding author.

E-mail address: chengjuncui2017@126.com

presents the detection method based on fuzzy decision one class support vector machine (FD-OCSVM), enhancing player detection effectiveness in the situation of automatically-obtained training samples.

2. Analysis of player detection problem and proposal of FD-OCSVM

Drawing lessons from general object detection approaches, researchers analyzing football video content analyses propose a few detection approaches based on statistical classifier. Of them, the method based on two class support vector machine (TCSVM) [11,2] and the one based on Adaboost are representative [6,12]. From the perspective of selection of classifier, those methods employ two kinds of classifier. The aim of training is to divide samples into two classes according to similarity of feature space. When data of two classes each have bigger intra-class similarity, two kinds of classifier can reach favorable classifying effect. For player detection, players have clear meaning and good within-class similarity, but non-player objects are not clearly defined with bad intra-class similarity. For training classifier, negative class samples (non-player object) are chosen randomly from the background [1]. The classifying performance of two classes of classifier acquired through the training of randomly selected samples will be somewhat affected.

The introduction of OCSVM can eliminate interference of messy non-player samples on the training, more in line with the requirement of automatic player detection. However, classical OCSVM method performs training through manual annotation of the training sample set. In order to realize the objective of automatic player detection, it is necessary to analyze the training and classifying process of OCSVM; then, based on features of automatic obtainment of training sample set, optimize the classifying process of OCSVM to get better player detection results [8,5,3].

In principle, the training process of OCSVM involves mapping training sample to position as far as possible from the original point in order to obtain the smallest hyper-plane boundary of a given training sample in the inter product space. For the feature space U of player sample, mark player training sample set $T = \{u_1, u_2, \dots, u_L\}$, and $\Phi(u_i)$ as mapping of U from inter product space. In order to obtain hyper-plane boundary of player training sample in inter product space, we can settle the quadratic optimization problem as shown in Equation 1 according to the training process of one class support vector machine.

$$\begin{aligned} & \text{Minimize } \frac{1}{2} \|w\|^2 - \rho + \frac{1}{vl} \sum_i s_i \\ & \text{s.t. } (w \cdot \Phi(u_i)) + s_i \geq \rho, s_i \geq 0 \end{aligned} \quad (1)$$

According to the duality problem of Lagrange's theorem (1):

$$\begin{aligned} & \text{Minimize } \frac{1}{2} \sum_{i,j} a_i a_j K(x_i, x_j) \\ & \text{s.t. } 0 \leq a_i \leq \frac{1}{vl} \end{aligned} \quad (2)$$

The two optimization problems of the Formula 2 can be used to determine the nonzero Lagrange multipliers a_i and support vector s_i . Thus, the OCSVM decision function $d(u^*)$ can be expressed as:

$$d(u^*) = \text{sign}(\sum_i a_i K(s_i, u^*) - \rho) \quad (3)$$

Classical OCSVM method uses the $\text{sign}(\cdot)$ function to determine the class of testing data. When training data are of sufficient quantity and good representation, the decision function can effectively predict testing data class. However, in player detection, the player training sample set, which is automatically obtained, can hardly meet the ideal condition. Training sample sets which are acquired automatically may have various problems, such as insufficiency, limited representation, and noise. If the acquired detector is trained with those training samples, the decision hyper-plane would deviate from its true classification hyper-plane. On that regard, fixed sign function utilized for decision-making will lead to wrong classification of partial testing sample; hence, this situation would appear in Equation 4:

$$\exists u_+^* \in U_+ : f(u_+^*) = \sum_i a_i K(s_i, u_+^*) - \rho < 0 \quad (4)$$

To conclude from above, training sample sets which are automatically obtained in player detection do not meet conditions like sufficient number and good representation and may cause offset of OCSVM decision hyper-plane. Since fixed $\text{sign}(\cdot)$ function is unable to address flexibly the offset of decision hyper-plane, applying the decision function will result in wrong class division of decision values around classification hyper-plane.

3. Automatic acquisition of training set

3.1. Partial difference criterion

In terms of material, play court consists of lawns. Thus, the field region must resemble color distribution. After converting to a grayscale image, pixels of play court generally have a consistent grayscale distribution. Meanwhile, for ornamental value, colors of other objects in the court, such as those of marking lines, the football and the players, differ largely from that of court pixels. Therefore, certain partial differences exist between the gray value of a player in the green field and that of court pixels, which are presented in Figure 1. Comparatively, foreground objects in the court have higher or lower grayscale value, with bigger partial differences. In this case, we regard partial difference of pixels as a criterion for foreground object detection. Based on the above analysis, this paper utilizes the filter in Equation 5 to reflect that part difference between objects and background in the court.

After convolution of Figure 1 by means of the filter, we get the image in Figure 2, where a) region and b) region belong respectively to the foreground object and background pixel area. Through in-depth observation, we note that the lower position of $\nabla G_{x,y}$ mostly belongs to the field area; foreground object area has higher $\nabla G_{x,y}$. It suggests the effectiveness of the proposed partial difference criterion and that the filter in Equation 5 effectively captures partial difference between court foreground object and the background, intensifying the identifiability of the foreground object in the image.

$$\nabla G_{x,y} = \max(G_{N_r}) - \min(G_{N_r}) \quad (5)$$

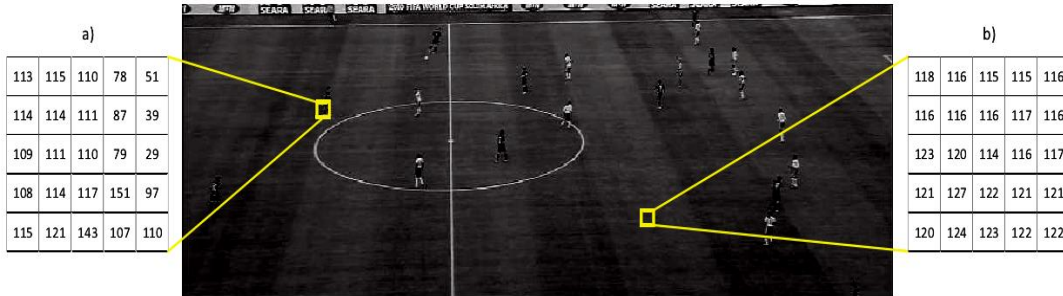


Figure 1. Characteristic of playfield pixels

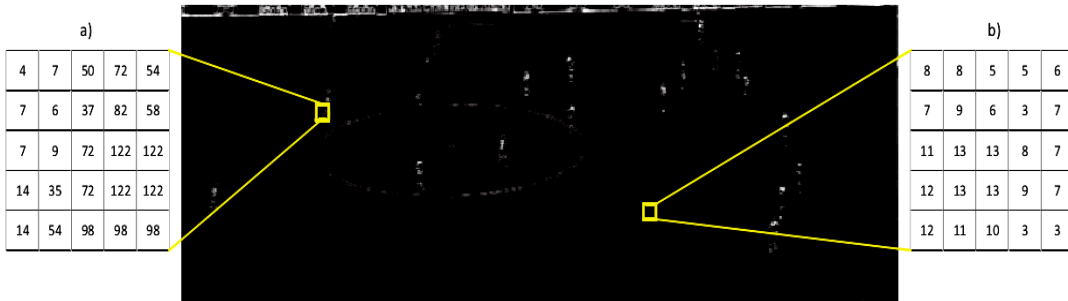


Figure 2. Example of the local range difference in playfield

Partial difference criterion strengthens the difference between object pixel and background pixel. With segmentation of intensified image by appropriate threshold, we can first separate foreground object area of the court. If the variance is bigger

between the foreground object area and the background region after separation, then the difference between them becomes bigger and wrong classification will happen less frequently. According to this principle, this paper chooses the criterion of maximum between-cluster to determine dividing threshold between court object pixel and background pixel after filtering. In order to avoid leak detection of foreground object area in the court, the paper binarizes the solution in Equation 6:

$$B_{x,y} = \begin{cases} 1, & \text{if } \nabla G_{x,y} \geq aT_0 \\ 0, & \text{if } \nabla G_{x,y} < aT_0 \end{cases} \quad (6)$$

The use of Formula 3 to obtain the value of the two imaginary examples is demonstrated in Figure 3. Careful observation of the map shows that the site area of the two-value image for the black area and the site of the foreground objects, especially the players, are almost all in the foreground area. This phenomenon shows the validity of the object detection method.

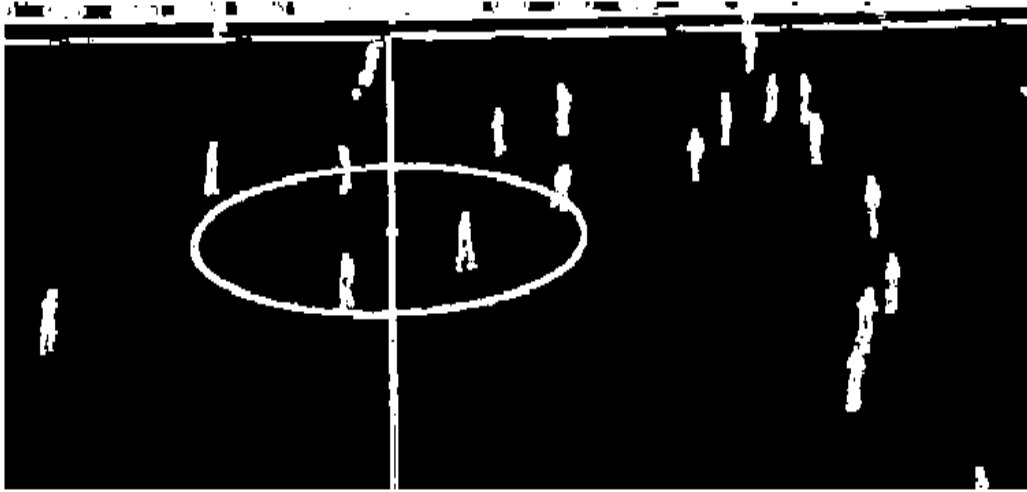


Figure 3. Example of the binarized local range image

3.2. Shape criterion

There is a significant difference in the shape of the individual players in the two-value image. Based on this difference, this paper distinguishes the players and other objects in the foreground objects by a set of shape rules.



Figure 4. An example of rule-based player detection

For ease of expression, w is the width of the region to be detected, h is the height of the area to be detected, θ_w is the width threshold of regional players, and θ_h is highly threshold regional players, area is the number of pixels to be detected. θ_{asp} is detection area aspect ratio, and θ_{min} and θ_{max} are area maximum and minimum values.

On the basis of this, the following shape criteria are defined:

- (1) $w > \theta_w$ is the regional players should have a certain width value,
- (2) $H > \theta_h$ is the regional players should have a certain height,
- (3) $w < h$, the width of the player area is greater than the height,
- (4) $\frac{w}{h} < \theta_{asp}$, namely the player area width and height ratio should be within a certain range,
- (5) $\theta_{min} < area < \theta_{max}$, the player area should be in a certain range.

The experiment using video resolution is 720*404. Typical value of above parameters is $\theta_w = 3$, $\theta_h = 3$, $\theta_{asp} = 0.8$, $\theta_{min} = 9$ and $\theta_{max} = 400$. The player area detected by the above criterion is shown in Figure 4.

Observing the figure carefully. We find that the set of rules can effectively detect independent player area. When a player is overlapping with other objects in the play court, its shape attribute will undergo big changes, thus missing the detection of the player. Overall, the method based on shape rule can detect a player who is not overlapping with other objects in the image.

3.3. Color criterion

The color of referee's clothing is very different from the color of player's clothing. In addition, the color of the player's clothing and the color of other objects in the field also have obvious differences. Therefore, color is one of the most important features in the players' area. Generally speaking, there are three main types of color in the field, and they belong to the two teams and the referees. In terms of quantity, the number of referees is far less than the number of players. Based on this analysis, the potential players in the field should meet the following criteria:

$$\begin{aligned} Num(P_1) &> Num(P_r) \\ Num(P_2) &> Num(P_r) \end{aligned} \quad (7)$$

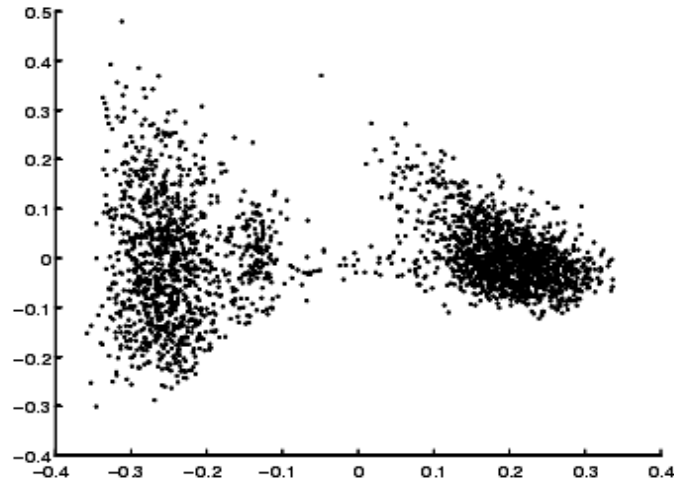
Based on the above criteria, this paper designs a method of selecting players based on color clustering. This method uses the differences in the colors of the players, divided into three categories of candidates in the sample: the team's players in the team, the team of two players, and the referees (including other objects). In this paper, the least number of results is classified as non-player objects and removed from the player samples.

Feature extraction is the premise of the division of the sample. In this paper, we use the HSV color histogram as the color feature of the cluster analysis:

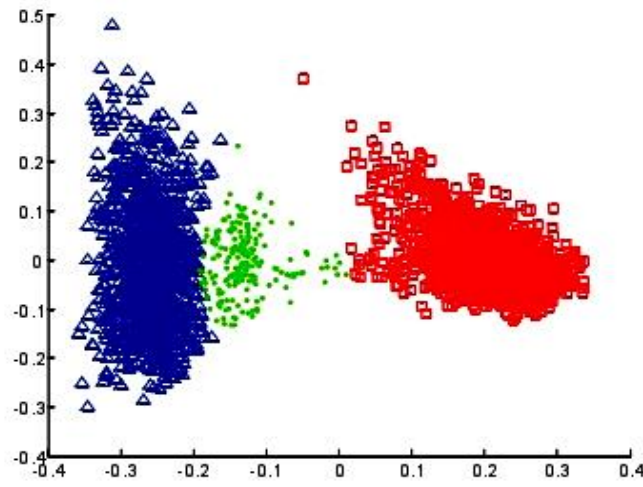
1. The region I_{rgb} is to be analyzed from RGB (Red-Green-Blue) color space into HSV (Hue-Saturation-Value) color space, denoted as I_{hsv} .
2. The color histogram is extracted on the three components of I_{hsv} , and the histogram of each component is quantized into 10 dimensions denoted as f_{hsv} .
3. Principal component analysis was performed for all the f_{hsv} , and the most representative of the first two dimensional features was used as clustering features.

According to the characteristics of the data distribution in the feature space, we use the Gauss mixture model (GMM) to estimate the distribution of the samples in the feature space. It is shown in Figure 5.

In general, GMM is maximized by the Expectation Maximization (EM) method to estimate the parameters of the model. But the EM method is sensitive to the initial value. In order to reduce this sensitivity, this paper adopts multiple clustering method to obtain the stable parameter estimation. The main process is described in algorithm 1.



a) Distribution of samples in HSV feature space



b) Cluster results obtained by Algo.1

Figure 5. An example of the player samples clustering

Algorithm 1 Color based player samples selection**Input:** X, Candidate sample set; T, iteration times**Output:** $GP_i, i = 1, 2, 3$, Sample after grouping

1. $j=1$
2. while $j \leq T$ do
3. Randomly initialize GMM parameters;
4. Using EM algorithm to estimate the GMM model G_j
5. Record each component center C_j in G_j ;
6. $j++$
7. End

4. Player detection based on FD-OCSVM

With training sample sets, it is possible to train OCSVM in order to get player detector. To detect candidate player region in one image, we can use the OCSVM detector acquired through training, i.e. we can obtain decision value collection in the candidate area. For convenience of analysis, we conduct range normalization of the decision value $f(u^*)$ of the same one frame image as per Formula 8. It is known from Figure 6 that range normalization does not affect the relative size of decisive value. Same as before normalization, the decision value after normalization is bigger, so that relative sample is highly similar with player.

$$f(u^*) = \frac{f(u^* - \min(f(u^*)))}{\max(f(u^*)) - \min(f(u^*))} \quad (8)$$

Manually constructed training sample set is chosen with more attention, containing training sample with abundant classes of better representation. Accordingly, the detector that completed training of such a training sample set has favorable generalizing ability. However, it requires huge man overheads to construct such a training sample set. Thus, the method of human constructing training sample is not practical. The paper presents a method which obtains the training sample set automatically. The method can greatly decrease the constructing cost of the training sample set and boost the practicability of player detection method. This is why we will use the training sample set acquired automatically to train player detector.

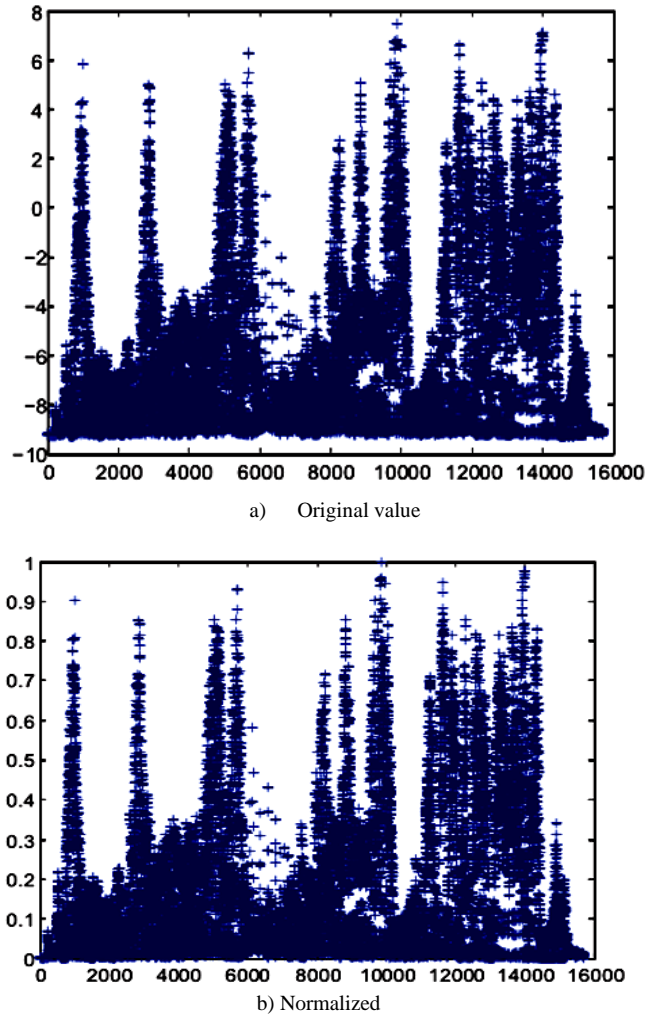


Figure 6. Examples of OCSVM decision values

Comparatively speaking, a training sample set acquired automatically is of bad representation. The training process of the proposed OCSVM is to find out a compact classification hyper-plane which includes target class samples. If training sample set is collected automatically, some pending players may differ from the used training sample. In the classification space of OCSVM, those players' relative data points maybe deviate from decision hyper-plane acquired through training. In order to avoid the error, this paper proposes FD-OCSVM to detect players.

According to Equation 5, the optimal division of FD-OCSVM is completed by FCM algorithm. Suppose the decision value given by OCSVM is composed of three parts: player, non-player and the gray region between them. From that, this paper takes advantage of the FCM algorithm which includes three components. Of them, the typical value of non-player, gray area and player center value is respectively 0.2, 0.5 and 0.8. Finally, this paper initializes membership matrix of FCM by following the method described in Equation 9.

$$V_{(i,k)} = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}} \quad (9)$$

x_i is the decision value of the i pixels, $V_{(i,k)}$ normalized by line. Membership matrix initialization results are shown in Figure 7. It can be found that the higher the decision value, the higher the level of the players.

In the above condition after initialization, after clustering by FCM algorithm, we can get membership value of all pending player areas to the three classes. Defining $U_{i,k}$ as the membership value of the x_i pixel area to the k th cluster, then the condition that the pixel region belongs to the player is defined as:

$$D(x_i) \in \text{player} = \begin{cases} 1, & \text{if } U_{i,3} > U_{i,2} + U_{i,1} \\ 0, & \text{if } U_{i,3} \leq U_{i,2} + U_{i,1} \end{cases} \quad (10)$$

5. Experimental Analysis and Results

A statistical classifier based on player detection method requires a certain number of player samples to verify its detection performance. This paper uses a sample of about 8000 players to test video images from more than 770 frames. In order to test the applicability of the method, four different game videos were used in the experiment, including 287 frames, 147 frames, 200 frames and 174 frames. Radio and TV game videos used in this experiment are from the 2014 World Cup, more specifically from games between South Africa and Mexico, Japan and Cameroon, Italy and Paraguay, and New Zealand and Slovakia. There are some differences in the background color of the selected game video in order to test the universality of the proposed method. At the same time, there are some similarities between the color and shape of the white football and the players in the white ball.

Therefore, the selection of the video game also includes the white ball players and the rapid movement caused by the vertical white football drag at the same time the situation. In order to facilitate processing, the resolution of the images used in the experiments was adjusted to $720 * 404$. At present, the player detection method based on statistical classifier is mainly based on TCSVM player detection method and player detection method based on Adaboost. Therefore, in this paper, the above two methods are used to compare the test results.

5.1. Visual contrasts of detection results by TCSVM and OCSVM

A few more missing players are found in the detection result by the TCSVM player detection method; for that reason, in TCSVM's training process both positive and negative training samples need to participate. The pattern of negative samples is chaotic and of bad intra-class similarity. Thus, when training sample set is of finite representation, the participation of negative samples brings down the classifying performance of TCSVM. Therefore, more wrong detection situations are found in the result. Less wrong detection is seen in the detection result by the OCSVM method, but missing detection of some players is also observed. This relates to the training process of OCSVM. OCSVM can obtain classification hyper-plane according to player sample which has better intra-class similarity and completes better modeling of player facade pattern in the training procedure, as to avoid interference by the chaotic patterns non-player samples. Under the circumstances of limited training sample set, it is possible to obtain better classification hyper-plane.

Basically, there are no wrong detections that occur in the detection result. However, constrained by the representativeness of the training sample, missing detection happens to the detection result by OCSVM method. Figure 8 a) shows examples of player detection results using FD-TCSVM player detecting method and FD-OCSVM player detecting method.

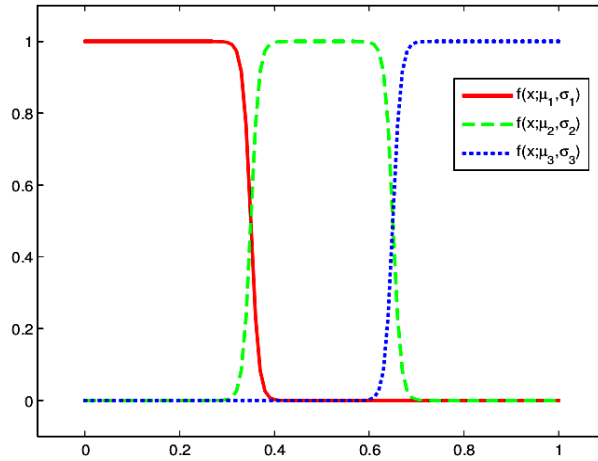


Figure 7. Illustration to the initialization of membership function

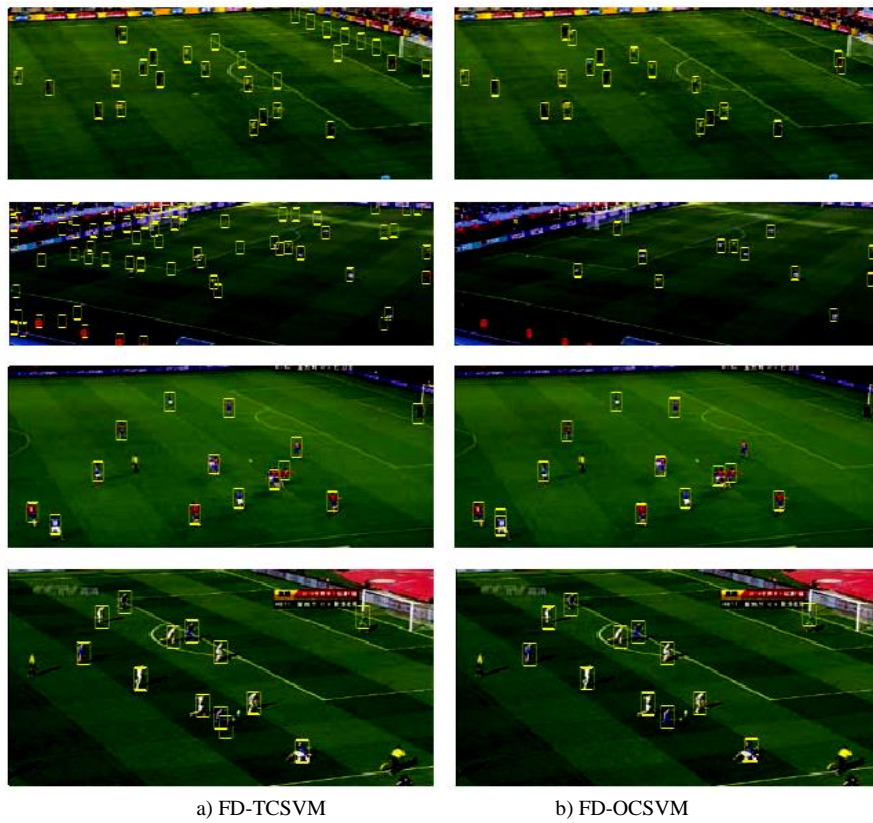


Figure 8. Examples of player detection result using FD-TCSVM player detecting method and FD-OCSVM player detecting method

5.2. Visual contrasts of detection results by FD-TCSVM and FD-OCSVM

Based on the principle of fuzzy division, we can improve decision function of OCSVM to get FD-OCSVM mentioned above. Based on a similar principle, we can improve the decision function of TCSVM. In order to compare the detection effect under similar conditions, we do similar improvements of TCSVM's decision function then compare player detection results by FD-TCSVM and FD-OCSVM. Examples of FD-TCSVM and FD-OCSVM test results are shown in Figure 9.

After decision functions are improved, there is more false detection of the result by FD-TCSVM. Influenced by the chaotic patterns of non-player samples, and the limited training sample, TCSVM works in predicting the decision value of the testing sample. Thus, when the decision is made as per the relative size of decision value, i.e. the decision is made through fuzzy division criterion, partial non-player area is falsely detected as player area. Comparatively, less false detection is seen

in the result by FD-OCSVM and leak detection problem is improved. The training process of OCSVM is interfered less by messy non-player patterns. Apparently, big differences exist between player area and non-player area in the decision value of OCSVM output. Thus, the introduction of the fuzzy division decision-making mechanism can overcome the declining detection performance as a result of bad representativeness of the training sample set.

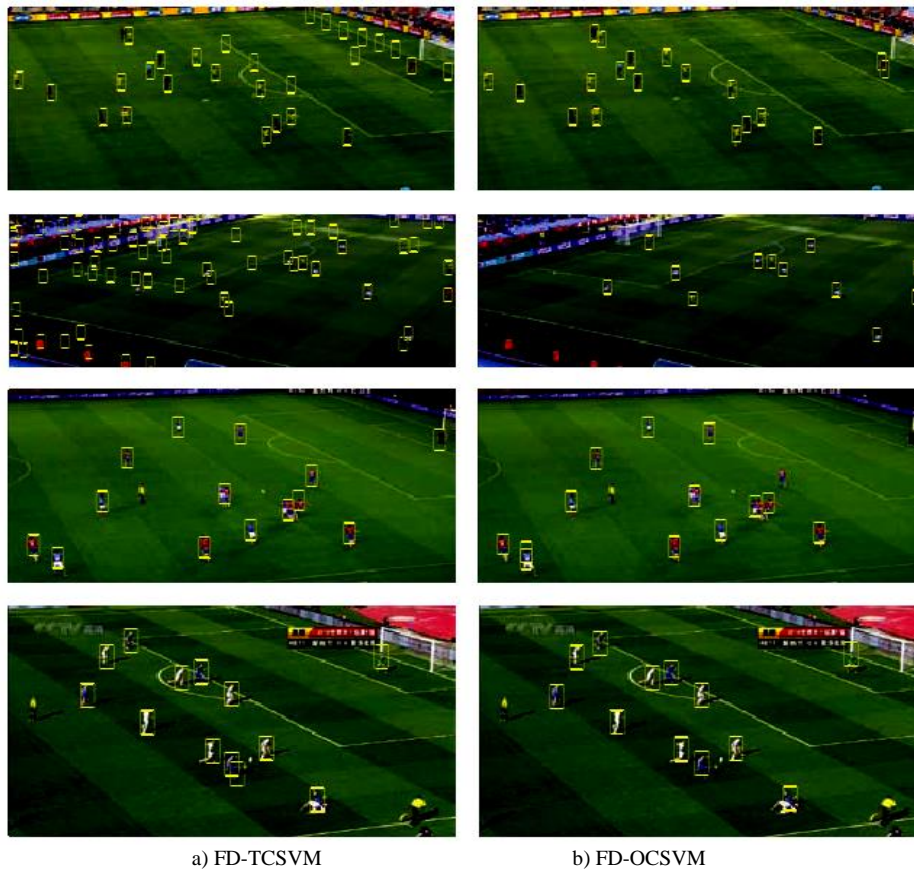


Figure 9. Examples of player detection result using FD-TCSVM player detecting method and FD-OCSVM player detecting method

5.3. Instance of Adaboost Player detection results

Likewise, with the training sample set which is acquired automatically, the detection result by Adaboost has a higher number of falsely detected players, which are obviously worse than methods based on TCSVM and OCSVM. From the training course of classifier, Adaboost classifier has higher requirements for the training sample set. When the training sample set is less representative, e.g. in the case of a training sample set acquired automatically, Adaboost classifier can hardly get weak classifier collection with stronger differentiation, degrading its generalizing performance. In the same condition, SVM can more effectively achieve modeling of player samples within the training sample set, and its generalizing capability is superior to the Adaboost method. Therefore, in the use of training sample sets acquired automatically, SVM based player detection method demonstrates better performance than that based on Adaboost. Figure 10 is based on the test results of Adaboost Player detection example.

6. Conclusions

The problem of player detection in broadcasted soccer videos is studied. This paper presents a method of automatic player detection based on fuzzy decision support vector machine (FD-OCSVM). The existing method of player detection based on the system classifier needs to use the training set to train the player detector. In order to solve this problem, firstly, the method of player detection based on statistical classifier is analyzed. This paper summarizes the shortcomings of the existing player detection methods in classifier selection and training sample collection. According to the decision value to reflect the characteristics of the test sample and the optimization of the training sample similarity decision function of OCSVM, a kind of support vector machine and fuzzy decision were put forward, thus improving decision hyper-plane offset by optimizing decision function. The experimental results show that the proposed method achieves high detection accuracy while saving labor costs and improving the automation of the detection methods.

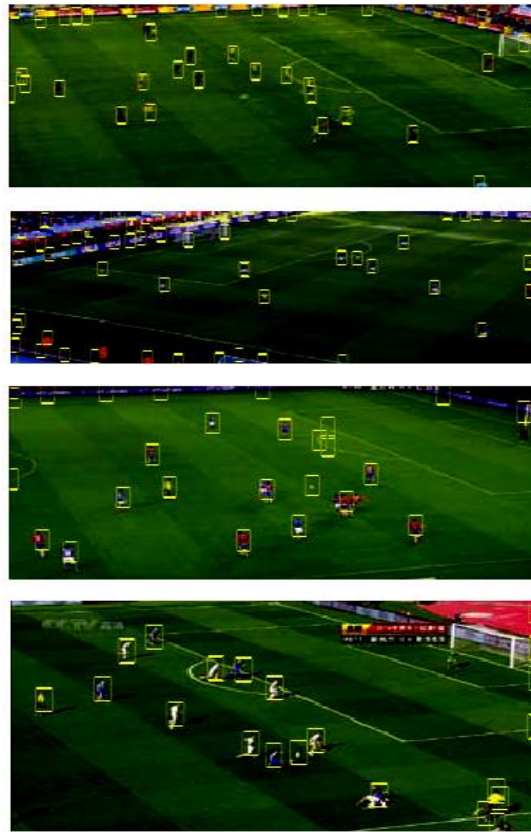


Figure 10. Detection results of Adaboost Player detection method

References

1. Qe Huang and Gen Zhu, "Trajectory Based Event Tactics Analysis in Broadcast Sports Video", *International Conference on Multimedia*. Santa Barbara, CA, USA: ACM, vol.4, pp.58–67, 2014.
2. Konieczny J, Kurc M and Mackowiak S, "A Complex System for Football Player Detection in Broadcasted Video", *International Conference on Signals and Electronic Systems. Gliwice, Poland: IEEE*, vol.6, pp.119–122, 2014.
3. Llach J and Bhagavathy S and Huang Y, "Players and Ball Detection in Soccer Videos Based on Color Segmentation and Shape Analysis", *International Workshop on Multimedia Content Analysis and Mining*, Weihai, China: Springer, vol.6, pp. 416–425, 2007.
4. Dermott J, Chun M and Kanwisher N, "The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception", *The Journal of Neuro-science*, vol.17, no.11, pp.4302–4311, 2007.
5. Miura K and Utsumi O, "An Object Detection Method for Describing Soccer Games from Video", *International Conference on Multimedia and Expo*. Lausanne, Switzerland: IEEE, vol.42, pp. 45–48, 2002.
6. Jin Liu and Xe Tong, "Automatic Player Detection, Labeling and Tracking in Broadcast Soccer Video", *Pattern Recognition Letters*, vol.30, no.2, pp.103 – 113, 2009.
7. Qiang Liu and Ye Zhang, "The Specificity of Category Information Processing in the Visual System", *Advances in Psychological Science*, vol.19, no.1, pp.42- 49, 2011.
8. Moya M and Hush R, "Network Constraints and Multi-Objective Optimization for One-class Classification", *Neural Networks*, vol.9, no.3, pp.463–474, 1996.
9. Kanwisher N, "What's in a Face", *Science Magazine*, vol.311, no.5761, pp.617–618, 2006.
10. Kanwisher N, "A Window into the Functional Architecture of the Mind", *Proceedings of the National Academy of Sciences of the United States of America*, vol.107, no.25, pp.11163–11170, 2010.
11. Huang Qu and Gi Zhu, "Event Tactic Analysis Based on Broadcast Sports Video", *IEEE Transactions on Multimedia*, vol.11, no.1, pp.49 –67, 2009.
12. Tong Xin and Jin Liu, "Automatic Player Labeling, Tracking and Field Registration and Trajectory Mapping in Broadcast Soccer video", *ACM Transactions on Intelligent Systems and Technology*, vol.2, no.2, pp.15–32,2013

Chengjun Cui received his B.S. degree in Physical Education from Yanbian University. He received his M.S degree and his Ph.D. in Measurement and Evaluation from Myongji University (Korea). He is a lecturer in the Department of Physical Education of Nanjing University of Science and Technology. His research interests include Measurement and Evaluation in Physical Education.