BRISK Local Descriptors for Heavily Occluded Ball Recognition

Pier Luigi Mazzeo, Paolo Spagnolo(✉), and Cosimo Distante

INO-Consiglio Nazionale delle Ricerche, Via della Liberta', 3,
73010 Arnesano, Lecce, Italy
paolo.spagnolo@cnr.it

Abstract. This paper focuses on the ball detection algorithm that analyzes candidate ball regions to detect the ball. Unfortunately, in the time of goal, the goal-posts (and sometimes also some players) partially occlude the ball or alter its appearance (due to their shadows cast on it). This often makes ineffective the traditional pattern recognition approaches and it forces the system to make the decision about the event based on estimates and not on the basis of the real ball position measurements. To overcome this drawback, this work compares different descriptors of the ball appearance, in particular it investigates on both different well known feature extraction approaches and the recent local descriptors BRISK in a soccer match context. This paper analyzes critical situations in which the ball is heavily occluded in order to measure robustness, accuracy and detection performances. The effectiveness of BRISK compared with other local descriptors is validated by a huge number of experiments on heavily occluded ball examples acquired under realistic conditions.

1 Introduction

In last decade sport video analysis has created great interest in the computer vision and multimedia technologies research communities [7]. This allowed to develop applications dealing with the analysis of different sports such as tennis, golf, American football, baseball, basketball, and hockey. Due to its worldwide viewership there has been an explosive growth in the research area of soccer video analysis [1], [6], and many different possible applications have been considered [4],[10],[12].

The main issue of any kind of the most important sport automatic systems is the detection of the ball; it is very difficult when images are taken from fixed or broadcast cameras with a large Field of View since the ball is represented by a small number of pixels and it can have different scales, textures and colors. Hence, automatic detection and localization of the ball in images is challenging as a great number of problems have to be managed: occlusions, shadowing, presence of very similar objects both near the field lines and on player's bodies (shoulders, legs, heads, ...), appearance modifications (for example when the ball is inside the goal it is faded by the net and it also experiences a significant amount of

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part I, LNCS 9279, pp. 172–182, 2015.
DOI: 10.1007/978-3-319-23231-7_16

deformation during collisions), unpredictable motion (for example when the ball is shot by players), and so on.

Most of the approaches proposed in literature use, at first, weak global information (size, color, shape) to detect the most likely ball regions and then motion hypotheses or template matching are introduced to filter candidates ([19],[22]. These approaches experience difficulties in ball candidate validation when many moving entities are simultaneously in the scene (the ball is not isolated) or when the ball abruptly changes its trajectory (for example in the case of rebounds or shots).

D'Orazio at al. [3] focus on ball pattern extraction and recognition: a neural architecture is proposed to discriminate the wavelet coefficients extracted from ball candidates and non-ball candidates previously selected by a modified version of the directional circular Hough transform (CHT). This approach fails to validate ball candidates in case of textured (invariance to rotation, scaling and illumination changes is limited) and occluded balls (wavelet description needs a large visible ball portion to supply significant coefficients). Moreover neural based classifier requires a rigorous selection of training examples and network parameters.

In [16] we have compared different local descriptors in order to face the ball recognition problem. Starting from these results we propose in this paper the novel local BRISK descriptors in order to deal with the heavily occluded balls.

Established leaders, in the field of the keypoints extractor, are the *SIFT* and *SURF* algorithms which exhibit great performance under a variety of image transformations. In [9] authors explored the use of scale-invariant feature transform (*SIFT*) to encode local information of the ball and to match it between ball instances. Experimental results point out some difficulties in detecting distinctive points on the ball and to make a decision by counting the number of correct point matches. Recently, Leutenegger [11] proposed a novel method for keypoints detection, description and matching called BRISK (Binary Robust Invariant Scalable Keypoints). BRISK, in contrast to well known algorithms with proven high performance, such as SIFT and SURF, offer a dramatically faster alternative at comparable matching performance. Considering these interesting peculiarities of BRISK, in this work we present a comparison of different features extraction approaches (Wavelet Transform - WT, Principal Component Analysis - PCA), and keypoints detectors and descriptors algorithms (SIFT and BRISK), in order to recognize soccer ball patterns.

The considered approaches were tested on a huge number of real ball images acquired in presence of translation, scaling, rotation, illumination changes, local geometric distortion, clutter, partial and heavy occlusion.

In the rest of the paper, section 2 gives a resume of the evaluation framework we use, including a short description of the implemented methodologies, whereas section 3 presents the experimental results, together with their comparison and discussion. Finally, in section 4, conclusions and future enhancement are drawn.

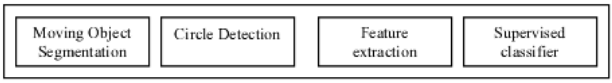


Fig. 1. A schematic diagram of the processing steps executed by each Node

2 Ball Recognition Evaluation

We have created an evaluation framework based on a GLT vision system which is able to detect the ball in an acquired soccer video sequences. The schematic diagram of whole processing path of vision system is shown in figure 1. The ball detection processing is composed by four main blocks: in the first one a background subtraction technique *[Moving Object Segmentation block]* is combined with a circle detection approach *[Circle detection block]* to extract ball candidate regions. Then a features extraction scheme is used to represent image patterns *[Feature extraction block]* and finally data classification is performed by using a supervised learning scheme *[Supervised Classifier block].*

2.1 Moving Object Segmentation

At the beginning of the image acquisition a background model has to be generated and later continuously updated to include lighting variations in the model. Then, a background subtraction algorithm distinguishes moving points from static ones. The implemented algorithm uses the mean and standard deviation to give a statistical model of the background. Detection is then performed by comparing the pixel current intensity value with its statistical parameters, as explained in several works on this topic (a good review can be found in [18]). Details about the implemented approach can be found in [20]. Finally, after the detection of moving pixels, a connected components analysis detects the blobs in the image by grouping neighboring pixels. After this step, regions with an area less than a given threshold are considered as noise and removed, whilst remaining regions are evaluated in the following blocks.

2.2 Circle Detection

The Circle Hough Transform (CHT) aims to find circular patterns of a given radius R within an image. Each edge point contributes a circle of radius R to an output accumulator space. The peak in the output accumulator space is detected where these contributed circles overlap at the center of the original circle. In order to reduce the computational burden and the number of false positives typical of the CHT, a number of modifications have been widely implemented in the last decade. The use of edge orientation information limits the possible positions of the center for each edge point. This way only an arc perpendicular to the edge orientation at a distance R from the edge point needs to be plotted. The CHT and also its modifications can be formulated as convolutions applied to an edge

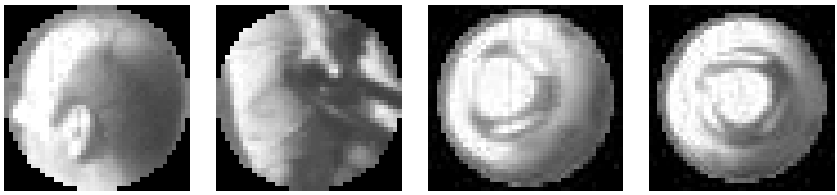


Fig. 2. Images of the training set: negative (first two) and positive examples of the ball

magnitude image (after a suitable edge detection). More details on this approach can be found in [16].

2.3 Feature Extraction

In this block the candidate ball regions are processed by different feature extraction methodologies in order to represent them only by coefficients containing the most discriminant information. A secondary aim is also to characterize the images with a small number of features in order to gain in computational time. Object recognition by using a learning-from-examples technique is in fact related to computational issues. In order to achieve real time performances the computational time to classify patterns should be small. The main parameter connected to high computational complexity is certainly the input space dimension. A reduction of the input size is the first step to successfully speed up the classification process. This requirement can be satisfied by using a feature extraction algorithm able to store all the important information about input patterns in a small set of coefficients. Wavelet Transform (WT), Principal Component Analysis (PCA), Scale Invariant Feature Transform (SIFT), Binary Robust Invariant Scalable Keypoints (BRISK) are different approaches allowing to significantly reduce the dimension of the input space, because they capture the significant variations of input patterns in a smaller number of coefficients. In the following four subsections we briefly review WT, PCA, SIFT, and BRISK approaches.

WT: Wavelet Transform. The WT is an extension of the Fourier transform that contains both frequency and spatial information [15]. Numerous filters can be used to implement WT: we have chosen Haar and Daubechies filters for their simplicity and orthogonality.

PCA: Principal Component Analysis. Principal component analysis (PCA) provides an efficient method to reduce the number of features to work with [8]. It transforms the original set of (possibly) correlated features into a small set of uncorrelated ones. In particular, PCA determines an orthogonal basis for a set of images involving an eigen analysis of their covariance matrix.

SIFT: Scale Invariant Feature Transform. The scale Invariant Feature Transform is a method for extracting distinctive invariant features from the images that can be used to perform reliable matching between different views of an object or a scene [13]. The features are invariant to image scale and rotation

and they provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination. The features are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many different images.

BRISK: Binary Robust Invariant Scalable Keypoints. It is a novel method for keypoints detection, description and matching [11], based on an application of a novel scale-space FAST-based detector in combination with the assembly of a bit-string descriptor from intensity comparison obtained by dedicated sampling of each keypoint neighborhood. BRISK uses a easily configurable circular sampling pattern from which it computes brightness comparisons to form a binary descriptor string. This detector can be useful for a wide spectrum of applications, in particular for tasks with hard real-time constraints or limited computational resources. In figure 3 are highlighted some matches among BRISK descriptors using the Hamming distance applied to the two balls extracted by the CHT showed in the second row of the figure 2.

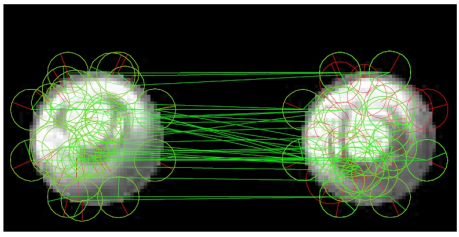


Fig. 3. BRISK descriptor matching

2.4 Supervised Classifier

The following step in the proposed framework aims at introducing an automatic method to distinguish between ball and no-ball instances on the basis of the feature vector extracted by one of the previous mentioned pre-processing strategies. To accomplish this task a probabilistic approach has been used. Probabilistic methods for pattern classification are very common in literature as reported by [21]. So-called *naive* Bayesian classification is the optimal method of supervised learning if the values of the attributes of an example are independent given the class of the example. Although this assumption is almost always violated in practice, recent works have shown that naive Bayesian learning is remarkably effective in practice and difficult to improve upon systematically. On many real-world example datasets naive Bayesian learning gives better test set accuracy than any other known method [2], [17]. In general a Naive Bayes classifier is also preferable for its computational efficiency.

Probabilistic approaches to classification typically involve modelling the conditional probability distribution $P(C|D)$, where C ranges over classes and D over

descriptions, in some language, of objects to be classified. Given a description d of a particular object, the class $argmaxcP(C = c|D = d)$ is assigned. A Bayesian approach splits this posterior distribution into a prior distribution $P(C)$ and a likelihood $P(D|C)$:

$$argmax_c P(C = c|D = d) = argmax_c \frac{(P(D = d|C = c)P(C = c))}{P(D = d)} \quad (1)$$

The key term in Equation1 is $P(D = d|C = c)$, the likelihood of the given description given the class (often abbreviated to $P(d|c)$). A Bayesian classifier estimates these likelihoods from training data. If the assumption that all attributes are independent given the class:

$$P(A_1 = a_1, ..., A_n = a_n|C = c) = \prod_{i=1}^{n} P(A_i = a_i|C = c) \quad (2)$$

then a Naive Bayesian Classifier (often abbreviated to *Naive Bayes*) is introduced. This means that a Naive Bayes classifier ignores interactions between attributes within individuals of the same class. We use a Gaussian Kernel as Naive Bayes likelihood with the attribute space size depending of the dimensions of the feature vectors.

Further details and discussions about the practical consequences of this assumption can be found in [5].

2.5 Experimental Setup

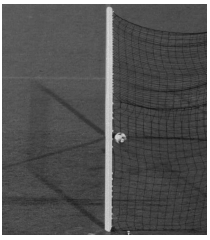
Here we, briefly, describe how many features were extracted from each of the different pre-processing techniques explained in 2.3. For the SIFT and BRISK methodologies the codebook of visual words (known as Bag of word representation) was built by quantizing (using K-means algorithm [14]) the 128-long feature vectors relative to the detected points of interests in the patches containing fully visible balls under good light conditions. The pre-processing strategies reduce the number of the coefficients that are the input of the Bayesian classifier. For the SIFT and BRISK local descriptors we used 20 coefficients that were the results of the k-means clustering algorithms. This because after many experiments we found that the descriptors length, with the best ball detection rate, was 20. For the Wavelet 'HAAR' and 'Daubechies3' methods we used the approximation coefficient at the second level of decomposition; this reduces the input coefficient to a vector of 64 (8x8) elements for 'HAAR' and to a vector of 121 (11x11) elements for 'Daubechies3'. In the PCA strategy we used the Single Value Decomposition (SVD) technique to solve the eigenstructure decomposition problem; the covariance matrix was evaluated on the entire set of training images, by obtaining 30 coefficients.

3 Experimental Results

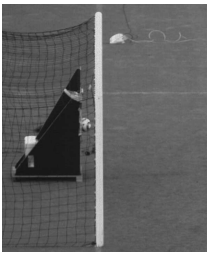
Experiments were carried out on image sequences acquired in a real soccer stadium by a *Mikroton EOSens MC1362 CCD camera*, equipped with a 135

mm focal length. The camera has the area around the goal in its field of view. Using this experimental setup the whole image size was 1280x1024 pixels whereas the ball corresponded to a circular region with radii in the range $[R\_MIN = 21, R\_MAX = 24]$ depending on the distance of the ball with respect the optical center of the camera. The camera frame rate was 200 fps with an exposure time of 1 msec in order to avoid blurring effect in the case of very high ball speed.

During the data acquisition session a number of shots on goal were performed: the shots differed in ball velocity and direction with respect to the goalmouth. This way, differently scaled (depending on the distance between the ball and the camera), occluded (by the goal posts) and faded (by the net) ball appearances were experienced. Moreover some shots hit a rigid wall placed on purpose inside the goal in order to acquire some instances of deformed ball and then to check the sensitiveness of the ball recognition system in the case of ball deformation. Figure 4 reports images acquired during the experimental phase and corresponding to two goal events; the first one refers to a goal event during the Free Shot session: the fading effect due to the presence of the Net is evident, but no occlusions and deformations are present. The second one was performed in presence of an Impact Wall: in this experiments, the ball is faded by the net, occluded by the goal posts, and deformed by the impact with the rigid wall.



Free shot - fading



Impact wall - deformation

Fig. 4. Some images acquired during experimental phase

During the acquisition session about *5,6M* of images where collected. Each image was processed to detect candidate ball regions by finding circular arrangement in the edge magnitude map: edge point contributes a circle to an output accumulator space and the peak in the output accumulator space is detected where these contributed circles overlap at the center of the original circle.

The main benefits of this method are the low computational cost and the high detection rate when the ball is visible in the image. The main drawback lies in the impossibility of detecting occlusions and ball absence in the image. The circle detection algorithm determines in every situation the highest peaks in the accumulator space. Then it is difficult to differentiate between images containing and not containing the ball. To reduce negative effects due to occlusions, we introduce BRISK feature descriptors and compare them with the other ones.

For each image a candidate ball region (of size $(2*R\_MAX+1)x(2*R\_MAX+1)$ i.e 42x42) was then extracted around the position corresponding to the highest value in the accumulator space.

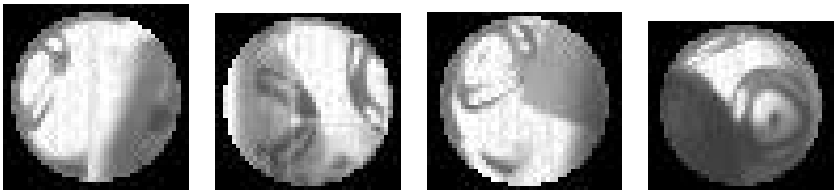


Fig. 5. Examples of occluded balls, respectively due to goal post, net, player's arm and head

After this preliminary step 3975 candidate ball regions were selected and manually labelled to form a test ground truth. This dataset has been built with the goal of provide about all possible appearances for the ball. This way, it contains balls acquired in both good and bad lighting conditions, in presence of both partial and heavy occlusions, and also in presence of cast shadows that strongly modify the ball appearance. In the acquired images these challenging conditions (occlusions and shadows) occurred either while the ball crosses the goal mouth appearing behind the goal posts or when a person interposes himself between the ball and the camera (some examples of occluded and shadowed balls are shown in figure 5).

Table 1. Ball Recognition results on the whole data set

Ball Descriptors	TP	FN	TN	FP	Detection Rate (%)
Wavelet HAAR	1890	95	1654	336	89.16%
Wavelet DB3	1870	117	1700	287	89.81%
BRISK	1027	930	1749	269	69.83%
SIFT	992	992	1702	289	67.77%
PCA	1595	383	1430	567	76.10%

Table 1 summarizes the ball recognition results obtained by comparing different feature descriptors on the whole data set. In the first column of the table 1 are itemized the used feature descriptors. From the second to fifth columns are presented respectively: the values of correct detection of the ball (TP:True Positive), the values of the errors in the ball detection (FN: False Negative), the values of correct detection of no-ball (TN: True Negative) and finally the values of errors in detection of ball in the candidate regions in which it does not exist (FP: False Positive). In the last column of the table 1 the overall detection performance of each methodology is shown. For all of the compared feature extraction methodologies we have used a Naive Bayes Classifier with the same training set

composed of 50 examples of fully visible balls, and 50 examples which did not contain the ball. All the training patches are extracted at the beginning of the acquisition session, with good lighting conditions. The proposed results are very encouraging: the detection rate is always very high, with the best results obtained by using Wavelet decomposition (in particular Daubechies family slightly better than Haar family). On the other hand, the detection rate of BRISK is good but not comparable with Wavelet. In this test, all training examples consisted only of patches extracted from images acquired under good light conditions, while the test images contained patches extracted from images acquired in different and challenging conditions (poor light condition, ball occluded or deformed). However, the selected descriptors were able to well characterize the ball patches, even in critical conditions. At the same time no-ball patches were well classified, avoiding a huge number of false ball validation.

Table 2. Occluded balls recognition results

Ball Descriptors	TP	FN	Detection Rate (%)
Wavelet HAAR	114	86	114/200 (57.00%)
Wavelet DB3	90	110	90/200 (45.00%)
BRISK	157	43	157/200 (78.50%)
SIFT	148	52	148/200 (74.00%)
PCA	69	131	69/200 (34.50%)

In the second experiment, we focus our attention on a subset of the whole ground truth dataset, specifically the subset composed by partially and heavily occluded balls. The final goal of our system, as remarked in section 1, is the detection of goal events during football matches. So, it is reasonable to highlight performance of the ball detection algorithms in presence of occlusions: a goal event surely happens very close to the goalposts, probably in presence of a goalkeeper, so, before crossing the line, the ball will be likely occluded by goalposts or player's body part. For these reasons we have analyzed performance of feature descriptors on a test subset consisting of *200* examples (only partially/heavy occluded balls). In table 2 the obtained results are outlined. As evident, SIFT and BRISK methodologies give the best results in terms of ball recognition rate in these critical situations. This can be explained by considering the characteristics of these features: they are local descriptors (differently from Wavelet and PCA that are global), so they are able to better represent local (both geometrical and textural) features of the balls.

In addition, considering the computational aspects, it results that BRISK is faster than SIFT. So, we can conclude that, in very critical situations, BRISK and SIFT outperform other approaches in terms of recognition rate, and computational issues encouraging the use of BRISK for the detection of the ball finalized to the goal event detection in real football contexts.

The results obtained in the experiments reported in section 3 deserve a thorough analysis because they open up new ways to further improve the system

aimed at automatically detecting goal events by using video inputs. From Table 1 it is clear that the traditional methods, which consider global features, are essential to recognize the ball in cases where it is clearly visible as well as free of external disturbances, such as shadows which substantially alter its appearance. In contrast, methods based on local descriptors are less reliable in case of coherent appearance but are much less sensitive to occlusions and to changes in lighting. That said, the results presented in this work, although preliminary and worthy of further investigation, allow us to see the possibility of introducing a substantial improvement to the traditional approaches used to recognize the ball in football images. In particular, they suggest to move from a single-step approach to a multi-step one in which different ball recognition strategies are used depending on the operative status of the system. Under this new point of view, global approaches like wavelet could be used to alert the system to the presence of the ball in the scene (since they have a higher recognition rate for not corrupted ball occurrences). Once the ball has been detected in the scene it can be tracked and, frame by frame, a double validation could be performed by using both global and local descriptors, especially when the ball is close to the goal post. This way, the 3D ball position can be retrieved on the basis of a more robust classification in each processing unit (a lower number of false negative occurrences is expected). This should increase the precision in 3D ball estimation and then in the number of correctly evaluated controversial situations about goal events.

4 Conclusion and Future Work

This paper investigated different feature descriptors to be used, as a part of a more complex system, in order to recognize football ball patterns. Experimental results on real ball examples under challenging conditions and a comparison with some of the more consolidated feature extraction methodologies demonstrated the suitability of the local descriptor (like SIFT and BRISK) to recognize ball pattern even in presence of occlusions of strong changes in appearance due to cast shadows. Future works will address computational aspects of the use of multi-step ball recognition strategies and an evaluation of the whole system after the integration of this new processing strategies.

References

1. Choi, K., Seo, Y.: Automatic initialization for 3d soccer player tracking. Pattern Recogn. Lett. **32**(9), 1274–1282 (2011)
2. Colas, F., Brazdil, P.: Comparison of SVM and some older classification algorithms in text classification tasks. In: Bramer, M. (ed.) Artificial Intelligence in Theory and Practice. IFIP AICT, vol. 217, pp. 169–178. Springer, Heidelberg (2006)
3. D'Orazio, T., Guaragnella, C., Leo, M., Distante, A.: A new algorithm for ball recognition using circle hough transform and neural classifier. Pattern Recognition **37**(3), 393–408 (2004)

4. Ekin, A., Tekalp, A.M.: Automatic soccer video analysis and summarization. IEEE Trans. on Image Processing **12**, 796–807 (2003)
5. Flach, P.A., Lachiche, N.: Naive bayesian classification of structured data. Machine Learning **57**(3), 233–269 (2004)
6. Gao, X., Niu, Z., Tao, D., Li, X.: Non-goal scene analysis for soccer video. Neurocomputing **74**(4), 540–548 (2011)
7. Hung, M.-H., Hsieh, C.-H., Kuo, C.-M., Pan, J.-S.: Generalized playfield segmentation of sport videos using color features. Patt. Recogn. Lett. **32**(7), 987–1000 (2011)
8. Jolliffe, I.T.: Principal Component Analysis. Springer (2002)
9. Leo, M., D'Orazio, T., Spagnolo, P., Mazzeo, P.L., Distante, A.: SIFT based ball recognition in soccer images. In: Elmoataz, A., Lezoray, O., Nouboud, F., Mammass, D. (eds.) ICISP 2008. LNCS, vol. 5099, pp. 263–272. Springer, Heidelberg (2008)
10. Leo, M., Mosca, N., Spagnolo, P., Mazzeo, P., D'Orazio, T., Distante, A.: A visual framework for interaction detection in soccer matches. International Journal of Pattern Recognition and Artificial Intelligence **24**(04), 499–530 (2010)
11. Leutenegger, S., Chli, M., Siegwart, R.Y.: Brisk: binary robust invariant scalable keypoints. In: 2011 IEEE International Conference on Proceedings of Computer Vision (ICCV), ICCV 2011, pp. 2548–2555 (2011)
12. Liu, J., Tong, X., Li, W., Wang, T., Zhang, Y., Wang, H.: Automatic player detection, labeling and tracking in broadcast soccer video. Pattern Recognition Letters **30**(2), 103–113 (2009)
13. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004)
14. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: Proc. of the 5th Berkeley Symposium on Mathem. Stat. and Prob., vol. 1, pp. 281–297. Univ. of Calif. Press (1967)
15. Mallat, S.: A Wavelet Tour of Signal Processing. AP Professional, London (1997)
16. Mazzeo, P.L., Leo, M., Spagnolo, P., Nitti, M.: Soccer ball detection by comparing different feature extraction methodologies. Adv. in Art. Int. (2012)
17. Petrović, N.I., Jovanov, L., Pižurica, A., Philips, W.: Object tracking using naive bayesian classifiers. In: Blanc-Talon, J., Bourennane, S., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2008. LNCS, vol. 5259, pp. 775–784. Springer, Heidelberg (2008)
18. Piccardi, M.: Background subtraction techniques: a review. In: IEEE SMC 2004 International Conference on Systems, Man and Cybernetics (2004)
19. Ren, J., Orwell, J., Jones, G.A., Xu, M.: Tracking the soccer ball using multiple fixed cameras. Computer Vision and Image Understanding **113**(5), 633–642 (2009)
20. Spagnolo, P., Mosca, N., Nitti, M., Distante, A.: An unsupervised approach for segmentation and clustering of soccer players. In: IMVIP Conference, pp. 133–142 (2007)
21. Tosic, I., Frossard, P.: Dictionary learning: What is the right representation for my signal? IEEE Signal Processing Magazine **28**(2), 27–38 (2011)
22. Yu, X., Leong, H., Xu, C., Tian, Q.: Trajectory-based ball detection and tracking in broadcast soccervideo. IEEE Transaction on Multimedia **8**(6), 1164–1178 (2006)