SIFT Based Ball Recognition in Soccer Images

Marco Leo, Tiziana D'Orazio, Paolo Spagnolo, Pier Luigi Mazzeo, and Arcangelo Distante

Institute of Intelligent Systems for Automation
via Amendola 122/D 70126 Bari (Italy)
{leo,dorazio,spagnolo,mazzeo,distante}@ba.issia.cnr.it

Abstract. In this paper a new method for ball recognition in soccer images is proposed. It combines Circular Hough Transform and Scale Invariant Feature Transform to recognize the ball in each acquired frame. The method is invariant to image scale, rotation, affine distortion, noise and changes in illumination. Compared with classical supervised approaches, it is not necessary to build different positive training sets to properly manage the great variance in ball appearances. Moreover, it does not require the construction of negative training sets that, in a context as soccer matches where many no-ball examples can be found, it can be a tedious and long work. The proposed approach has been tested on a number of image sequences acquired during real matches of the Italian Soccer "Serie A" championship. Experimental results demonstrate a satisfactory capability of the proposed approach to recognize the ball.

1 Introduction

During soccer matches a number of doubtful cases occur, especially for detecting the offside or the goal events. An automatic method that detects in each image of the sequence the ball position is the first and the most important step to build a (non invasive) vision based real time decision support tool for the referee committee.

In the last decade different methods to automatically recognize the ball in soccer game have been proposed. They could be conveniently divided in two categories: direct methods and indirect methods. Indirect methods do not search the ball in each frame but they distinguish the ball from other moving objects by means of a priori knowledge about its motion. In [1] a strategy based on no-ball candidate elimination is applied using a coarse-to-fine process. A 'condensation' algorithm is utilized in ball tracking and a confidence measure representing the ball region's reliability is presented to guide possible ball re-detection for continuous tracking. The ball recognition task is achieved in [2] by a trajectory verification procedure based on Kalman filter rather than the low-level feature. Two different procedures run iteratively (trajectory discrimination and extension) to identify the ball trajectory. Soccer ball estimation and tracking using trajectory modeling from multiple image sequences is proposed in [3]. Indirect methods seem to be well suited for video indexing applications but not for real time event detection (as, for example, to solve goal line crossing problem) since they are not able to asses the ball position in each frame.

Direct methods overcome this drawback since they recognize the ball in each frame by using appearance information as shape, size and color. Direct methods classify the pattern images after a suitable pre-processing and they make use of

A. Elmoataz et al. (Eds.): ICISP 2008, LNCS 5099, pp. 263–272, 2008.

computational paradigms generally used in many other contexts as face recognition, people detection, character recognition [6-10]. In [14-15] supervised classification techniques based on support vector machine and neural network have been applied in the soccer context on both textured ball and non-textured ball. In [4-5] Wavelet Transform and Independent Component Analysis are used to obtain a suitable representation of the ball to give as input to a neural network.

In this paper a new direct ball recognition approach is proposed. It consists of two steps: first of all the Circle Hough Transform (CHT) [12] is applied on the moving objects in the scene to find the region having the most circular shape inside a given radius; then, the Scale Invariant Feature Transform (SIFT) [11] is applied on the selected region to verify, by analyzing its appearance, if it really contains the ball.

The CHT has been already used for ball recognition purposes [4,5,13] to select regions with circular shape that are candidate to contain the ball. However this step is just preliminary for the actual ball recognition task. Indeed the CHT always produces a maximum value on a region with a circular shape, whatever object is contained inside. Nor the usage of threshold can be considered to select the ball since many objects with circular shape produces high values of the CHT. A further step is necessary to carry out the actual ball recognition process.

The main contribution of this paper is the use of Scale Invariant Feature Transform to validate selected regions. Previous approaches made use of supervised algorithms that required long and tedious learning procedure based on multiple positive training sets (selected manually trying to cover quite all the possible appearances of the ball) to assure acceptable ball recognition performance. Moreover, these methods required a set of negative examples (no-ball examples) to guarantee a more reliable learning phase: unfortunately, the selection of no-ball examples is generally not a trivial task considering that regions with a quite circular shape such as player's socks, pants or shirts, advertising posters, are very common in the soccer images.

The SIFT application allows, in fact, to overcome these drawbacks avoiding the difficulties to build generalized models of the ball and no-ball instances. The proposed approach, taking advantage of the SIFT property of being invariant to image scale, rotation, affine distortion, addition of noise and changes in illumination allows to easily build only one appearance model of the "ball concept" using a small set of ball examples.

In this paper we present a large number of experiments that were carried out on real image sequences acquired during the Italian Soccer "Serie A" championship. We demonstrated that satisfactory ball recognition results can be obtained using only few positive training images.

The rest of the paper is organized as follows: section 2 gives an overview of the proposed method; section 3 describes the experimental setup ; section 4 reports the results of our experiments on real images. Finally, discussion and conclusions are reported in section 5.

2 System Overview

The proposed method consists of two steps (see fig. 1): first of all the Circle Hough Transform (implemented as convolutions on the edge magnitude image) is applied on

the moving regions to select the area that best fits the sought pattern. The region producing the highest peak in the CHT accumulation space is selected as the one having the best circular shape (see [4] for major details).

Then, in the second step a validation procedure is used to verify, by analyzing the appearance, if the selected region really contains the ball. This validation procedure is necessary since the CHT always produces a maximum value in the image, independently from the presence or not of the ball inside the image. The validation procedure proposed in this paper consists in using the SIFT algorithm proposed by Lowe in [11] that extracts distinctive invariant features (the SIFT *keypoints*) by a four stages filtering approach. During an initialization phase a small set of reference ball images is selected and the corresponding SIFT keypoints are stored in a database. This step is carried out just at the beginning during the calibration of the experimental setup and it remains valid for all the experiments.

Then, the SIFT keypoints are evaluated on the region selected by the CHT preprocessing step. These features are compared, by a proper matching procedure (the *keypoint matching procedure* is well described in [11]), to each set of features stored in the database. The final decision about Ball/Non-Ball occurrence in the region is done on the basis of the average number *N* of correct key-point matches on the stored sets. If *N* is greater than a fixed threshold *Th,* experimentally set, the testing image is classified by the system as containing the ball otherwise it is discarded.

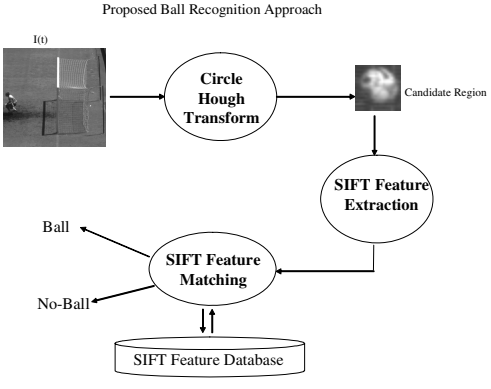


Fig. 1. The Ball recognition system. I(t) is the frame acquired at the time t.

2.1 Scale Invariant Feature Transform (SIFT)

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

The algorithm consists of four main steps:

1. Scale-space extrema detection;
2. Keypoints localization;
3. Orientation assignment;
4. Keypoint description.

The first stage identifies locations and scales that can be repeatably assigned under differing views of the same object. Detecting locations that are invariant to scale change of the image can be accomplished by searching for stable features across all possible scales, using a continuous function of scale known as scale space.

Under a variety of reasonable assumptions the only possible scale-space kernel is the Gaussian function. Therefore, the scale space of an image is defined as a function, $L(x,y,\sigma)$, that is produced from the convolution of a variable-scale Gaussian, $G(x,y,\sigma)$, with an input image, $I(x; y)$ i.e.:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

where * is the convolution operation and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} .$$

The keypoints are detected using scale-space extrema in the difference-of-Gaussian function D convolved with the image $I(x; y)$:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$

where k is the constant multiplicative factor which separates two nearby scales. In order to detect the local maxima and minima of $D(x;y;\sigma)$, each sample point is compared to its eight neighbors in the current image and to its nine neighbors in the scale above and below. It is selected only if it is larger than all of these neighbours or smaller than all of them.

Once a keypoint candidate has been found by comparing a pixel to its neighbours, the next step is to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. This information allows points to be rejected if they have low contrast (and are therefore sensitive to noise) or are poorly localized along an edge.
A 3D quadratic function is fitted to the local sample points. The approach starts with the Taylor expansion (up to the quadratic terms) with sample point as the origin

$$D(\mathrm{X}) = D + \frac{\partial D^T}{\partial \mathrm{X}} \mathrm{X} + \frac{1}{2} \mathrm{X}^T \frac{\partial^2 D}{\partial \mathrm{X}^2} \mathrm{X}$$

where D and its derivatives are evaluated at the sample point $\mathrm{X}=(x,y,\sigma)^{\mathrm{T}}$. The location of the extremum is obtained taking the derivative with respect to X, and setting it to 0, giving

$$\hat{\mathrm{X}} = - \frac{\partial^2 D}{\partial \mathrm{X}^2}^{-1} \frac{\partial D}{\partial \mathrm{X}}$$

that is a 3x3 linear system, easily solvable.
The function value at the extremum

$$D(\hat{\mathrm{X}}) = D + \frac{1}{2} \frac{\partial D^T}{\partial \mathrm{X}} \hat{X}$$

is useful for rejecting unstable extrema with low contrast.

At this point the algorithm rejects also keypoints with poorly defined peaks i.e those points having, in the difference-of-Gaussian function a large principal curvature across the edge but a small one in the perpendicular direction.

By assigning a consistent orientation, based on local image properties, the keypoint descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation.

For each image sample, $L(x; y)$, at this scale, the gradient magnitude, $m(x; y)$, and orientation, $\theta(x;y)$, is pre-computed using pixel differences:

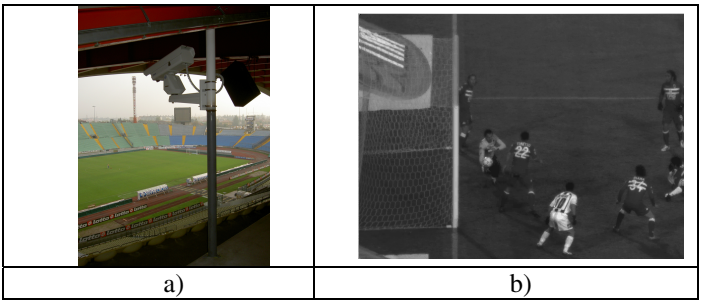
$$m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2}$$
$$\theta(x,y)=a\tan 2((L(x,y+1)-L(x,y-1))/(L(x+1,y)-L(x-1,y)))$$

An orientation histogram is formed from the gradient orientations of sample points within a region around the keypoint.

Finally a keypoint description is created by first computing the gradient magnitude and orientation at each image sample point in a region around the keypoint location. These sample points are weighted by a Gaussian window and then accumulated into a 8 bins orientation histograms summarizing the contents over 4x4 subregions, with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. The final descriptor size is 4x4x8=128 and a normalization procedure is applied to the descriptor to allow invariance to illumination change.

3 Experimental Setup

Experiments were performed on real image sequences acquired at the Friuli Stadium in Udine (Italy) during different matches of the "Serie A" Italian Soccer Championship 2006/2007. Images were acquired using a DALSA TM.6740 monochrome cameras able to record up to 200 frames/sec with a resolution of 640x480 pixels. The cameras were placed on the stands of the stadium with theirs optical axis lying on goal-mouth plane (see fig. 2). Different matches in different



a) b)

Fig. 2. a) The DALSA TM.6740 camera placed on the stands of the stadium and used to acquire the image sequence during real soccer matches. In figure it is protected by an enclosure. b) an image acquired by the camera during a match.

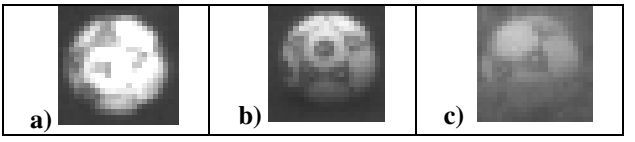


Fig. 3. Three different ball appearances. a) The ball in a sunny day. b) The ball during an evening match. c) The Ball in the goal post. In this case the grid of the goal post is between the camera and the ball.

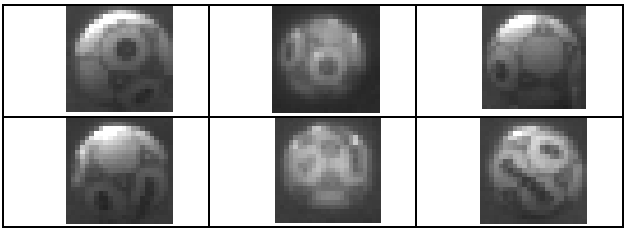


Fig. 4. 6 of the 17 training images used in the experimental phase

lighting conditions (evening matches with artificial lights, afternoon matches in both cloudy and sunny days) were acquired. The acquired images demonstrated the great variance in the appearance of the ball depending on lighting conditions, ball speed, ball position etc. In figure 3 three different ball images are shown.

In the acquired images, the ball radius varied from 9 to 11 pixels (depending on the distance from the camera) so two convolution masks of dimension 23x23 pixels were used to perform Circle Hough Transform and, consequently, a candidate region having size 23x23 pixels was given as input to the validation step based on Scale Invariant Feature Transform.

In the calibration phase, a training set was generated consisting in 17 ball examples acquired during an evening match and chosen to make the system effective also when the ball texture changed under different views. If a uniformly textured ball was used, the training set would be reduced to just one image. The training set had not to take into account possible scale variance of the ball, light condition changes and eventual noise addition. These images were processed to extract the SIFT keypoint features that were saved in a database for further analysis.

In figure 4, some of the 17 training images are reported.

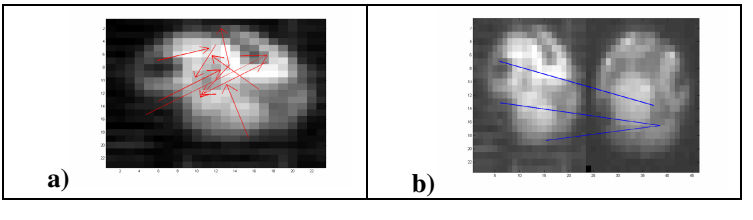


Fig. 5. a) The keypoints localized on a ball image. b) 3 keyponts matching between a test patch and a training patch.

The regions, selected by the CHT step, were classified as a ball instance by the validation step if the average number of key-points matching with the training set keypoints was greater than 3.

In figure 5a) the keypoints detected on a ball image are represented by red arrows; in figure 5b) some keypoints correctly matched between a test image (on the left) and a training image (on the right) are connected by blue lines.

4 Experimental Results

In this paper two sets of experiments were performed: in the first set, the proposed ball recognition approach was evaluated on some sequences acquired during evening matches (with the stadium artificial lights switched on); in the second set of experiments the images were acquired with natural light conditions (both cloudy and sunny conditions).

From the evening matches a set of 3560 images was selected; 1945 of these images contained the ball together with some players and the goal-keeper; the remaining 1615 did not contain the ball but only some players and the goal-keeper.

All these images were firstly processed by the CHT algorithm that, for each of them, selected the region having most circular shape: 1921 of the selected regions really contained the ball, whereas the remaining 1639 did not.

Figure 6 reports some regions selected by CHT algorithm. The two regions on the left contain two ball examples, the two central images contain two examples of partially occluded ball and finally the two images on the right contain two no-ball examples as the shoulder of the goal-keeper and of the shoe of a player. All the regions selected by the CHT algorithm (1921 ball images and 1639 no-ball images), were provided as input to the SIFT based classifier for the validation step that had the main task of separating ball instances from no ball instances.

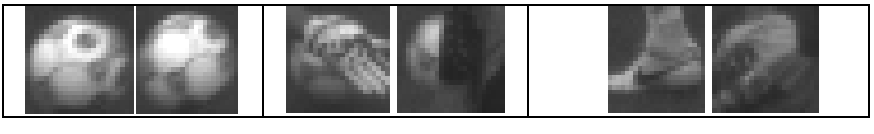


Fig. 6. Some regions selected by CHT algorithm: two examples of whole ball (left column); two examples of partially occluded ball (central column) and two no-ball examples (right column)

In Table 1 the scatter matrix of the results obtained using SIFT based classifier in the first set of experiment is reported. More than 90% of candidate regions containing the ball were correctly validated. At the same time, almost 89% of candidate regions that did not contain the ball were correctly discarded. In figure 7, a case of wrong ball validation is shown. Some keypoints relative to the texture of a player's shoe erroneously matched some keypoints of the ball training images and the proposed approach failed.

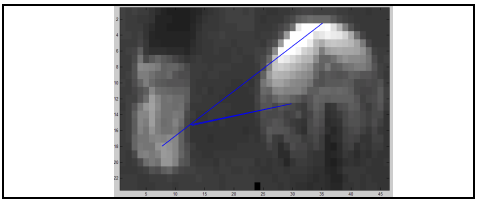


Fig. 7. An example of wrong validation of selected region: the texture of the player's shoe matched some areas of the ball

Table 1. The scatter matrix reporting the performance of the proposed SIFT based classifier in the set of experiments on images acquired during evening matches

		System Results	
		Ball	No Ball
Ground Truth	Ball (1921)	**90.3 % (1734/1921)**	**9.7% (187/1921)**
	No-Ball (1639)	**10.93% (179/1639)**	**89.07% (1460/1639)**

It should be noted that the CHT algorithm was not able to select the ball in 24 images over the total of 1945 containing the ball. This means that the ball contours were not so clear to produce the maximum value in the CHT accumulation space. In these cases the SIFT classifier is not able to recover the ball, but it can only confirm the absence of the ball in the region erroneously selected by the CHT preprocessing step.

In the second set of experiments, 2147 images were acquired with natural light conditions; 1034 of them contained the ball together with some players and the goal-keeper and the remaining 1113 did not contain the ball but only some players and the goal-keeper.

As in the first experiment these images were processed by the CHT algorithm that for each of them selected the region having the most circular shape: the result was that 1005 of the selected regions really contained the ball, whereas the remaining 1142 did not. Also in this case 29 regions containing the ball were lost by the CHT algorithm that provided maximum values on regions having a more circular appearance than the ball. A careful analysis of these images demonstrated that the ball was not so clear since the images were unfocused (for high speed shot) or the ball was behind the net of the door.

Table 2. The scatter matrix reporting the performance of the proposed SIFT based classifier in the set of experiments on images acquired with natural light conditions

		System Results	
		Ball	No Ball
Ground Truth	Ball (1005)	**83.08 % (835/1005)**	**16.91% (170/1005)**
	No-Ball (1142)	**10.07% (115/1142)**	**89.92% (1027/1142)**

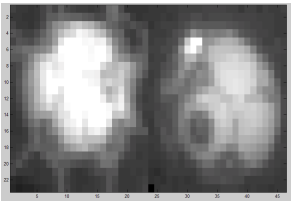


Fig. 8. One example in which the of misrecognized the ball due to reflection effects on the ball surface

In Table 2 the scatter matrix reports the performance results of the SIFT validation step. The presence of self shadows and the saturation of some areas of the ball appearance due to the sun reflection reduced the ball recognition performance with respect to the table 1.

In figure 8 one example of misrecognition of the ball due to the ball saturation effect is shown. The keypoints of the testing image (on the left) did not match with those of the ball in the training set (on the right).

However the general performances remain satisfactory (more than 80% for true positive and true negative detection) also considering that the set of training examples did not contain ball images acquired with natural light conditions. This is, in our opinion, a very pleasant result that makes the proposed approach preferable to those who requires careful selection of training sets.

5 Discussion and Conclusions

In this paper a new method for ball recognition in soccer images is proposed. It combines Circular Hough Transform and Scale Invariant Feature Transform to recognize the ball in each acquired frame. The method is invariant to image scale, rotation, affine distortion, noise, and changes in illumination. It has the main advantage compared with classical supervised approaches, of not requiring different positive training sets to properly manage the great variance in ball appearances. Moreover, it does not require the construction of negative training sets that, in a context as soccer matches where many no-ball examples can be found, it can be a tedious and long work.

Experimental results executed on real images acquired both with natural and artificial lighting conditions, demonstrated the capability of the proposed approach to recognize ball instances. In the reported experiment only one set of 17 ball training images, all acquired during an evening match, was used to performs ball recognition in any lighting condition. This is a very pleasant characteristic for a ball recognition system considering that the ball appearance can greatly change and it is practically impossible to build a generalized ball model.

In conclusion, the proposed approach seems to be a proper trade off between performance, portability and easiness to start up. Future work will be addressed to improve classification performance both using new vision tools able to avoid saturation effect on ball surface in sunny days and introducing more robust strategies to reduce false validations of the ball.

References

[1] Tong, X.-F., Lu, H.-Q., Liu, Q.-S.: An effective and fast soccer ball detection and tracking method, Pattern Recognition, 2004. In: ICPR 2004. Proceedings of the 17th International Conference, August 23-26, 2004, vol. 4, pp. 795–798 (2004)
[2] Yu, X., Leong, H.W., Xu, C., Tian, Q.: Trajectory-based ball detection and tracking in broadcast soccer video. IEEE Transactions on Multimedia 8(6), 1164–1178 (2006)
[3] Ren, J., Orwell, J., Jones, G.A., Xu, M.: A general framework for 3d soccer ball estimation and tracking. In: ICIP 2004, pp. 1935–1938 (2004)
[4] 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, 393–408 (2004)
[5] Leo, M., DOrazio, T., Distante, A.: Independent Component Analysis for Ball Recognition in Soccer Images. In: Proceeding of the Intelligent Systems and Control ~ISC 2003, Salzburg, Austria (2003)
[6] Murase, H.: Visual Learning and Recognition of 3-D Objects from Appearance. International Journal of Computer Vision 14, 5–24 (1995)
[7] Papageorgiou, C., Oren, M., Poggio, T.: A general framework for Object Detection. In: Proc. of Intern Conference for Computer Vision (January 1998)
[8] Rowley, H., Baluja, S., Kanade, T.: Neural Network-Based Face Detection. IEEE Trans. On Pattern analysis and Machine Intelligence 20(1), 23–38 (1998)
[9] Jones, M., Poggio, T.: Mode- based Matching by Linear Combinations of Prototypes. In: Proc. of Image Understanding workshop (1997)
[10] Mohan, A., Papageorgoiu, C., Poggio, T.: Example-based Object Detection in Images by Components. IEEE Trans. On Pattern analysis and Machine Intelligence 23(4), 349–361 (2001)
[11] Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
[12] Atherton, T.J., Kerbyson, d.J.: Size invariant circle detection. Image and video Computing 17, 795–803 (1999)
[13] Matsumoto, K., Sudo, S., Saito, H., Ozawa, S.: Optimized Camera Viewpoint Determination System for Soccer Game Broadcasting. In: Seo, Y., Choi, S., Kim, H., Hong, K.S. (eds.) Proc. IAPR Workshop on Machine Vision Applications, Tokyo, pp. 115–118 (2000)
[14] Ancona, N., Cicirelli, G., Branca, A., Distante, A.: Ball recognition in images for detecting goal in football. In: ANNIE 2001, St. Louis, MO (November 2001)
[15] Ancona, N., Cicirelli, G., Branca, A., Distante, A.: Goal detection in football by using support vector machine for classification. In: International Joint INNS–IEEE Conference on Neural Network, Washington, DC (July 2001)