SIFT APPROACH FOR BALL RECOGNITION IN SOCCER IMAGES

M. Leo, T. D'Orazio, N. Mosca and A. Distante
Institute of Intelligent Systems for Automation, Via Amendola 122/D, 70126 Bari, Italy

Keywords: Pattern Recognition, Hough Transform, Motion Analysis, Invariant Features.

Abstract: In this paper a new method for ball recognition in soccer images is proposed. It detects the ball position in each frame but, differently from related previous approaches, it does not require a long and tedious phase to build different positive training sets in order to properly manage the great variance in ball appearance. Moreover it does not need any negative training set, avoiding the difficulties to build it that occur when, as in the soccer context, negative examples abound. A large number of experiments have been carried out on image sequences acquired during real matches of the Italian "Serie A" soccer championship. The reported experiments demonstrate the satisfactory capability of the proposed approach to recognize the ball.

1 INTRODUCTION

Automatic ball recognition in image sequences is a fundamental task: a number of doubtful cases occurs during the game, as for example detecting the outside event or the goal event. 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 decision support tool for the referee committee during the game.

In the last decade different methods to automatically recognize the ball have been proposed. They could be conveniently divided in two categories: appearance based methods and motion based methods.

Motion based methods do not search the ball in each frame but distinguish the ball from other objects by means of a priori knowledge about its motion.

In (Tong, Lu & Liu, 2004) a strategy based on non-ball elimination is applied using a coarse-to-fine process. 'Condensation' algorithm is used in ball tracking and a confidence measure representing the ball region's reliability is presented to guide possible ball re-detection for continuous tracking.

In (Yu, Leong, Xu & Tian, 2006) the ball recognition task is achieved by a trajectory verification procedure based on Kalman filter rather than the low-level features. Two different procedures run iteratively (trajectory discrimination and extension) to identify the ball trajectory.

In (Ren, Orwell, Jones & Xu, 2004) soccer ball estimation and tracking using trajectory modeling from multiple image sequences is proposed.

Motion based methods seem to be well suited for video indexing applications but not for real time events detection (as, for example, to solve the goal line crossing problem) due to their intrinsic characteristic to not detect the ball in each frame but to analyze all the objects in the scene during a period of time and to recognize, at the end, the ball on the basis of some a-priori assumption about its motion.

Appearance based methods, instead, perform ball recognition in each frame using information such as shape, size and color.

From this point of view the ball recognition in soccer images is one of the applications of the most general problem of object recognition, where the mainly used approach is based on classifying the pattern images after a suitable pre-processing.

This approach has been applied in many contexts (face recognition, people detection, car detection,…) using histogram equalization, wavelet based preprocessing, parametric eigenspace decomposition (Murase, 1995; Papageorgiou, Oren & Poggio, 1998; Rowley, Baluja & Kanade, 1998; Jones & Poggio, 1997; Mohan, Papageorgiou & Poggio, 2001) to obtain a new target representation in a more suitable vector space.

Leo M., D'Orazio T., Mosca N. and Distante A. (2008).
SIFT APPROACH FOR BALL RECOGNITION IN SOCCER IMAGES.
In *Proceedings of the Tenth International Conference on Enterprise Information Systems - AIDSS*, pages 207-212
DOI: 10.5220/0001688702070212

In particular in (D'Orazio, Guaragnella, Leo & Distante, 2004; Leo, D'Orazio & Distante, 2003) wavelet and independent component analysis (ICA) were used in the soccer context in combination with a Neural Network classifier.

Unfortunately existing appearance methods for ball recognition are limited (in accuracy) by several inherent difficulties unless a long and tedious learning procedure based on multiple positive training sets (covering all the possible appearances of the ball) is optimized accurately selecting by hand representative patterns for each set.

This drawback is due to the great variance of the ball appearance over frames depending on many factors including the view, the speed of the ball, the lighting conditions, the possible partial occlusion.

Moreover appearance based methods could require negative examples (no-ball examples) that usually abound in soccer fields (player's socks, pants or shirts, advertising posters, etc.). For this reason collecting negative training examples requires hard work and caution: negative training examples must uniformly represent the universal set excluding the positive class and, at the same time, they should not heavily outnumber the positive examples.

In this paper a new appearance based ball detection approach is proposed. It consists of three steps: at first a circle detection algorithm, CHT, (Atherton & Kerbyson, 1999) is applied on the moving objects in the scene to select the region that is the best candidate to contain the ball considering only the edge information; then Scale Invariant Feature Transform, SIFT, (Lowe, 2004) is applied on the candidate region to extract representative feature vectors (keypoints); finally they are compared by nearest neighbour with those contained in the database of keypoints extracted from a set of few positive images.

Considering that both CHT and SIFT are invariant to image scale, rotation, affine distortion (only in a limited range), addition of noise, and changes in illumination, the proposed method does not require long and tedious work to build different and huge positive training sets. Moreover it does not require any negative training set.

A large number of experiments have been carried out on real image sequences acquired during Italian "Serie A" soccer championship and satisfactory ball recognition results were obtained using only few positive training images.

The rest of the paper is organized as follows: section 2 gives an overview of the proposed system, section 3 explains experimental setup and section 4 reports experimental results. Finally, discussion and conclusions are reported in section 5.

2 SYSTEM OVERVIEW

Our system operates in three stages (see Figure 1): it first applies a circle detection (implemented as convolutions on the edge magnitude image) on all the moving regions of the whole image to select the area that best fits the sought pattern as proposed by (D'Orazio, Guaragnella, Leo & Distante, 2004).

Then a two steps validation procedure is used to detect if the pattern corresponding to the highest peak in the accumulation space, is really a ball or a wrong pattern has been erroneously detected. This procedure is necessary due to the impossibility of the Hough Transform to detect occlusions and ball absence in the image, so that the ball position is wrongly determined in such situations in correspondence of the highest peak in the accumulator space.

In the first step of the validation procedure, from the sub-image containing the result of the previous detection procedure, distinctive invariant features (SIFT keypoints) are extracted by a cascade filtering approach consisting of four stages as proposed by (Lowe, 2004).

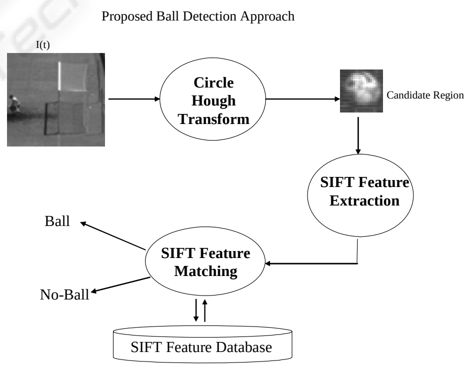

Figure 1: The Ball recognition system. I(t) is the frame acquired at the time t.

After Scale Invariant Feature Transform application, each extracted set of features is individually compared to those extracted from a small set of reference ball images stored in a database (keypoint matching).

The matching procedure is based on nearest neighbor algorithm and incorrect matching are determined by taking the ratio of distance from the closest neighbour to the distance of the second closest. If this value is greater than a proper threshold the keypoint match is considered correct, otherwise it is discarded.

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 camera able to record up to 200 frames/sec with a resolution of 640x480 pixels.

The camera was placed on the stands of the stadium with its optical axis lying on goal-mouth plane (see Figure 2).

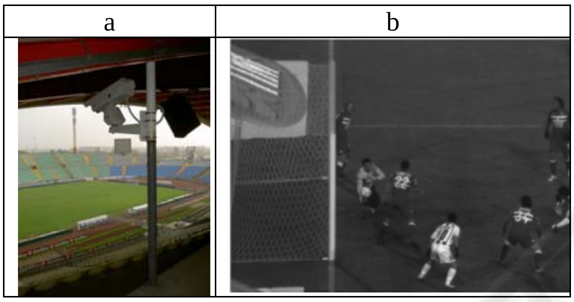

Figure 2: a) The DALSA TM 6740 camera placed on the stands of the stadium and used to acquire the image sequence during real soccer matches. In the figure it is protected by an enclosure. b) An image acquired by the camera during a match.

Different matches in different light conditions (evening matches with artificial lighting conditions, afternoon matches in both cloudy and sunny days) were acquired.

The acquired images demonstrate the great variance in the appearance of the ball depending on lighting conditions, ball speed, ball position etc. as visible in Figure 3 where three different ball appearances are shown.

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

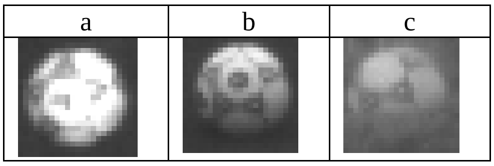

Figure 3: Three different ball appearances. a) The ball in a sunny day. b) The ball during an evening match. c) The Ball in the goal post. In this case the grid of the goal post is between the camera and the ball.

In the validation step, the positive training set consists of only 17 ball patches acquired during an evening match.

The system needs several positive ball examples (not just one) due to the different texture of the considered ball under different views. Theoretically, using a uniformly textured ball, the training set could be reduced to just one image.

In Figure 4, six of seventeen of the training images are reported.

For each test image the keypoint descriptors (the descriptor length is 128 and its elements are normalized to unit length) are compared with those of the reference images by means of the matching procedure.

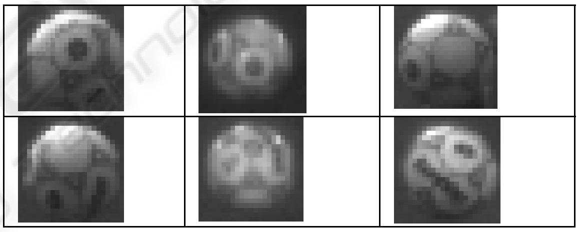

Figure 4: Six of the seventeen training images used in the experiments.

Keypoint descriptors matching procedure is based on Euclidean distance. Matching procedure yields, for each test image, the number of matched keypoints for each training image.

Finally, to validate the tested patch, the mean value of successful matches between the tested patch and the 17 training patches is used. If the mean value is higher than a experimental threshold then the test image is labelled as ball otherwise it is labelled as no-ball.

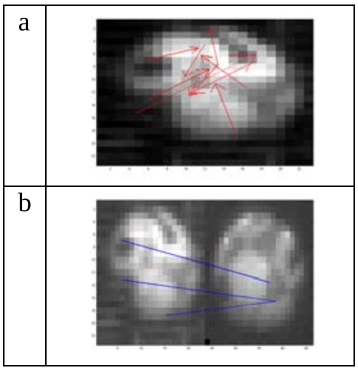

Figure 5: a) The keypoints localized on a ball image. b) Three keyponts matching between a test patch and a training patch.

In Figure 5.a the keypoints localized on a ball image are shown.

In Figure 5.b the keyponts matched between a test patch and a training patch are connected.

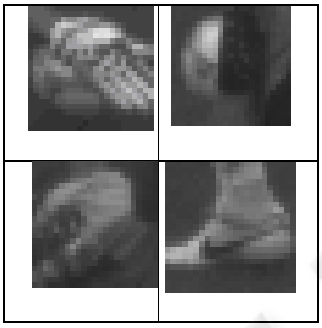

Figure 6: Two ball instances not detected by CHT (first row) and the corresponding regions erroneously chosen as ball candidates (second row).

4 EXPERIMENTAL RESULTS

The experimental phase has been divided in two parts: in the first one the ball recognition approach has been evaluated on the sequences acquired during evening matches (with artificial light conditions); in the second part the proposed approach has been applied on images acquired with natural light conditions. In the first experiment a set of 3560 images was used; 1945 of these images contain the ball (and nearly always some players and the goal-keeper) and the remaining 1615 do not contain the ball but only some players and the goal-keeper.

Table 1 reports a scatter matrix explaining the performance of the Circle Hough Transform to correctly extract the region containing the ball (when it is present in the scene).

Table 1: The performance of the Circle Hough Transform for ball candidate detection in soccer images acquired during evening matches.

	Extracted Patches containing the Ball	Extracted Patches not containing the Ball
Images containing the Ball (1945)	**98.76 % (1921/1945)**	**1.23% (24/1945)**
Images not containing the Ball (1615)	**0% (0/1615)**	**100% (1615/1615)**

When the ball was in the scene the CHT was almost always able to correctly extract the patch around it as a candidate ball region. The CHT failed only when the ball was heavily occluded.

Figure 6 reports two occurrences where the ball region was not identified by CHT (first row) and the corresponding regions erroneously chosen as ball candidates (second row). In these cases the peaks in the accumulation space associated to the objects in the second row (a shoulder of the goal-keeper and a shoe of a player) overcomes the one associated to the ball caused by occlusion. In Table 2 a scatter matrix explaining the performance of the validation step by matching of SIFT are reported.

Table 2: The performance of the SIFT Keypoints matching for ball candidate validation in soccer images acquired during evening matches.

	Ball	No-Ball
Ball (1921)	**90.3 % (1734/1921)**	**9.7% (187/1921)**
No-Ball (1639)	**10.92% (179/1639)**	**89.08% (1460/1639)**

More than 90% of candidate regions containing the ball were correctly validated. At the same time, almost 89% of candidate regions that do not contain the ball were correctly discarded.

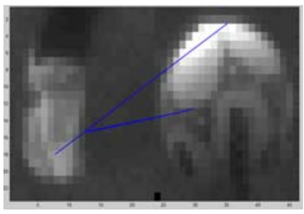

Figure 7: A false positive occurs when, in under a certain perspective, the texture of the player’s shoe matches some areas of the ball.

In Figure 7 a case of incorrect validation is shown. Keypoints relative to the texture of the player’s shoe matches some areas of the ball causing an error obtained by using the proposed approach.

Table 3 summarizes ball recognition performance by the proposed approach with artificial light condition. It combines the performance of the circle Hough transform for candidate ball detection (Table 1) and SIFT matching for candidate ball validation (Table 2). As reported the global performance of the system are almost to reach 90% of correct recognition of the ball in the images.

Table 3: The global performance of the proposed approach for recognizing the ball in soccer images acquired during evening matches.

	Ball	No Ball
Ball (1945)	**89.15 % (1734/1945)**	**10.84% (211/1945)**
No Ball (1615)	**11.08% (179/1615)**	**88.91% (1436/1615)**

In the second experiment a set of 2147 images acquired with natural light conditions was used. 1034 of these images contain the ball (and nearly always some players and the goal-keeper) and the remaining 1113 do not contain the ball but only some players and the goal-keeper.

Table 4 reports the scatter matrix explaining the performance of the Circle Hough Transform to correctly extract the region containing the ball (when it is present in the scene).

Table 4: The performance of the Circle Hough Transform for ball candidate detection in soccer images acquired with natural light conditions.

	Extracted Patches containing the Ball	Extracted Patches not containing the Ball
Images containing the Ball (1034)	**97.19 % (1005/1034)**	**2.80% (29/1034)**
Images not containing the Ball (1113)	**0% (0/1113)**	**100% (1113/1113)**

Comparing Table 1 and table 4 it is possible to conclude that performance are similar to those obtained in the first experiment even if some additional misdetections of the ball regions occur due to the presence of the self shadow on the ball that reduces edge matching in the convolution with the oriented masks.

In Table 5 the performance of the validation step by matching of Scale Invariant Feature are reported. The presence of self shadows and the saturation of some areas of the ball appearance due to the sunrise reflection (see Figure 2.a) reduce ball recognition performance with respect to Table 2.

Table 5: The performance of the SIFT’s Keypoints matching for ball candidate validation in soccer images acquired with natural light conditions.

	Ball	No Ball
Ball (1005)	**83.08 % (835/1005)**	**16.91% (170/1005)**
No-Ball (1142)	**10.07% (115/1142)**	**89.93% (1027/1142)**

Table 6 summarizes ball recognition performance by the proposed approach with natural light conditions. It combines the performance of the circle Hough transform for candidate ball detection (Table 4) and SIFT matching for candidate ball validation (Table 5).

Table 6: The global performance of the proposed approach for recognizing the ball in soccer image acquired with natural light condition.

	Ball	No Ball
Ball (1034)	**80.75 % (835/1034)**	**19.25% (199/1034)**
No Ball (1113)	**10.33% (115/1113)**	**89.67% (998/1113)**

In Figure 8 an example of misrecognition of the ball due to reflection effects on the ball surface is shown. Anyhow, performance remain satisfactory (more than 80%) even if the small training set does not contains balls acquired in the tested natural lighting conditions.

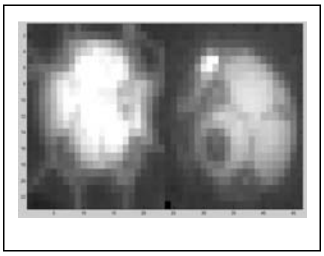

Figure 8: An example of misrecognition of the ball due to reflection effects on the ball surface.

As will be highlighted in the conclusions, this is, in our opinion, a very pleasant result, because the manual extraction of a small set of positive examples guarantee very good results even in light conditions that were not well represented or not considered at all. This makes the proposed approach sometimes preferable with respect to methods that are more accurate but also more demanding in user intervention.

5 DISCUSSION AND CONCLUSIONS

Experimental results demonstrate the capability of the proposed approach to recognize the ball in image soccer sequences acquired in different light conditions.

Ball recognition performance are comparable to those obtained in previous works by using appearance based methods involving more complex training procedures.

Differently from the other appearance based methods that can be found in the literature, the proposed one does not require, nevertheless, a long and tedious phase to build different training sets to manage different light conditions. Moreover it does not require any negative training set, avoiding the difficulties relative to the balancing of the number of negative and positive examples that occurs when, as in the soccer context, negative examples abound (player's socks, pants or shirts, advertising posters, etc.).

The usefulness of such approach allows to use the method in real systems with little user intervention during the setup phase of the system installation. Last but not least, the fact that it poses less emphasis in the acquisition of negative examples and the balancing with the positive ones, means that it is less prone to errors when dealing with previously unknown external objects.

In the reported experiments only one set of 17 ball images acquired in 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 ball texture is not uniform and moreover ball appearance can change also depending on the stadium.

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 newest vision tools able to avoid saturation effects on ball surface and introducing different keypoint matching strategies.

REFERENCES

Atherton, T. J., Kerbyson, d. J., 1999. Size invariant circle detection. *Image and video Computing.*

D'Orazio, T., Guaragnella, C., Leo, M., Distante, A., 2004. A new algorithm for ball recognition using circle Hough Transform and neural classifier. *Pattern Recognition, 37*, 393-408.

Jones, M., Poggio, T., 1997. Model based Matching by Linear Combinations of Prototypes. *Proceedings of Image Understanding workshop.*

Leo, M., D'Orazio, T., Distante, A., 2003. Independent Component Analysis for Ball Recognition in Soccer Images. *Proceedings of the Intelligent Systems and Control.*

Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision, 60, 2*, 91-110.

Mohan, A., Papageorgoiu, C., Poggio, T., 2001. Example-based Object Detection in Images by Components. *IEEE Transactions On Pattern Analysis and Machine Intelligence, 23, 4*, 349-361.

Murase, H., 1995. Visual Learning and Recognition of 3-D Objects from Appearance. *International Journal of Computer Vision, 14*, 5-24.

Papageorgiou, C., Oren, M., Poggio, T., 1998. A general framework for Object Detection. *Proceedings of the International Conference for Computer Vision.*

Ren, J., Orwell, J., Jones, G. A., Xu, M., 2004. A general framework for 3d soccer ball estimation and tracking, *ICIP 2004*, 1935-1938.

Rowley, H., Baluja, S., Kanade, T., 1998. Neural Network-Based Face Detection. *IEEE Trans. On Pattern analysis and Machine Intelligence, 20, 1*, 23-38.

Tong, X. F., Lu H. Q., Liu Q. S., 2004. An effective and fast soccer ball detection and tracking method. *Pattern Recognition, ICPR 2004, Proceedings of the 17th International Conference, 4*, 795-798.

Yu, X., Leong, H.W., Xu C., Tian, Q., 2006. Trajectory-based ball detection and tracking in broadcast soccer video. *IEEE Transactions on Multimedia, 8, 6*, 1164-1178.