Aerial ball perception based on the use of a single perspective camera

João Silva, Mário Antunes, Nuno Lau, António J. R. Neves, Luís Seabra Lopes
{joao.m.silva,mario.antunes,nunolau,an,lsl}@ua.pt

IEETA / Department of Electronics, Telecommunications and Informatics
University of Aveiro, Portugal

Abstract. The detection of the ball when it is not on the ground is an important research line within the Middle Size League of RoboCup. A correct detection of airborne balls is particularly important for goal keepers, since shots to goal are usually made that way. To tackle this problem on the CAMBADA team , we installed a perspective camera on the robot. This paper presents an analysis of the scenario and assumptions about the use of a single perspective camera for the purpose of 3D ball perception. The algorithm is based on physical properties of the perspective vision system and an heuristic that relates the size and position of the ball detected in the image and its position in the space relative to the camera. Regarding the ball detection, we attempt an approach based on a hybrid process of color segmentation to select regions of interest and statistical analysis of a global shape context histogram. This analysis attempts to classify the candidates as round or not round. Preliminary results are presented regarding the ball detection approach that confirms its effectiveness in uncontrolled environments. Moreover, experimental results are also presented for the ball position estimation and a sensor fusion proposal is described to merge the information of the ball into the worldstate of the robot.

1 Introduction

In the context of the Middle Size League (MSL) of RoboCup where the ball is shot through the air when the robots try to score goals, it is important to have some estimation of the ball path when it is in the air.

The CAMBADA team robots are equipped with an omnidirectional vision system which is capable of detecting the ball when it is on the ground, but fails to detect the ball as soon as it goes higher than themselves. In this paper, we present a new proposal to achieve a perception of the ball on the air using a single perspective camera, installed in the robots.

To achieve this objective, we explore the physical properties of the vision system and the correspondent geometric approximations to relate the position of the detected ball on the image and its position over the field. The ball detection is based on a hybrid approach. This approach is based on color segmentation for

Region Of Interest (ROI) selection and subsequent global shape context analysis for circularity estimation.

In section 2, the problem is briefly exposed and some related work overviewed. Section 3 describes the used vision system and section 4 presents the detection and estimation of ball candidates on the image. In section 5, the algorithm for estimating the position of the ball candidates on the space in front of the camera is presented and section 6 introduces some guidelines for the integration of information from both the cameras of the robots. Section 7 presents a brief comment on the work. Finally, in section 8, some final remarks are presented as future guidelines for this problem.

2 Problem statement and related work

The work presented in this document is focused on the perception of a ball in a robotic soccer scenario. The soccer ball to detect is a size 5 FIFA ball, which has approximately 22 cm of diameter. According to the rules, it has a known predominant color for each tournament. Most teams take advantage of this restriction while the ball is on the ground, since in that case, the environment is very controlled (green floor with some white lines and black robots) and the ball candidates can be expected to be surrounded by this reduced set of colors. In the air, these restrictions are completely lost and thus the approach should be more shape based. The existing shape based approaches are mainly aiming at detecting a ball through shape on the omnidirectional camera.

Several teams already presented preliminary work on 3D ball detection using information from several sources, either two cameras or other robots information [1,2]. However, these approaches rely on the ball being visible by more than one source at the same time, either two cameras with overlapping visible areas or two robots, and then triangulate it. This is not possible if the ball is above the robot omnidirectional camera.

Regarding the detection of arbitrary balls, the MSL league is the most advanced one. Many of the algorithms proposed during previous research work showed promising results but, unfortunately, in some of them, the processing time do not allow its use during a game, being in some cases over one second per video frame [3].

Hanek *et al.* [4] proposed a Contracting Curve Density algorithm to recognize the ball without color labeling. This algorithm fits parametric curve models to the image data by using local criteria based on local image statistics to separate adjacent regions. The author claims that this method can extract the contour of the ball even in cluttered environments under different illumination, but the vague position of the ball should be known in advance. The global detection cannot be realized by this method.

Treptow *et al.* [5] proposed a method for detecting and tracking the ball in a RoboCup scenario without the need for color information. It uses Haar-like features trained by an adaboost algorithm to get a colorless representation of the ball. Tracking is performed by a particle filter. The author claims that the algorithm is able to track the ball with 25 fps using images with a resolution of 320×240 pixels. However, the training process is too complex and the algorithm cannot perform in real time for images with higher resolutions. Moreover, the results still show the detection of some false positives.

Mitri *et al.* [6] presented a scheme for color invariant ball detection, in which the edged filtered images serve as the input of an Adaboost learning procedure that constructs a cascade of classification and regression trees. This method can detect different soccer balls in different environments, but the false positive rate is high when there are other round objects in the environment.

Lu *et al.* [7] considered that the ball on the field can be approximated by an ellipse. They scan the color variation to search for the possible major and minor axes of the ellipse, using radial and rotary scanning, respectively. A ball is considered if the middle points of a possible major axis and a possible minor axis are very close to each other in the image. However, this method has a processing time that can achieve 150 ms if the tracking algorithm fails.

More recently, Neves *et al.* [8] proposed an algorithm based on the use of an edge detector, followed by the circular Hough transform and a validation algorithm. The average processing time of this approach was approximately 15 ms. However, to use this approach in real time it is necessary to know the size of the ball along the image, which is simple when considering the ground plane in a omnidirectional vision system. This is not the case when the ball is in the air, in a completely unknown environment without any defined plane.

3 The Perspective camera

The used camera contains a CCD of $6.26 \times 5.01 mm$ and pixel size $3.75 \mu m$. The maximum resolution is 1296×964 and the used lens has a $4mm$ focal length. The camera is fixed to the robot in such a way that the axis normal to the CCD plane is parallel to the ground (Fig. 1).

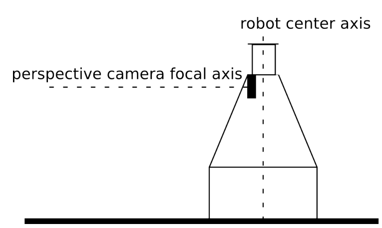


Fig. 1. Illustration of the perspective camera positioning on the robot. It is placed in such a way that the camera focal axis is parallel to the ground and the focal point is slightly displaced from the robot center axis.

Based on the CCD size and the focal length, we can derive the opening angle, both along the horizontal and vertical axis. We will use α to identify the horizontal opening angle and β to identify the vertical opening angle.

Given α and β, we can also estimate the theoretical relation of the field of view (FOV) of the camera at different distances (these calculations do not take into account any distortion that may be caused by the lens). For a given distance, we can then estimate the width of the plane parallel to the camera CCD (as illustrated in Fig. 2) by:

$$tan(\alpha) = \frac{1}{2}\frac{hFOV}{Y} \Rightarrow hFOV = 2 \times Y \times tan(\alpha) \tag{1}$$

where $hFOV$ is the width of the FOV plane and Y is the distance of the FOV plane to the camera focal point.

For the height of the same plane, the analysis is similar in every way, now considering the ccd height and a β angle.

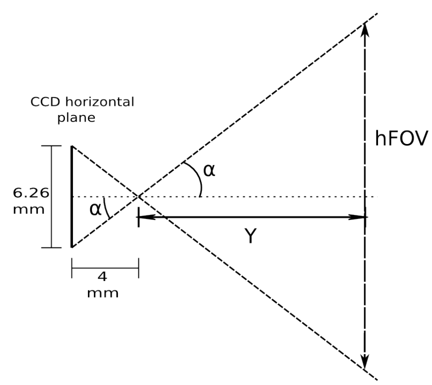


Fig. 2. Scheme of the relation between the CCD width and focal length with the opening angle. The geometric relation is used for estimating the FOV at a given distance.

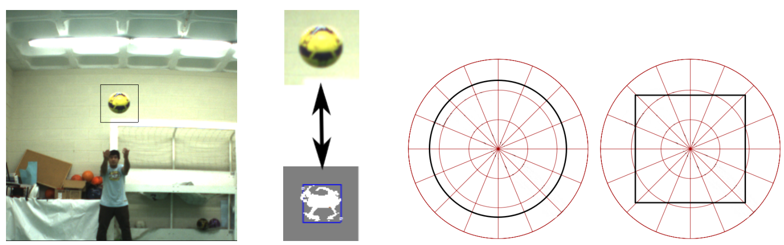
Due to the position of the camera relative on the robot, the analysis of the FOV and associated geometric characteristics of the camera image have a more direct application. The idea of this application is to provide an approximation of the ball position mainly when it is in the air. The camera captures the images using format7, which allows to use the full resolution of the camera to capture the images but also allows to get and use only a specific ROI. For the objectives of the described work, we opted to use the full horizontal size of the image, while the vertical size was cropped to the first 500 lines. This value was defined to cope with the choice that the camera is used to detect aerial balls and thus it only needs the image above the horizon line. Since the maximum vertical resolution is 964, we use the top of the image, with a small margin.

4 Ball visual detection and validation

Our proposal to detect ball candidates in the air is to use a hybrid approach of color segmentation and statistical analysis of a global shape context histogram.

On a first phase, and since the ball main color is known, a color segmentation of the image is made in order to obtain blobs of the ball color. This is achieved by a horizontal scan of the image rows. On each row, the ball color is detected and a blob is built row by row. The generated blobs have some properties which are immediately analyzed. Based on thresholds for minimum size and solidity of the blob convex hull, each solid blob is selected as a candidate for ball while blobs with very low values of solidity are discarded. An image of the ROI with the blob is created (Fig. 3b). The method to calibrate the colors and camera parameters is the same as the one used for the team omnidirectional camera and is described in [9].

These images are then analyzed by a modified global shape context classifier [10]. Each image is pre-processed with an edge detector and a polar histogram is created. This histogram is then statistically analyzed and returns a measure of circularity of the candidate. The edges image is divided in n layers and m angles, creating the polar histogram, as defined in [11]. The analysis of the histogram is made layer by layer, covering all the angles. An estimation of the average number of edge points on each slice and its standard deviation allows a rough discrimination between circular and non-circular contours, as exemplified in Fig. 3c. A ratio of edge points is also part of the statistics of the candidate.



a) **b)** **c)**

Fig. 3. a): image from the camera with a ROI around the ball, which is the portion of the image used for the histogram analysis; **b)**: the ROI created based on the correspondent color blob; **c)**: rough representation of the polar histogram. The histogram creation fits its radius with the outer points of the edge image, which is not fully represented in these pictures. *Left*: A perfect circular edge on the polar histogram would look something like this. All the edge points are on the same layer and each of its slices have a similar number of points; *Right*: A square on the polar histogram. Due to the fitting properties of the histogram itself, the square edge points should be divided in more than one layer, which would not yield good statistics as circles.

The previously described step always returns the layer with the best statistics, which is currently the layer that has higher average value with minimum standard deviation (the maximum difference between average and standard deviation). This should represent the layer with the most consistent number of edge pixels and thus should be the rounder layer. The next step must then select which of these statistics make sense. Currently, three characteristics are analysed:

- The ratio of edge points on the layer must be within a given range. This range was empirically estimated through testing of several examples of ball and no ball candidates.
- The order of magnitude of the mean should be greater than or equal to the order of magnitude of the standard deviation.
- The candidate diameter estimated by color segmentation and the diameter estimated by the classifier must be coherent. Since the radius of the histogram is dependent on the number of layers, the coherence between the measures is dependent on an error margin based on the histogram radius.

4.1 Experimental results

Some experiments were performed by throwing the ball through the air from a position approximately 6 meters away from the camera and in its direction. In the acquired videos the ball is always present and performs a single lob shot.

As expected, since the ball is moving almost directly to the camera, the variation of the ball center column on the image is very small (Fig. 4). The row of the ball center, however, was expected to vary. Initially the ball was being held low on the image (meaning the row of the image was a high value) and as it was thrown, it was expected that it went up on the image, then down again. Fig. 4 allows us to verify that this behavior was also observed as expected.

On the other hand, since the ball is coming closer to the camera every frame, it was also expectable that its size on the image would be constantly growing. The correct evaluation of the width of the ball is important for the position estimation described in the next section.

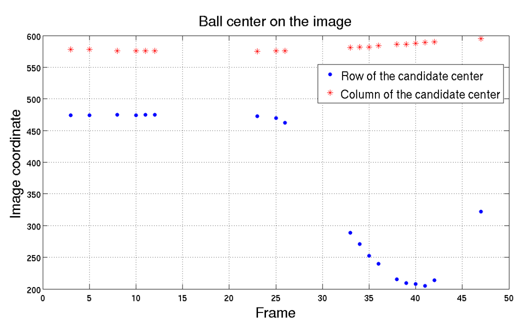


Fig. 4. Results from the image analysis of a ball being thrown in the camera direction: values of the row (Blue dots) and column (Red stars) where the ball center was detected on the image.

The main contribution of this color/shape hybrid approach, however, is the reliability of the acquired data, respecting the real time constraints of the application. A test scenario was created, where the ball was thrown in such a way that it was visible in every frame of the videos, and several runs were made. Although the results are strongly affected by the environment around the ball, we obtained promising preliminary results with relatively high precision, even if the recall has shown lower results. A pure color approach yielded better results in a very controlled environment, but in a more uncontrolled environment we obtained a very low precision. We consider that having a higher precision is more advantageous, even when facing the loss of some recall. The processing time for the algorithm was 11.9 ± 2.8 ms which is still within our time restrictions.

5 Ball position estimation

After having the candidates selected as balls, there is the need to estimate their position. To accomplish that, we first analyze the candidate radius in pixels. The size that each pixel represents at each distance increases with distance to the camera focal point. This is due to the fact that the resolution is constant but the FOV is not. With the FOV width relation, we can estimate the size that each pixel represents at each given distance, by a relation of the estimated distance and the horizontal resolution:

$$pS = \frac{hFOV}{hR} \tag{2}$$

and thus, since we know that the ball has $0.22m$, we can estimate the number of pixel expected for the blob width at a given distance:

$$pW = \frac{0.22}{pS} \tag{3}$$

where pS is the pixel size in the plane with the defined horizontal FOV($hFOV$), hR is the CCD horizontal resolution and pW is the expected ball width, in pixels, for the plane with the given pixel size.

For the same setpoint distances as before the ball width, in pixels, was estimated. Table 1 presents those results.

Distance to camera	1	2	3	4	5	6	7	8	9
Expected ball width	182	91	61	46	36	30	26	23	20

Table 1. Table with the theoretical ball width at several distances. Distances are in meters, ball width are in pixels.

From this analysis, Equation 3 can be developed using Equations 2 and 1:

$$pW = \frac{0.22 \times hR}{hFOV} = \frac{0.22 \times hR}{2 \times Y \times tan(\alpha)} \tag{4}$$

from which we get an inverse relation function of pixel width pW and distance to camera focal point Y.

Given the known distance of the ball candidate, which is our YY coordinate, and the linear relation of the pixel size, we can estimate the XX coordinate.

This is true due to the camera positioning on the robot that, besides having the focal axis parallel to the ground, it is also coincident with the robot YY axis. To accomplish the estimation of the XX coordinate we have to analyze the horizontal pixel coordinate of the ball center from the image center and apply the calculated pixel size.

We can thus obtain the XX and YY coordinates of the ball on the ground plane, relative to the perspective camera, from which we know the relative coordinates from the center of the robot.

5.1 Experimental results

An experimental analysis was performed to verify the relation between the detected ball width in pixels and the distance it is from the camera.

Unfortunately, and like most practical scenario, it was verified that the expected theoretical values of the ball pixel width according to the distance was not verified in practice. To verify the ball width according to the distance from it to the camera, an experimental setup was mounted.

The experiment was performed by placing the camera with its axis along a field line and placing the ball in front of it. The ball was on top off a support, which maintained ball height, and was placed at the several setpoint distances from the camera (from one to nine meters). These distances were measured with tape and used as ground truth data. The ball pixel width was measured by manually stopping the video at frames corresponding to the setpoint distances and verifying the estimate width of the detected ball. The results are presented in Table 2.

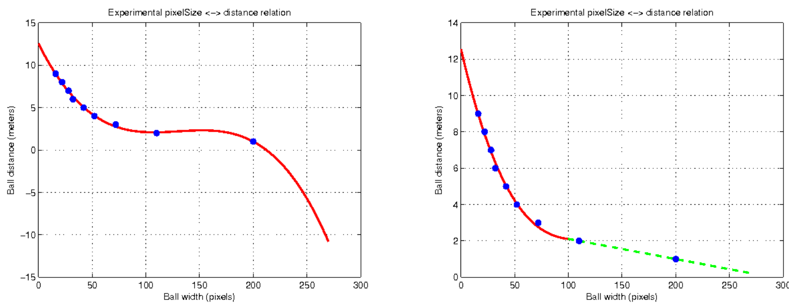
Distance to camera	1	2	3	4	5	6	7	8	9
Measured ball width	200	110	72	52	42	32	28	22	16

Table 2. Table with the measured ball width at several distances. Distances are in meters, ball width are in pixels.

Based on the values for the relation between the ball pixel width and the distance to camera, a 3rd degree polynomial function is used to, given a ball blob's width, estimate its distance to camera (the YY coordinate). This relation is slightly different from the theoretical relation presented in Equation 4, due to factors like lens distortion that were not accounted in the previous analysis.

To make an approximation for this data, we can estimate a polynomial function that, given a pixel width of a given candidate, returns the distance at which that candidate is from the camera. To keep computational time affordable, we do not wish to have high degree polynomial functions, and thus we tested the fit of functions up to 4th degree, and verified that a 3rd degree function would fit the data acceptably. However, the function behavior at shorter distances is not proper (Fig. 5a).

For that reason, and given the short distances, a linear approximation of the data would fit the correspondent data in a better way. Fig. 5b represents the two polynomial functions considered. The used separation point was empirically estimated.



a) **b)**

Fig. 5. Left **a)**: Third degree polynomial (red line) fitting the defined experimental setpoints (blue dots). Although the polynomial function fits acceptably for sizes corresponding to distances above 2 meters, closer distances sizes do not seem to evolve according to the same curve; Right **b)**: Third degree polynomial function (red line) fitting the defined experimental setpoints (blue dots) and linear polynomial function (green dashed) for the sizes corresponding to closer distances.

In the same experiment of throwing the ball from 6 meters away from the camera, described in Section 4.1, the results of the positions evaluated by the previously described algorithm were captured. Fig. 6 depicts these results. The path formed by the estimated XY positions approximates the path of the ball arc through the air, projected on the ground. This data allows a robot equipped with such camera to estimate the path of the incoming airborne ball and place itself in front of the ball, for defending in the case of the goal keeper. Given the nature of the task, there is no need for an excellent precision on the estimations, just a general direction which provides the target for the robot. The perspective vision process, from capture to the production of the XX and YY coordinates took an average time of around 12.5 ms to execute in a computer with an Intel Core 2 duo at 2.0 GHz. The tests were made for the perspective camera process running standalone at 30 frames per second.

6 Ball integration

Being the ball the main element of a soccer game, its information is very important and needs to be as precise as possible. Failure on its detection can have very negative impact on the team performance. Probably even worse than failing to detect the ball (situation on which the robots can strategically move in search for it) is the identification of false positives on the ball. This can deviate the attention of a robot or even the entire team from the real ball, which can be catastrophic. To avoid false positives and keep coherence on the information of the ball, several contextual details are taken into account.

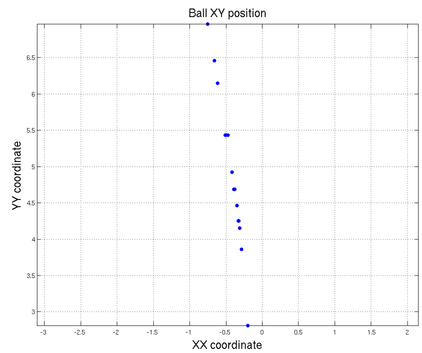


Fig. 6. Picture of a data capture of a ball kicking test. The ball was thrown by the air from a position around (-0.5, 6.0) in the approximate direction of the camera (which is the origin of the referential). The blue dots represent the estimated ball positions.

Given the several possible sources of information, the priority for ball position is the omnidirectional camera. Details about the visual ball detection and validation on the omnidirectional camera can be found in [9]. The next source to use is the shared information between team mates, because if they see the ball on the ground, there is no need to check the perspective camera information. Finally, the agent tries to fit the information from the perspective camera into the worldstate (Fig. 7). This is an improvement of the integration algorithm for the ball position information presented in [12].

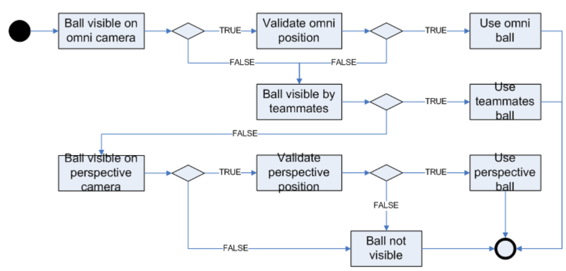


Fig. 7. Ball integration diagram.

At this point, the visual information is a list of candidates that can be the ball.

The validation of a perspective ball candidate depends on some context of its detection, based on the analysis of information known at this point:

– The first analysis to make is whether the ball candidate is inside the field or not. If it is outside the field of play, there is no need to use it, since even if it is truly the ball, the game would be stopped.
– To maintain coherence between the perspective vision ball and the omni vision ball, an analysis of the last omni camera ball positions is made and a perspective candidate is considered only if the this candidate position is in

a given vicinity of the omni camera ball position. Since we are interested in getting the general ball path, the angular difference is the measure considered for defining this vicinity. A candidate from the perspective camera is only accepted if the difference to the last omni camera ball position is below a given threshold. This validation is performed on the first detections by the frontal camera, when the ball has also just became or is becoming not visible for the omni directional camera.
- Another filtering that is done is an analysis of the number of cycles with the ball visible on the perspective camera. Again, the objective of the perspective camera is to detect aerial balls. During the game, the ball leaves the ground only on short periods. When a kick raises the ball, it will inevitably be in the air for only a few instants, which can be periodic if the ball bounces several times, but still the appearances are short. A ball constantly detected for more than a given amount of time is then discarded, since it is probably a false positive or, for instance, a stop game situation and the referee is holding the ball on his hands.

7 Conclusions

This paper presents a new approach for aerial ball perception based on the use of a single perspective camera. This approach is based on three main steps, the visual ball detection followed by an estimation of the ball position based on a geometric analysis of the vision system and finally a sensor fusion approach of this information with other sources of information.

The hybrid approach for visual detection of the ball uses a fast color segmentation based algorithm combined with the application of a polar histogram analysis. Although a pure shape based algorithm could provide more accurate results, the fact that this application has real-time restrictions, lead us to include the color segmentation based algorithm to reduce the shape analysis to limited small size ROIs.

8 Future work

The initial analysis of the performance of this approach showed that there are some limitations to its use on a real game scenario, mainly due to the fact that the object of interest, the ball, moves at a very high speed. In many frames of a video capture, it is verified that the distortion blur is very high and thus, the shape analysis is compromised, forcing us to wide the range of detection, thus lowering the effectiveness of the algorithm.

As future approaches, we intent to explore two scenarios to try to deal with this problem:

- to export and use the detection approach on high speed cameras, which would probably provide us frames with a reduced blur effect (even if we could/should not process all the frames)
- to try a new approach based on 3D Kinect camera to detect aerial objects.

Aknowledgements

This work was developed in the Institute of Electronic and Telematic Engineering of University of Aveiro and was partially funded by FEDER through the Operational Program Competitiveness Factors - COMPETE and by National Funds through FCT - Foundation for Science and Technology in the context of the project FCOMP-01-0124-FEDER-022682 (FCT reference PEst-C/EEI/UI0127/2011).

References

1. Voigtlrande, A., Lange, S., Lauer, M., Riedmiller, M.: Real-time 3D ball recognition using perspective and catadioptric cameras. In: Proc. of the 3rd European Conference on Mobile Robots, Freiburg, Germany (2007)
2. Burla, A.: 3D ball detection in robocup. Master's thesis, University of Stuttgart (2007)
3. Mitri, S., Frintrop, S., Pervolz, K., Surmann, H., Nuchter, A.: Robust object detection at regions of interest with an application in ball recognition. In: Proc. of the 2005 IEEE International Conference on Robotics and Automation, ICRA 2005, Barcelona, Spain (2005) 125–130
4. Hanek, R., Schmitt, T., Buck, S.: Fast image-based object localization in natural scenes. In: Proc. of the 2002 IEEE/RSJ Int. Conference on Intelligent Robotics and Systems, Lausanne, Switzerland (2002) 116–122
5. Treptow, A., Zell, A.: Real-time object tracking for soccer-robots without color information. Robotics and Autonomous Systems **48** (2004) 41–48
6. Mitri, S., Pervolz, K., Surmann, H., Nuchter, A.: Fast color independent ball detection for mobile robots. In: Proc. of the 2004 IEEE Int. Conference on Mechatronics and Robotics, Aechen, Germany (2004) 900–905
7. Lu, H., Zheng, Z., Liu, F., Wang, X.: A robust object recognition method for soccer robots. In: Proc. of the 7th World Congress on Intelligent Control and Automation, Chongqing, China (2008)
8. Neves, A.J.R., Azevedo, J.L., Cunha, B., Lau, N., Silva, J., Santos, F., Corrente, G., Martins, D.A., Figueiredo, N., Pereira, A., Almeida, L., Lopes, L.S., Pinho, A.J., Rodrigues, J., Pedreiras, P.: CAMBADA soccer team: from robot architecture to multiagent coordination. In Papic, V., ed.: Robot Soccer. INTECH (2010) 19–45
9. Neves, A.J., Pinho, A.J., Martins, D.A., Cunha, B.: An efficient omnidirectional vision system for soccer robots: From calibration to object detection. Mechatronics **21** (2011) 399–410
10. Pereira, R., Lopes, L.S.: Learning visual object categories with global descriptors and local features. In: Progress in Artificial Intelligence. Volume 5816., Aveiro, Portugal (2009) 225–236
11. Antunes, M.L.P.: Semantic vision agent for robotics. Master's thesis, University of Aveiro (2011)
12. Silva, J., Lau, N., Neves, A.J.R., Rodrigues, J., Azevedo, J.L.: World modeling on an MSL robotic soccer team. Mechatronics **21** (2011) 411–422