A Visual Landmark Recognition System for Topological Navigation of Mobile Robots.

CONFERENCE PAPER *in* PROCEEDINGS - IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION · JANUARY 2001

DOI: 10.1109/ROBOT.2001.932762 · Source: DBLP

4 AUTHORS:

M. Mata
European University of Madrid
24 PUBLICATIONS **307** CITATIONS
SEE PROFILE

J.M. Armingol
University Carlos III de Madrid
97 PUBLICATIONS **1,094** CITATIONS
SEE PROFILE

Arturo de la Escalera
University Carlos III de Madrid
100 PUBLICATIONS **1,135** CITATIONS
SEE PROFILE

Miguel Angel Salichs
University Carlos III de Madrid
122 PUBLICATIONS **992** CITATIONS
SEE PROFILE

A Visual Landmark Recognition System for Topological Navigation of Mobile Robots

M. Mata, J. M. Armingol, A. de la Escalera and M. A. Salichs

Universidad Carlos III de Madrid, Division of Systems Engineering and Automation.
C/ Butarque 15, 28911 Leganés (Madrid) SPAIN
{mmata, armingol, escalera, salichs}@ing.uc3m.es

Abstract

This paper describes a vision-based landmark recognition system for use with mobile robot navigation tasks. A search algorithm based on genetic techniques for pattern recognition in digital images is presented. The developed system allows the topologic localization of a mobile robot using natural and artificial landmarks. Text strings inside landmarks can be read and interpreted, if present.

The resulting system has been tested onboard a B21 mobile robot and proved useful. The presented experimental results show the effectiveness of the proposed algorithm.

1 Introduction

A prerequisite for geometric navigation of a mobile robot is a position-finding method. As a mobile robot moves through its environment, its actual position and orientation always differ from the position and orientation that it is commanded to hold, being wheel slippage a major source of error [4]. Errors accumulate and the localization uncertainty increases over time.

Topological navigation allows to overcome some of the classical problems of geometric navigation in mobile robots, such as simultaneously reducing the uncertainty of localization and perception of the environment [9].

On the other hand, topological navigation is heavily dependent on a powerful perception system to identify elements of the environment. Chosen elements should be simple enough in order to permit an easy identification from different view angles and distances.

There are two major approaches in the use of landmarks for topological navigation in related literature. One approach uses as landmarks regions of the environment that can be recognized later, although they are not a single object. For example, in [2] a spatial navigation system based on visual templates is presented. Templates are created by selecting a number of high-contrast features in the image and storing them together with their relative spatial locations in the image. Franz *et al.* [6] have developed a vision based system for topological navigation in open environments. This system represents selected places by local 360º views of the surrounding scenes. The second approach uses objects of the environment as landmarks, with perception algorithms designed specifically for each object. Beccari *et al.* [3] describe a series of motor and perceptual behaviors used for indoor navigation of mobile robot; walls, doors and corridors are used as landmarks.

This paper presents a visual landmark recognition system for topological navigation. A landmark is a localized physical feature that the robot can sense and use to estimate its own position in relation to a "map" that contains the landmark's relative position and/or other mark characterization.

One of the major contributions of this work is that the visual system is able to work with any 2D (or nearly 2D) landmark. This system is not specifically developed for only one object. In the experiments carried out, two very different landmarks have been used with the same searching algorithm: an artificial landmark (green circles placed on the walls) and a natural landmark (office's nameplates attached at the entrance of each room), shown in Figure 1. The search algorithm is based on genetic techniques for pattern recognition in digital images. This genetic search algorithm is able to handle landmark perspective deformation problems.

Figure 1. Artificial and natural landmarks.

2 Pattern recognition method

Pattern search is done using the 2D Pattern Search Engine presented in this paper, and designed for general application. Once a landmark is found, the related information extraction stage depends on each mark, since they contain different types and amounts of information. However, the topological event which is generated with a successful landmark recognition is independent from the selected landmark, except for the opportunity of "high level" localization which implies the interpretation of the contents of an office's nameplate. A brief description of each of this stages is given below.

2.1. 2D pattern search engine

There is a large collection of 2D pattern search techniques in the literature [8]. In our application we use a classical one: normalized correlation with an image of the pattern to find (usually named *model*).

The advantages and drawbacks of this technique are well known. The strongest drawback is its high sensitivity to pattern aspect changes (size, perspective and illumination), which makes this method unpractical in most cases. A two step modified method is proposed for overcoming this problem. First, in a segmentation stage, relevant regions in the image are highlighted; then the regions found (if any) are used for initializing the genetic pattern search process.

2.1.1. Extraction of regions of interest.

A simple and quick segmentation is done on the target image, in order to establish Regions of Interest (ROI). These are zones where the selected model has a relevant probability of being found. The data used for this segmentation is extracted from a database, previously constructed with the segmentation and correlation related information for the models of interest.

Segmentation is done by thresholding in HLS space followed by some morphologic transformations. Figure 2 shows a physical interpretation of this space. The *hue* component specifies the "perceptual" color property of the considered pixel, and is highly independent of illumination conditions (ambient light, shadows, ..). The *saturation* component indicates how much of the particular color does the pixel have or, in other words, how far from gray scale (S=0) is the pixel. The *luminance* component can be interpreted as the gray scale version of the image, as it measures the amount of light which has arrived at each pixel. The strongest drawback of HSL space is that hue is ill-defined with S→0; therefore great care should be taken to avoid using hue in low saturation regions. More detailed information about HSL space can be found in [7].

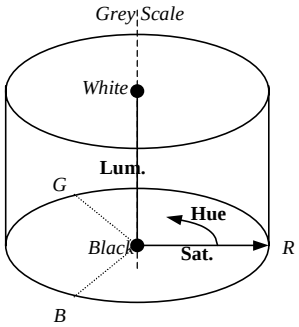
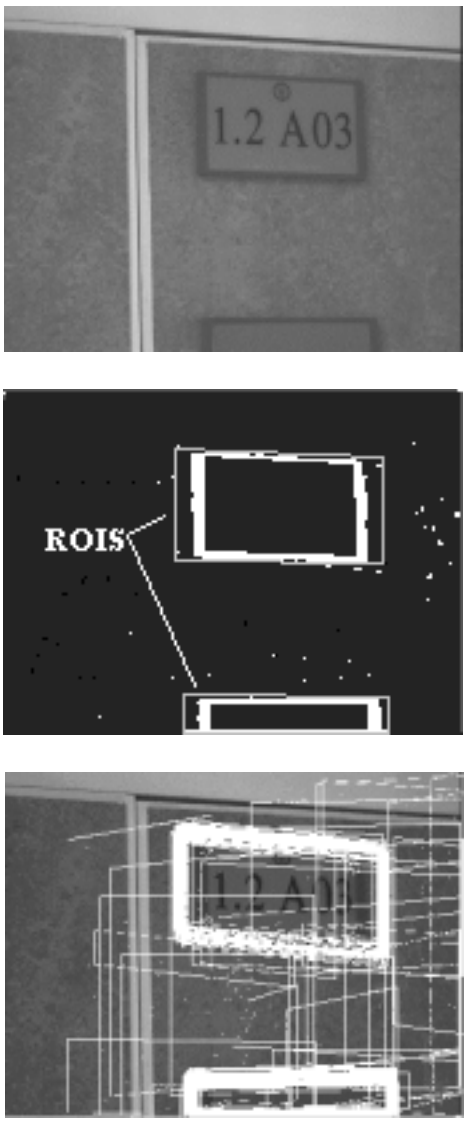


Figure 2. Cylindrical interpretation of HSL space.

In this segmented image, blobs with the appropriate geometry are select as ROI; this ROI may be considered as model location hypotheses. ROI marked in a real example are shown in Figure 3 (b).



(a) (b) (c)

Figure 3. a) original image, b) ROIs, c) model search.

2.1.2. Pattern Search.

Next, a Genetic Algorithm (GA) is used to confirm or reject the ROI hypotheses. Each individual's genome is made of 5 genes (or variables): the individual's cartesian coordinates (x, y) in the image, its horizontal and vertical size in pixels (ΔX, ΔY) and a measure of its vertical perspective distortion (SkewY). Figure 4 shows the geometrical interpretation of an individual.

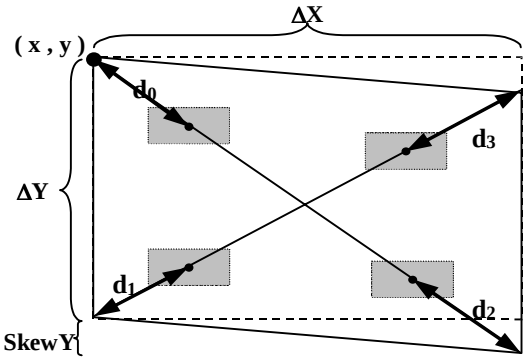


Figure 4. Individual's geometric interpretation.

The individual's health is estimated by a fitness function using the normalized correlation results (on the luminance component of the target image) of four little pattern-windows. These pattern-windows are located at fixed distances along the geometric diagonals of the individual (Figure 4). The correlation for each window ρ_i is calculated only in a very small (about 7 pixels) neighborhood of the pixel in the target image which matches the pattern-window's center position, for real-time computation purpose.

The use of four small pattern-windows has enormous advantages over the classical use of one big pattern image for correlation. The relative position of the pattern-windows inside the individual can be modified during the search process. This idea is the basis of the proposed algorithm, as it makes it possible to find landmarks with very different apparent sizes and perspective deformations in the image. Furthermore, the pattern-windows for one landmark does not need to be rotated or scaled before correlation (assuming that only perspective transformation are present), due to their small size. Finally, computation time for one search is much lower for the correlation of the four pattern-windows than for the correlation of one big pattern.

The selected fitness function is:

$$error = \left(\frac{3 - (\rho_0 * \rho_1 + \rho_2 * \rho_3 + \rho_0 * \rho_1 * \rho_2 * \rho_3)}{3} \right)^K \qquad (1)$$

$$Health = \frac{1.0}{0.1 + error} - K, \qquad 0 \le K < 0.909$$

where ρ_i are the pattern-windows correlations and K is a constant which affects the "gravity force" of dominant individuals (that is, the attraction of dominant individuals over the rest of the population). This parameter is dynamically changed in order to keep a good balance between exploration of new image regions (dispersing individuals) and refining the match found (concentrating individuals). This change is a function of the best individual health, preventing this dominant individual from attracting too much the other ones. The selected fitness function includes one term that takes into account the four pattern-windows correlations, for encouraging the individuals to match the full extent of the object in the image. On the other hand, there are two terms with the product of correlations of pattern-windows belonging to the same side of the individual; this allows the fitness of the individual not to drop drastically if partial occlusion is present. However, a non-occluded mark will always have better correlation than a partially occluded one.

The relative position of the diagonals (and the pattern-windows attached to them) change with the individual's deformation, in order to match the perspective distortion of the objects in the image. The individual's deformation is achieved by the variations of ΔX, ΔY and SkewY. The location of the pattern-windows inside an individual is shown in Figure 4 highlighted in gray. Here, the adimensional values d_0 ,…, d_4 are fixed. They are obtained as the ratio between the distance from the corner of the individual to the center of the window and the diagonal's size. This pattern-windows are selected off-line in the relevant locations and stored in the model's database for their further use. In Figure 5 the pattern-windows for the green circle and the office's nameplate models are shown. These windows have been taken from real images; the only restriction is that the window centers must be on the diagonals of a parallelepiped.


Figure 5. Pattern-windows.

Although four pattern-windows is the minimum number which ensures that the individual covers the full extent of the object in the image, a higher number of pattern-windows can be used if needed for more complex landmarks without increasing significantly computation time.

The GA is initialized by centering a big part of the total population around the previously established ROI; this allows working with very few individuals (less than a hundred) and a quick convergence if the ROI is really a valid hypothesis (one or two generations are usually enough). On the other hand, a false hypothesis makes the population go away from the ROI and to explore other image regions. Finally, the individual's health is a good measure of the match certainty. Therefore, if the health of any of the individuals goes over the *certainty threshold*, it is taken as a valid match, while the lack of convergence in a few generations should be considered as if there is no

reliable enough model in the target image. Figure 6 represents the health of an individual versus the average correlation of its four pattern-windows. Two thresholds have been empirically selected. When a match reaches the *certainty threshold*, the search ends with a very good result; on the other hand, any match must have an average correlation over the *acceptance threshold* to be considered as a valid one.

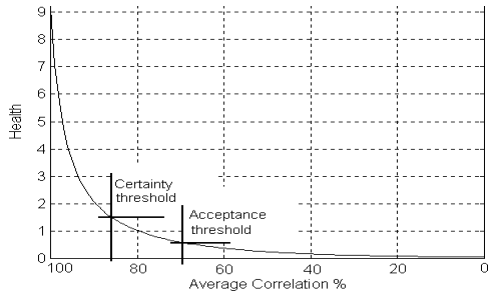


Figure 6. Health vs. average correlation.

In Figures 3 and 7, a) shows a real image, b) shows the selected ROI and c) presents a snapshot of the GA's evolution (each white polygon marks one individual), for the two landmarks used, in a real exercise.

2.2. Extraction of landmark associated information.

If a new landmark is found in the target image, then the relevant information for the localization process is extracted. For topological navigation, often the only information needed from a landmark is its presence or absence in the robot's immediate environment. However, more information may be needed for other navigation strategies, regardless of their topologic or geometric nature. For general application, the centroid, size and perspective distortion of each landmark are extracted. Furthermore, if the landmark found is an office's nameplate, the next step is *reading* its contents. This ability is widely used by humans, and other research approaches have been done recently in this sense [10]. In our work, a simple Optical Character Recognition (OCR) algorithm has been designed for the reading task, briefly discussed below.

2.2.1. OCR Engine.

Once an office's nameplate is found in the target image, its inner region (which contains the office's label) is extracted from the mark. Then this inner region is segmented using an adaptive threshold. Blobs which possibly correspond to individual characters are then selected. Each of this blobs is analyzed in order to ensure it has the right size: relatively big blobs (usually means some characters merged in the segmentation process) are split recursively in two new characters, and relatively small blobs (fragments of characters broken in the segmentation process, or punctuation marks) are merged to one of their neighbors. Then this blob-characters are grouped in text lines, and each text line is split in words (each word is then a group of one or more blob-characters).

Finally, the blobs are normalized in size and sent to a classifier (a feed-forward neural network with one hidden layer) which assigns the corresponding alphanumeric label to each blob, and provides an estimation of the recognition's confidence. This simple network has proved to have a very good ratio between recognition ability and speed compared to more complex neural networks. It has also proved to be more robust than conventional classifiers (only size normalization of the character patterns is done, the neural network handles the possible rotation and skew). This network is trained offline using the *quickpropagation* algorithm, described in [5].

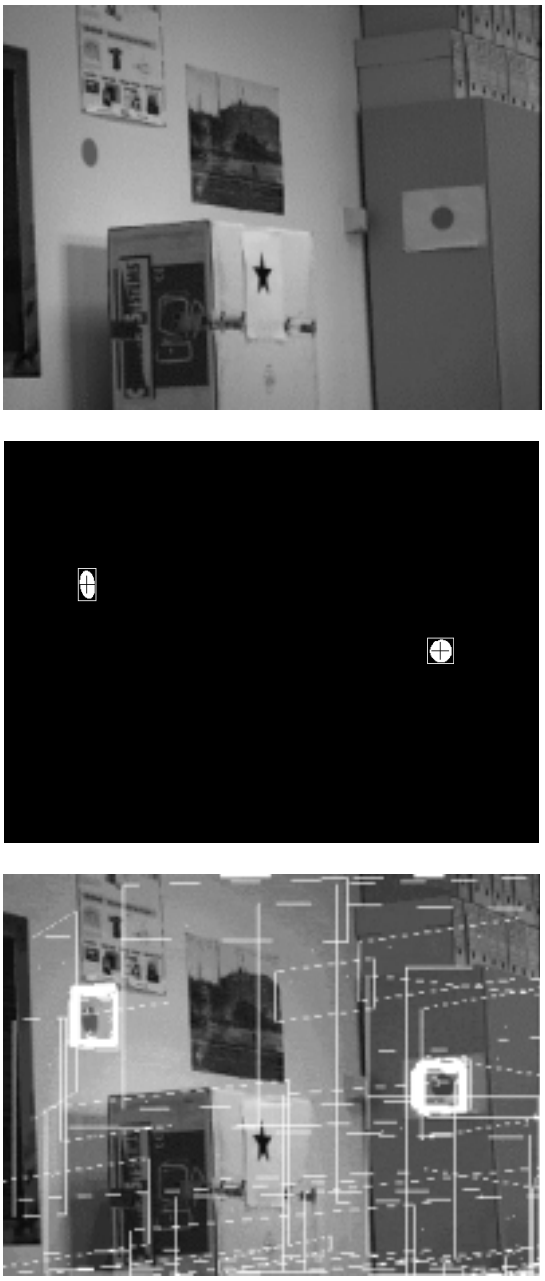


Figure 7. Search of an artificial landmark.

The recognized text string (with '?' characters representing the ones with insufficient certainty) is sent to the location module to be compared with the known office text labels. Figure 8 a) shows the inner region of an office's nameplate found in a real image; in b) blobs considered as possible characters are shown, and in c) binary size-normalized images, that the neural network has to recognize, are included. In this example, recognition confidence is over 85% for every character.

Figure 8. OCR stage.

3 Experimental results

Experiments have been conducted on a B21-RWI mobile vehicle (Figure 9), in the facilities of the System Engineering and Automation Division at the Carlos III University [1]. This implementation uses a JAI CV-M70 progressive scan color camera and a Matrox Meteor II frame grabber plugged inside a standard Pentium III personal computer mounted onboard the robot.

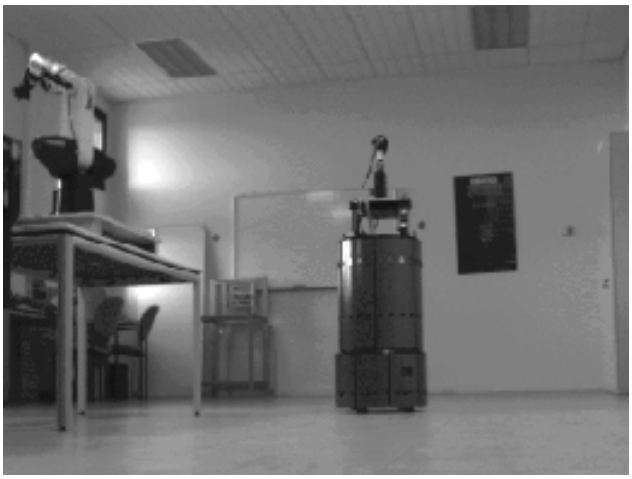


Figure 9. B21 mobile robot.

The pattern recognition stage has shown good robustness with both landmarks used. Figure 10 summarizes some of the test results. The curves show the average correlation obtained with the landmarks situated at different distances and angles of view from the robot, under uncontrolled illumination conditions.

A "possible recognition" zone in the vicinity of any landmark can be extracted from the data on this plot. This means that there is a very good chance of finding a landmark if the robot enters inside the oval defined by angles and distances over the acceptance threshold line in the graph. These results were obtained using a 25 mm fixed optic. When a motorized zoom is used with the camera, it is possible to modify the recognition zone at will.

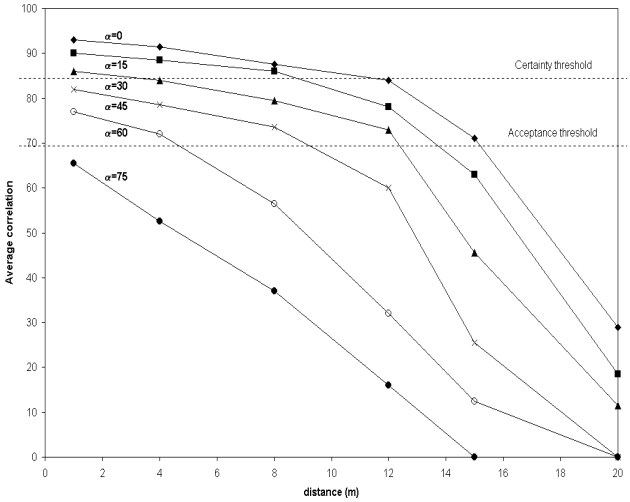


Figure 10. Recognition results.

The robot is able to localize itself successfully using the standard University's nameplates, and using the artificial landmarks placed in large rooms (Figure 9). The ability of reading nameplates means that there is no need for the robot initial positioning. The robot can move around searching for a nameplate and then use the text inside to realize its "absolute" position. The system can actually process up to 4 frames per second when searching for a landmark, while the text reading process requires about half a second to be completed (once the plate is within range).

Since the nameplates can be detected at larger distances and angles of view than those minimum needed for successfully reading its contents, a simple approach trajectory is launched when the robot detects a plate. This approach trajectory does not need to be accurate since, in practice, the text inside plates can be read with angles of view up to 45 degrees. Once this approach movement is completed, the robot tries to read the nameplate's content. If the reading is not good enough, or the interpreted text is not any of the expected, a closer approach is launched before discarding the landmark and starting a new search.

In Figure 11 a real situation is presented. Nine artificial landmarks are placed inside room 1 and four natural landmarks are situated along the hall. The frame captured by the camera (25 mm focal distance and 14,5º horizontal angle of view) is shown in Figure 7 a), where two artificial landmarks are successfully detected after only one iteration of the genetic search.

Figure 12 illustrates the case where both kinds of landmarks were present in the captured image; in this case two runs of the algorithm were needed to identify both landmarks.

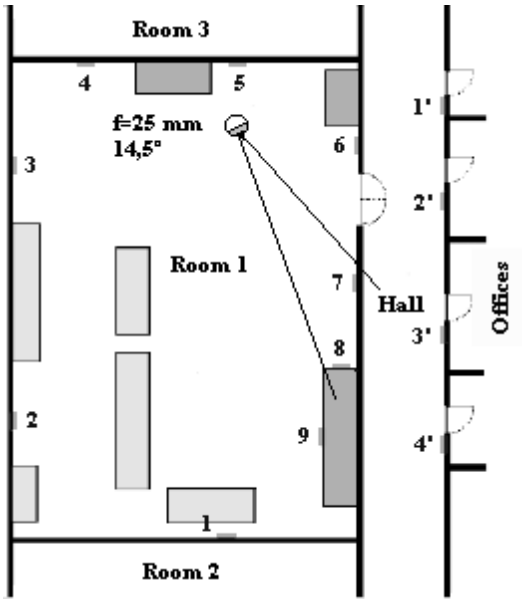


Figure 11. Real mission example 1.

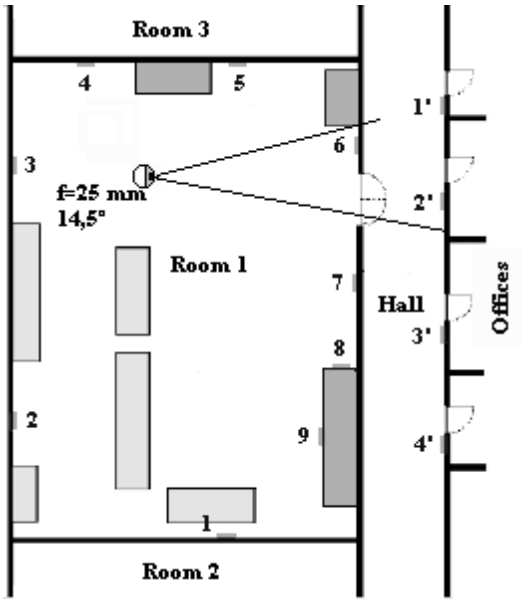


Figure 12. Real mission example 2.

4 Conclusions

In this paper a new color vision-based landmark recognition system for mobile robots is presented. It has been shown that the proposed technique is able to work with both artificial and natural landmarks, which can contain written text. In this case, the text can be read and used later for any task, not only localization but also in general decision making processes. The system can be easily adapted to handle different landmarks (updating the database with the new pattern-windows) and text styles (retraining the classifier weights). The natural application environments of the system are big public buildings and industrial buildings (factories, stores) where the pre-existent wall signals may be used.

Acknowledgments

The authors gratefully acknowledge the funds provided by the Spanish Government through the CICYT project TAP99-0214.

References

[1] Armingol, J.M; Moreno, L.; Escalera, A. de la; Salichs, M.A. (1998) *"Landmark Perception Planning for Mobile Robot Localization"*. 1998 IEEE International Conference on Robotics and Automation, vol. 3, pp. 3425-30

[2] Balkenius, C. (1998) *"Spatial learning with perceptually grounded representations"*. Robotics and Autonomous Systems, vol. 25, pp. 165-175.

[3] Beccari, G., Caselli, S. Zanichelli, F. (1998) *"Qualitative spatial representations from task-oriented perception and exploratory behaviors"*. Robotics and Autonomous Systems, vol. 25, pp. 165-175.

[4] Borenstein J.; Feng L. (1996) *"Measurement and correction of systematic odometry errors in mobile robots"*. IEEE Transactions on Robotics and Automation, vol. 12, n 5.

[5] Fahlman, S. E. (1998) *"An empirical study of learning speed in back-propagation networks"*. CMU-CS-88-162.

[6] Franz, Matthias O. (1998) *"Learning view graphs for robot navigation"*. Autonomous robots, vol. 5, pp. 111-125

[7] Perez, F.; Koch, C. (1994) *"Toward color image segmentation in analog VLSI: algorithm and hardware"*. International Journal of Computer Vision, 12:1 pp. 17-42.

[8] Rosenfeld, A. (2000) *"Image analysis and computer vision: 1999 [survey]"*. Computer Vision and Image Understanding. vol.78, no.2; pp. 222-302.

[9] Salichs, M.A., Moreno, L. (2000) *"Navigation of mobile robots: open questions"*. Robotica. vol.18, pp. 227-234.

[10] Tomono, M., Yuta, S. (2000) *"Mobile robot navigation in indoor environments using object and Character Recognition"*. IEEE International Conf. on Robotics and Automation. San Francisco, pp. 313-320.