IAV2004 - PREPRINTS
5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles
Instituto Superior Técnico, Lisboa, Portugal
July 5-7, 2004

FEATURE EXTRACTION AND SELECTION FOR MOBILE ROBOT NAVIGATION IN UNSTRUCTURED ENVIRONMENTS

Alberto Vale * **José Miguel Lucas** **
M. Isabel Ribeiro *

* *Institute for Systems and Robotics / Instituto Superior Técnico, Av. Rovisco Pais 1, 1049 - 001 Lisboa, Portugal*
{vale,mir}@isr.ist.utl.pt
** *University of Murcia, Dept. of Information and Communication Engineering, Campus de Espinardo, 30100 Espinardo (Murcia), Spain, jmlucas@um.es*

Abstract: Feature extraction and selection are important steps in the construction of a map to support the navigation of mobile robots in outdoor environments. The large amount of data acquired by the on-board sensors has to be reduced to retain only the crucial information for the navigation purpose. This procedure should be robust given the rough, dynamic and unpredictable conditions provided by outdoor scenarios. The paper discusses the use of different types of features and proposes a feature selection criteria aiming at building a topological map of the environment. The approach was tested with real data acquired in different types of outdoor scenarios.

Keywords: Feature Extraction, Feature Selection, Outdoor Environments, Mobile Robots

1. INTRODUCTION

One application of mobile robots is to carry out tasks in unstructured environments often without or with a reduced a priori knowledge of the scene map. To accomplish this type of mission, mobile robots have to adapt, recognize, localize and navigate simultaneously, while moving towards the desired target goal.

If no environment representation is available, mobile robots have to build it based on the observations acquired by on-board sensors. Large scenarios, as those in outdoor environments, rides to a large amount of information to store, including that required to accomplish the map, but also the localization and navigation algorithms. Topological representations, as the ones considered in this paper, are based on landmarks characterized by features. Therefore a feature extraction procedure, reducing the data acquired by the sensors but retaining the crucial information, is required. Features have to support different scenarios but not every type of feature is essential to a particular scenario, this requiring a feature selection criteria.

In natural scenes, there are several features: corners, distinctive features such as buildings, streetlamps, placards, vertex and lines junctions, colors (Kasprzak and Szynkiewiez, 2003), textures, vertical edges (Kang and Jo, 2003). The key question is how to determine the best set of features in an outdoor scene, aiming at producing a topological map that supports the navigation of a mobile robot.

In this paper we address the problem of feature extraction and selection to build a topological map. Given a set of images acquired by the on-board sensors, the paper's novelty is the choice of the best features, according to a statistical criteria, that fits on the scenario representation. These features are used in the mapping algorithm described in (Vale and Ribeiro, 2003), that supports the localization and navigation of a mobile robot in an outdoor environment. Our approach uses the data acquired by a standard camera installed on top of the mobile robot. It can also be applied with a different vision system geometry, namely an omni-directional camera.

This paper is organized as follows. Section 1 presents the paper motivation. In Section 2 an overview of related work on features and world representations is described. Section 3 introduces the notation and the algorithms proposed for feature extraction. The feature selection procedure is described in Section 4. Experimental results obtained with images acquired in a real outdoor environment are presented in Section 5. Section 6 concludes the paper and presents directions for further developments.

2. RELATED WORK

To obtain a topological representation from a real environment, it is necessary to perform a feature extraction process from data acquired by on-board sensors (e.g., vision camera, laser range finder). Some works deal with this issue.

Santos-Victor *et al.*(Santos-Victor and Bernardino, 2002) proposed a vision-based navigation which takes into account special spatial representations and visual geometries. The navigation problem is based on the decomposition of sub-goals, identified by recognizable landmarks.

Ulrich *et al.*(Ulrich and Nourbakhsh, 2000) presented an appearance-based place recognition system for topological localization. This work focuses on color images to distinguish the places. Hähnel *et al.*(Hähnel *et al.*, 2003) discussed the problem of creating maps in dynamic environments, using a technique to identify dynamic objects.

Zhou *et al.*(Zhou and Huang, 2001) proposed structural features for content-based image retrieval (CBIR), especially edge/structure features extracted from edge maps. They describe a new algorithm to extract information embedded in the edges. Experiments show that the new features can catch salient edge/structure information and improve the retrieval performance.

The Principal Component Analysis (PCA) obtains a feature set and provides a mathematical model for the loss of information (Jolliffe, 1986). It supplies a linear representation of the original data using the least number of components with the minimum mean-squared error. PCA has been successfully used in several robotic applications for finding linear features from intensity data. It can also be applied on laser range data (Wallner *et al.*, 1998). Thrun (Thrun, 1998) presented a method to learn what features/landmarks are best suited for localization, using neural networks. Vlassis *et al.*(Vlassis *et al.*, 2001) proposed a method for an appearance based modeling of the environment, using linear image features extracted using PCA.

Our paper addresses some of the issues described in the reviewed literature. In particular, we take the best characteristics of different types of features (e.g., vertical edges, color histograms, PCA) and propose the choice of the best features for mapping, based on a statistical criteria.

3. FEATURE EXTRACTION

A feature extraction procedure corresponds to the projection of high-dimensional data onto a low dimensional subspace leading, in most cases, to a loss of information. Any feature extraction method must satisfy the following properties, (Vlassis *et al.*, 2001): *i)* robustness to small displacements, *ii)* invariant to lighting conditions, *iii)* invariant to occlusion, *iv)* fast computation and *v)* capacity to compress the images as much as possible while retaining pertinent information.

Different feature types can be used to solve the mobile robot localization problem, in particular geometrical features (lines, corners, edges, shapes), color, textures and whatever can be discriminated as a landmark. Given the huge amount of information but also the requirement to support localization it is necessary to choose the best features to represent a landmark and to carry out a feature selection procedure.

An important goal to support robot navigation is to achieve a good and optimized representation of features to improve the performance of the matching required in the mapping procedure, as described in (Vale and Ribeiro, 2003).

The notation used throughout the paper is:

- o_t is a p-dimensional observation vector acquired at time instant t,
- f_t is a m-dimensional vector of features referred to time instant t, where $f_t(i)$ is the i-th feature value, $i = 1 \dots m$,
- $f = \{f_{t_0}, f_{t_1}, \dots, f_{t_N}\}$ is a sequence of feature vectors from t_0 to t_N.

The observation o_t integrates all the sensor information available. The feature vector, f_t, is extracted at each time instant t from the observation data o_t by a nonlinear function FE, $f_t = FE(o_t)$, where $FE : I\!R^p \longrightarrow I\!R^m$. The extraction function FE reduces the amount of data, retaining only the essential information of sensor data. For that reason, $FE^{-1}(f_t) \supset o_t$, which means that different observation vectors could lead to the same feature. When this happens, it is important to identify if the observations were acquired in the same place, or in places where the distinction among them is not important. The performance of FE is addressed in Section 4, where the best features, which are time independent, are chosen.

3.1 Edges and Hough-Transform

As described in (Finlayson *et al.*, 1998), the image dependencies due to lighting source and illuminance, mainly in outdoor environments, require a color image normalization procedure. This drawback points towards the use of edge-based features to support environment representation and robot navigation (Mata *et al.*, 2003), (Lamon *et al.*, 2001), (Zhou and Huang, 2001), (Vlassis *et al.*, 2001).

To extract edges from an image a specific filter (e.g., Sobel, Prewitt, Roberts, Gaussian, or other) is applied. In outdoor environments where the scenario is unstructured, it is important to detect the vertical ones. Moreover, the edges present noisy information and therefore it is necessary to remove or, at least, reduce the superfluous data, applying the Hough Transform (HT) to the edges (Hansen and Andersen, 1997). This technique yields an histogram of straight lines for different directions, as shown in Fig. 1-c), where the brightness corresponds to the amount of pixels that belong to a specific line.

A straight line is defined by (r, θ), $x \cos\theta + y \sin\theta = r$, where (x, y) are the coordinates of an image pixel. To select only the vertical or near to vertical edges, the directions are chosen around 0 and 180 degrees, as exemplified in Fig. 1-d). The k_{Edges} straight lines with the larger number of pixels (high level on the histogram, represented by '+') are selected and considered as the edges' features extracted from the image. The image in Fig. 1-b) shows the result. To reduce the dependence of r_i from the sensor's orientation, we only record the distance between two consecutive straight lines, $d_i = r_{i+1} - r_i$, yielding

$$f_t^{Edges} = FE(o_t) = \begin{bmatrix} d_1 & d_2 & \cdots & d_{k_{edges}} \\ \theta_1 & \theta_2 & \cdots & \theta_{k_{edges}} \end{bmatrix}. \quad (1)$$

3.2 Histogram parameterizations

Even with the light and geometric dependencies, the color is still an important source of information. Applying a normalization procedure, as suggested by (Finlayson *et al.*, 1998), or simply, using the HSV colormap in spite of RGB, color histogram are important features. However, histograms provide large amount of information that could be parameterized as exemplified in Fig. 2. We tested the parameterization of Hue and Saturation histograms using a polynomium and a sum

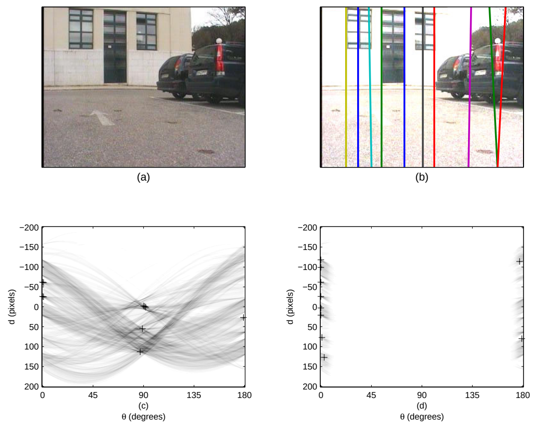


Fig. 1. Example of vertical edges detection. (a) original image; (b) vertical and near vertical edges; (c) the Hough Transform of edges; (d) the same as in c) but only around vertical directions

of Gaussian functions. A parameterization using a polynomium of order n requires $n+1$ parameters $(a_0, a_1, \ldots, a_n)$, while by a sum of n $\mathcal{N}(\mu, \sigma)$, requires $3n$ parameters (weights, means and variances). The parameterization error is evaluated by the square error of the original and the parameterized histograms. We carried out experimental tests with a large amount of images acquired in different places of outdoor environments. The corresponding Gaussian parameterization errors are significantly lower when the number of parameters are equal or larger then 6, as shown in Table 1.

♯ of parameters	Gaussians		Polynomials	
	H color	S color	H color	S color
3	3214	1087	2401	1284
6	2010	830	2200	960
9	1537	534	2035	681
12	714	478	1871	539

Table 1. Parameterization error using Gaussian and polynomial functions

According to the experimental results, the Gaussian parameterization yields better histogram representations for the considered outdoor scenarios. Consequently, in our work, the features extracted are the Gaussian parameterization: the weights c_i, the means μ_i and the variances σ_i,

$$f_t^{Hist} = FE(o_t) = \begin{bmatrix} \begin{bmatrix} c_1 & c_2 & \cdots & c_{k_{Hist}} \\ \mu_1 & \mu_2 & \cdots & \mu_{k_{Hist}} \\ \sigma_1^2 & \sigma_2^2 & \cdots & \sigma_{k_{Hist}}^2 \end{bmatrix}_{Hue} \\ \begin{bmatrix} c_1 & c_2 & \cdots & c_{k_{Hist}} \\ \mu_1 & \mu_2 & \cdots & \mu_{k_{Hist}} \\ \sigma_1^2 & \sigma_2^2 & \cdots & \sigma_{k_{Hist}}^2 \end{bmatrix}_{Sat.} \end{bmatrix}, (2)$$

where Hue and Sat. correspond to the Hue and Saturation components. The features extracted from the Parzen windows, (Parzen, 1962), return similar results and are computationally faster.

3.3 2D histogram and image segmentation

Based on histograms it is possible to identify regions on the image with similar colors. We per-

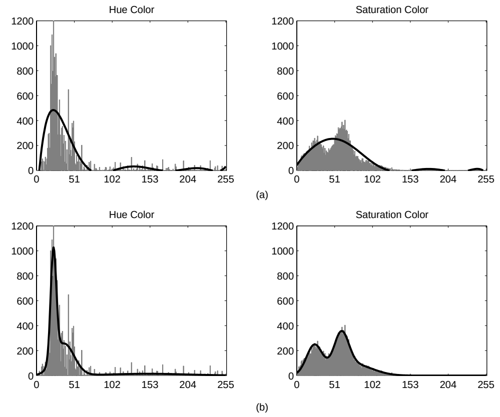


Fig. 2. Example of histograms parameterization using 9 parameters: (a)- by a polynomial of order 8; (b)- by 3 Gaussians

formed the bi-directional histogram along Hue-Saturation colors and selected the k_{2Dhist} most significant colors $(Hue, Saturation)$. For each significant color, the smallest boundary-box that fits all the pixels with the same color defines a region. The features extracted from each boundary-box are the width and height, the amount of pixels and the color, i.e.,

$$f_t^{2Dhist} = FE(o_t) = [box_1 box_2 \cdots box_{k_{2Dhist}}]$$
$$= \begin{bmatrix} width_1 & width_2 & \cdots & width_{k_{2Dhist}} \\ height_1 & height_2 & \cdots & height_{k_{2Dhist}} \\ pixels_1 & pixels_2 & \cdots & pixels_{k_{2Dhist}} \\ color_1 & color_2 & \cdots & color_{k_{2Dhist}} \end{bmatrix} \quad (3)$$

The position of the boundary-box on the image is not recorded, since it is much dependent on the point of view, (Kasprzak and Szynkiewiez, 2003).

3.4 PCA and ICA

A common approach to extract the essential information from images is the Principal Component Analysis (PCA), (Jolliffe, 1986). A similar technique where the components are orthogonal is the Independent Component Analysis (ICA), (Wachtler *et al.*, 2001). Both techniques extract a base, $B = \{B_1, B_2, \ldots, B_{k_{comp}}\}$, from a training set of images. We will refer each B_i as a component of the base.

Given the size of images, Wachtler (Wachtler *et al.*, 2001) proposes an implementation optimization dividing the images into sub-images. This is useful, since the original images present common areas (e.g., the ground, the sky), as illustrated in Fig. 3 and Fig. 5. Consequently, the training set increases according to the number of divisions (e.g., 4× if each image is sub-divided in 4 sub-images) and the base also changes.

The projection of the training set into each base B provides different energy distribution. The PCA results condense the energy into the first components (usually the first 2 retain more than 90%), as exemplified in Fig. 4, using the first 25 principal components of the base evaluated from the training set with the 12×16 images obtained by sub-dividing in 16 each of the 12 images of Fig. 3.

The features are the projection of the observed images, o_t, on the base, B, or equivalently:

$$f_t^{PCA} = FE(o_t) = \left[< o_t, B_1 > \cdots < o_t, B_{k_{comp}} >\right]. (4)$$

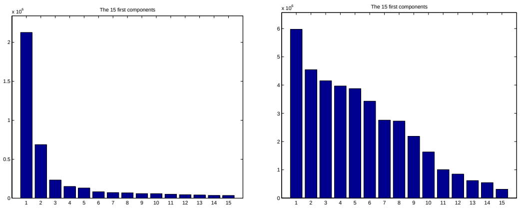
The features f_t^{ICA} are similarly extracted if the basis results from the ICA procedure. Both techniques, PCA and ICA, can be applied to the images in RGB or HSV format. However, the two colors Hue and Saturation are the most important as explained in Section 3.2.

Fig. 3. A training set of images

The relation between the basis, the number of components and the number of sub-images is non linear, as exemplified by the results in Table 2. This table presents the image reconstruction error using PCA and ICA, with 5 to 25 components. The columns correspond to the sub-divisions of the images (1-no division, 4,16-divides the image into 4 and 16 sub-images respectively, as illustrated in Fig. 5) with a training set of 12 images. The error is an average for all pixels (each pixel changes between 0 and 255). When the number of components increases, the error decreases. For instance, the reconstruction error is 0 when the number of components is larger than 12 since the training set has 12 images. However, the reconstruction using 5 or 10 components and images divided into 4 sub-images provides an error larger than the one obtained with images divided into 16 or not divided. For more than 10 components the error decreases. This results from the fact that, when the original images are divided into 16, the sub-images coincide with the ground, the sky or the buildings.

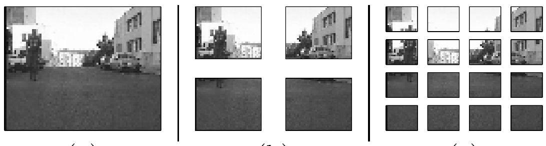
♯ of com-	PCA			ICA		
ponents	1	4	16	1	4	16
5	6.6	7.3	6.1	11.5	12.2	8.9
10	1.7	5.7	5.1	2.3	11.0	8.7
15	0	4.4	4.5	0	10.4	8.5
20	0	3.2	4.1	0	8.9	8.3
25	0	2.0	3.8	0	8.0	8.1

Table 2. Image reconstruction error using PCA and ICA



(a) (b)

Fig. 4. L_2 norms of the basis functions using (a) PCA (b) ICA



(a) (b) (c)

Fig. 5. (a)- The entire image; (b)- Image divided into 4 sub-images; (c)- Divided into 16

4. FEATURE SELECTION

As soon as features are extracted it is necessary to select the ones that will be used for mapping, this requiring a selection criteria. The quality of a feature has to be analyzed along two different perspectives: time/space and correlation with other features as illustrated in the following example. Consider that there are two types of features: "colors" and "geometric forms" and that the mobile robot navigates along three distinct places. If all the places are identified by the same color, the feature "color" is useless, independent of the "geometric" information. If the two first places are "red" (the same value for the feature "color") and the third place is "blue", the feature "color" can identify some places, but the ambiguity is still present. In this case, if the geometric form is the same for the "red" places, but different for the "blue" place, the two features are redundant, i.e., the correlation between features is too high.

To evaluate the quality explained above, the correlation between features is required. In the sequel we formalize this correlation concept. Let the set of features given by

$$f = \{f^{Edges}, f^{Hist}, f^{2Dhist}, f^{PCA}, f^{ICA}\}, \quad (5)$$

where f^a is extracted from $N+1$ observations, i.e.,

$$f^a = \{f^a_{t_0}, f^a_{t_1}, \ldots, f^a_{t_N}\} \quad (6)$$

with $a \in \{Edges, Hist, 2Dhist, PCA, ICA\}$.

Define μ_{f^a}, μ_{f^b} as $\mu_{f^i} = \frac{1}{N+1}\sum_{t=t_0}^{t_N} f^i_t, i = a, b$, let M be the $N_a \times N_b$ matrix,

$$M = \frac{1}{N+1}\sum_{t=t_0}^{t_N} \left[f^a_t - \mu_{f^a}\right]\left[f^b_t - \mu_{f^b}\right]^T \quad (7)$$

with N_a and N_b such that $f^a_t \in I\!R^{N_a}$ and $f^b_t \in I\!R^{N_b}$.

The correlation between two different types of features, f^a and f^b, with $a \neq b$, is evaluated as

$$corr(f^a, f^b) \propto \sum_{i=1}^{N_a}\sum_{j=1}^{N_b} |M_{ij}|\,. \quad (8)$$

For the map construction, the algorithm uses the n lowest correlated features as they convey the highest degree of environment information. The dimension of n is selected according to the desired accuracy of the mapping.

5. EXPERIMENTAL RESULTS

The experimental results were obtained from images acquired by a camera mounted on top of the mobile robot displayed in Fig. 6. The experiments were made in outdoor environments in three different scenarios: "Exp.1" is a mixed of buildings, trees, cars, walking people and diversity of light conditions; "Exp.2" was carried out in a garden with a couple of trees; "Exp.3" run in a parking place (a structured area). The features used were vertical edges, the Hue/Saturation-colors histograms parameterization (using 3 Gaussians), 2D histograms (the 4 first boundary-boxes), the PCA (images subdivided in 14 sub-images and building a base with 15 components) and ICA. Feature extraction yields data compression as rep-

Fig. 6. Mobile Robot ATRV-Jr with a Pan and Tilt camera Sony EVI-D31

resented in Table 3 for a set of 80 images (JPG) requiring 1055KBytes of storage space. The *edges, histograms and 2D histograms* yields higher compression than PCA or ICA, since the two last features require a base of images.

Edges	Hist	2Dhist	PCA	ICA
98.83%	98.94%	99.11%	85.80%	85.80%

Table 3. Compression from observations to features

The correlation differs between experiences and features. The edges have low correlation with other features in Exp.1 and 3, since the buildings contain very well defined edges, while in Exp.2, the trees and the waved terrain provides rough edges. The correlation between any feature extracted with PCA or ICA is small except between themselves (PCA and ICA). This exception is also

	Exp.1	Exp.2	Exp.3
Edges & Hist	0.0492	0.513	0.265
Edges & 2Dhist	0.1944	0.442	0.226
Edges & PCA	0.189	0.474	0.409
Edges & ICA	0.013	0.288	0.074
Hist & 2Dhist	0.138	0.381	0.569
Hist & PCA	0.237	0.236	0.110
Hist & ICA	0.033	0.098	0.108
2Dhist & PCA	0.213	0.472	0.296
2Dhist & ICA	0.058	0.315	0.093
PCA & ICA	0.832	0.452	0.861

Table 4. Correlation between features

verified between PCA (or ICA) and 2Dhist of Exp.2, since the boxes of similar colors coincides with some image components. According to these results, the Edges and Hist or Edges and ICA are the less correlated features in Exp.1, while Hist and ICA, Hist and PCA are the less correlated features in Exp.2 and Exp.3, respectively.

From the results presented in Table 4 we conclude that the edges and ICA or Edges and Hist (or even the three type of features simultaneously) are the best features to build a map in Exp.1. The Hist and ICA or Hist and PCA, but not the three simultaneously (since PCA and ICA are high correlated) should be used to build a map in Exp.2. Similarly for edges and ICA or 2Dhist and ICA in Exp.3.

Each experience exploits a good combination of two (or more) types of features corresponding to the lowest correlated features. The work (Vale and Ribeiro, 2003) describes the clustering technique of the features selected along the lines explained in this paper, aiming a topological representation of the environment.

6. CONCLUSIONS AND FUTURE WORK

Future research includes texture extraction using Gabor Filters and/or Nonlinear Operator and a tuning criteria for the dimension of each feature (e.g., optimize the appropriate number of edges, k_{edges}, or the number of gaussians to parameterize the Hist, k_{hist}, or the number of boundary-boxes, k_{2Dhist}). However, non correlated or weakly correlated features are not necessarily essential features. Different observations in the same place (or near) must lead to similar features. To compare the features' vector from different observations, f_i and f_j, it is also necessary to define a distance criteria and improve the feature selection.

Future work also includes the integration of the feature extraction/selection procedure into the mapping algorithm, described in (Vale and Ribeiro, 2003) in the frame of project RESCUE. According to the goals of the project, it would be interesting to take advantage of using distributed feature extraction and selection to achieve cooperative localization and mapping in outdoor environments using several mobile robots.

ACKNOWLEDGMENT

Work supported by the FCT project RESCUE - Cooperative Navigation for Rescue Robots (SRI/32546/99-00) under Programa Operacional Sociedade de Informação (POSI) in the frame of QCAIII. The first author acknowledges the Ph.D. grant SFRH/BD/929/2000 from FCT. The second author acknowledges the Ph.D. grant supported by the project TIC2001-0245-C02-01 from Spanish Science and Technology Ministry.

REFERENCES

Finlayson, Graham D., Bernt Schiele and James L. Crowley (1998). Comprehensive colour image normalization. *Lecture Notes in Computer Science Journal,* pp. 475–490.

Hähnel, Dirk, Rudoph Triebel, Wolfram Burgard and Sebastian Thrun (2003). Map building with mobile robots in dynamic environments. *Proc. IEEE ICRA,* pp. 1557–1563.

Hansen, Klaus and Jens Damgaard Andersen (1997). Understanding the hough transform: Hough cell support and its utilization. *Journal IVC* **15**, 205–218.

Jolliffe, I. T. (1986). *Principal Component Analysis.* Springer-Verlag. New York.

Kang, Hyun-Deok and Kang-Hyun Jo (2003). Self-localization of mobile robot using omnidirectional vision. *Proc. 7th Korea-Russia Int. Symposium.*

Kasprzak, W. and W. Szynkiewiez (2003). Using color image features in discrete self-localization of a mobile robot. *Int. IEEE Conference on Methods and Models in Automation and Robotics,* pp. 1101–1106.

Lamon, Pierre, Illah Nourbakhsh, Björn Jensen and Roland Siegwart (2001). Deriving and matching image fingerprint sequences for mobile robot localization. *Proc. IEEE ICRA, Seoul, Korea,* pp. 1609–1614.

Mata, M., J. M. Armingol, A. de la Escalera and M. A. Salichs (2003). Using learned visual landmarks for intelligent topological navigation of mobile robots. *Proc. IEEE ICRA,* pp. 1324–1329.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics* **33**, 1065–1076.

Santos-Victor, José and Alexandre Bernardino (2002). Vision-based navigation, environmental representations and imaging geometries. *Robótica, Portugal* **47**, 21–29.

Thrun, Sebastian (1998). Bayesian landmark learning for mobile robot localization. *Machine Learning* **33**(1), 41–76.

Ulrich, Iwan and Illah Nourbakhsh (2000). Appearance-based place recognition for topological localization. *Proc. IEEE ICRA, San Francisco,* pp. 1023–1029.

Vale, Alberto and M. Isabel Ribeiro (2003). Environment mapping as a topological representation. *Proc. 11th ICAR, Coimbra, Portugal,* pp. 29–34.

Vlassis, Nikos, Yoichi Motomura, Isao Hara, Hideki Asoh and Toshihiro Matsui (2001). Edge-based features from omnidirectional images for robot localization. *Proc. IEEE ICRA, Seoul, Korea,* pp. 1579–1584.

Wachtler, Thomas, Te-Won Lee and T. Sejnowski (2001). The chromatic structure of natural scenes. *Journal of the Optical Society of America* **18**(1), 65–77.

Wallner, Frank, Bernt Schiele and James L. Crowley (1998). Position estimation for a mobile robot from principal components of laser range data. *Proc. IEEE ICRA, Leuven, Belgium,* pp. 3121–3128.

Zhou, Xiang Sean and Thomas S. Huang (2001). Edge-based structural features for content-based image retrieval. *Pattern Recognition Letters,* pp. 457–468.