LEARNING VISUAL LANDMARKS FOR MOBILE ROBOT NAVIGATION

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 learning and recognition system for use in mobile robot navigation tasks. The system uses genetic algorithms (GA) for both learning and recognition processes. It is able to learn new landmarks with very little human intervention. The recognition system can read text inside landmarks, when present.

This learning ability has been tested with two very different landmarks that have been successfully used for indoor topological robot navigation. In addition, some new landmarks are learnt, that will be tested for indoor-outdoor navigation in future works.

The presented experimental results show the performances of the proposed algorithm. Copyright © 2002 IFAC.

Keywords: Landmark-based Navigation, Mobile Robots, Computer Vision, Genetic Algorithms, Learning systems

1 INTRODUCTION.

A mobile autonomous robot must have a reliable navigation system for avoiding objects in its path and recognizing important objects of the environment to identify places. Machine learning techniques are being applied with remarkable success to several problems of computer vision and perception, but most of these problems have been fairly simple in nature and still cannot handle with real-time requirements (Ohyama, 1995; Bhandarkar et al., 1997; Luo, Potlapalli, 1994). The difficulty with scaling up to complex tasks is that inductive learning methods require a very large number of training patterns in order to generalize correctly from high density sensor information (like video cameras). However, recent results in mobile robot learning have demonstrated that robots can learn simple objects identification from very little initial knowledge information in restricted environments (Mahadevan, Teocharous, 1998; Bin-Ran et al. 1998).

Topological navigation allows overcoming some of the classical problems of geometric navigation in mobile robots, such as simultaneously reducing the uncertainty of localization and perception of the environment. On the other hand, topological navigation is heavily dependent on a powerful perception system to identify elements of the environment. The landmarks chosen for recognition 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 (Balkenius, 1998) 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 (Franz, 1998) has developed a vision based system for topological navigation in open environments. This system represents selected places by local 360° 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. (Beccari et al., 1998) 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 learning and recognition system for topological navigation. Here, 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, several different landmarks have been learned. Two of them have been tested in a mobile robot navigation application, employing the same searching algorithm: an artificial landmark (green circles placed on the walls) and a natural landmarks (office’s nameplates attached at the entrance of each room), shown in Figure 1 a) and b). All of them has been automatically learned by the system, with very little human intervention (only several training images with the landmarks to learn remarked must be provided). The searching and learning algorithm are based on genetic techniques for pattern recognition in digital images. The genetic searching algorithm is able to handle landmark’s perspective deformation problems. Finally, although the presented system is being tested in mobile robot topological navigation, it is general enough for its direct use in a wide range of applications, such as geometric navigation, inspection and surveillance systems etc.

![Figure 1. Some of the landmarks learned.](image)

### 2 PATTERN RECOGNITION METHOD

Pattern search is done using the general application 2D Pattern Search Engine presented in (Mata et al., 2001), briefly explained below.

#### 2.1. 2D pattern search engine

There is a large collection of 2D pattern search techniques in the literature (Rosenfeld, 2000). 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.

**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 learned database. This database is previously constructed during the training stage, including segmentation and correlation related information for the models of interest.

Segmentation is done by thresholding in a corrected HLS space, followed by some morphological transformations. Pixels with very low saturation values are marked as ‘achromatic’, and they are taken into account separately in the segmentation process. Upper and lower thresholds for each component are selected during the training phase.

In the 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 2 (b).

![Figure 2. a) original images, b) ROIs, c) model search.](image)
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 3). The correlation for each window $\rho_i$ 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, allowing real-time computation.

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 $\Delta X$, $\Delta Y$ and SkewY. The location of the pattern-windows inside an individual is shown in Figure 4 highlighted in gray. Here, the adimensional values $d_0$, $\ldots$, $d_4$ 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 and their positions are selected in the training process and stored in the model’s database for their further use. In Figure 4 the pattern-windows learned for the green circle and the office’s nameplate models are shown. These windows have been autonomously taken from real images.

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. This quick convergence of rejection makes mutation of little use. However, mutation is still necessary when using the system for an intensive search in the whole image, for example in map generation tasks if no ROI are found. In Figures 2 a) shows real images, 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.

The GA parameters and their values are the usual ones. Only the low number of individuals is somehow special. However, when no ROI are found and still the whole image has to be explored, this number must be increased.

2.2. Extraction of landmark associated information.

When a new landmark is found in the target image, 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 (Tomono, Yuta, 2000). In our work, a simple Optical Character Recognition (OCR) algorithm has been designed for the reading task. It uses a multilayer backpropagation neural network to classify segmented blobs (those suitable to be characters). More detail on this subject can be found in (Mata et al. 2001).

3 LEARNING RECOGNITION’S PARAMETERS.

For the learning process, the human teacher must provide several training images, where the landmark(s) to learn are bounded by rectangular boxes (this boxes will be referred as target boxes from now on). There are no a priori restrictions for the set of training provided. Of course, the wider the conditions this set of images covers (illumination, background, perspective distortions, etc), the best results will obtain the learned parameters on real situations.

As previously established, the recognition process can be sequentially divided in two steps: candidate hypotheses generation (through segmentation) and hypotheses verification (using correlation). Consequently, the learning process for a new landmark is too divided in two stages. In the first step, thresholding levels for segmentation are found. The

![Figure 3. Individual’s geometric interpretation.](image3)

![Figure 4. Pattern-windows.](image4)
second step is dedicated to determine the location of the correlation-windows inside one individual.

### 3.1. Learning segmentation parameters.

In the first training step, the system has to learn the best threshold values for the segmentation of the landmark. Upper and lower thresholds for Hue, Saturation and Luminance components must be estimated. This six values will conform the genome of the individuals of a GA that will search through the training image space.

The fitness function for the GA must encourage that the segmented regions each individual generates matches the target boxes defined in the training images. On the other hand, there should not appear segmented regions outside the target boxes, if possible. The ideal segmentation result should be a binary black image with the target boxes corresponding zones in white. This "ideal image" can be matched with the binary image resulting from the individual’s genome, using a XOR logical function. Pixels that survive this operation are missclassified, since they have been included in segmented regions while they do not have to (or its inverse). The count of white pixels after the XOR pass is then a good measure of the segmentation error for the considered training image. The total segmentation error for one individual is obtained by repeating this operation for all the training image set and accumulating the missclassified pixels in each image. The fitness function is then chosen as an inverse function of this total error.

Previously to the learning run, a coarse initialization of the GA is done, in order to decrease search time. A set of initial H, L and S threshold values are obtained from any of the training images, using local histograms. Two sets of three histograms (one for each component) are computed from the inner and outer regions adjacent to the target box’ boundaries (fig. 5). The inner histograms contain information from the landmark, the background, and noise. The outer histograms contain information from the background, other objects and noise. For each component, its outer histogram is subtracted from the corresponding inner histogram, with negative values forced to zero. The resulting difference histogram (denoted \( LDH \), from Local Difference Histogram) will contain only information belonging to the desired landmark and not present in the outer region. Finally, initialization values for thresholding are taken from a peak search over the LDH (various peaks are selected for each component).

This way several values for H, L and S thresholds are estimated, and their possible combinations generate a good part of the GA’ s initial population. The rest of the population is randomly generated. This initialization speeds up considerably the training process; training time results in the order of five seconds by training image (as will be described later).

![Figure 5. LDH regions.](image)

### 3.2. Learning correlation windows.

The second training step deals with the location of the four correlation-windows inside the individual. Once an individual is situated over a ROI (or a target box in a training image), the problem translates to find the adimensional values \( d_0, \ldots , d_4 \). Again a GA is used to find these four values (that will compose each individual’s genome).

Once again, a coarse GA initialization can be easily done in order to decrease training time. Intuitively, the relevant positions where the correlation-windows should be placed are those having strong local variations in the image components (H, L and/or S). A simple method is used to find locations like that. The diagonal lines of the diagonal box of a training image (which will match a theoretical individual’s ones) are scanned to H, L and S vectors. Inside this vectors, a local estimate of the derivative is calculated. Then pixels having a high local derivative value are chosen to compute possible initial values for the \( d_i \) parameters. Figure 6 shows this process, where the plot represents the derivative estimation for the marked diagonal (starting from the top left corner), while the vertical bars over the graph indicate the selected initial \( d_i \) values.
4 EXPERIMENTAL RESULTS

Experiments have been conducted on a B21-RWI mobile vehicle, in the facilities of the System Engineering and Automation Division at the Carlos III University (Armingol et al., 1998).

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.

The pattern recognition stage has shown good robustness with both landmarks tested in a real application. Figure 7 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, with a 25 mm. fixed optic. Matches above the certainty thresholds are good ones with a probability over 95% (85% for the acceptance threshold).

In Figure 8 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 14.5° horizontal angle of view) is shown in Figure 2 a), where two artificial landmarks are successfully detected after only one iteration of the genetic search.

The learning stage has successfully dealt with all the proposed landmarks. Apart from the two landmark tested, other natural landmarks have been learned for testing this stage: university’s large informative panels (fig. 1d)), different kinds of traffic signs (fig. 1e) y f) and highway’s large informative direction panels (fig. 1c, 5, 6). In all cases, adequate recognition parameters have been found. Figure 9 shows several ROI obtained with learned segmentation parameters for office’s nameplates and road informative panels.

In future work, all the learnt landmark will be tested for their utility in real navigation tasks.
Fig. 9. Learned segmentation results.

4 CONCLUSIONS

In this paper a new color vision-based landmark learning and 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 new learned pattern-windows) and text styles (retraining the classifier weights). This generalization ability is the relevant advantage from classical rigid methods. 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, and outside environments with well-defined landmarks such as streets and roads.

ACKNOWLEDGMENTS

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

REFERENCES


