A Vision System for Environment Representation: From Landscapes to Landmarks

Rafael Murrieta-Cid[1,⋆], Carlos Parra[2,⋆⋆], Michel Devy[3], Benjamín Tovar[1], and Claudia Esteves[1]

[1] ITESM Campus Ciudad de México Calle del puente 222, Tlalpan, México D.F.
{rmurriet,betovar,cesteves}@campus.ccm.itesm.mx
[2] Pontificia Universidad Javeriana Cra 7 No 40-62 Bogotá D.C., Colombia
carlos.parra@javeriana.edu.co
[3] Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS-CNRS)
7, Avenue du Colonel Roche, 31077 Toulouse Cedex 4, France
michel@laas.fr

Abstract. In this paper a complete strategy for scene modeling from sensory data acquired in a natural environment is defined. This strategy is applied to outdoor mobile robotics and goes from environment recognition to landmark extraction. In this work, environment is understood as a specific kind of landscape, for instance, a prairie, a forest, a desert, etc. A landmark is defined as a remarkable object in the environment. In the context of outdoor mobile robotics a landmark has to be useful to accomplish localization and navigation tasks.

1 Introduction

This paper deals with the perception functions required to accomplish the exploration of a natural environment with an autonomous robot. From a sequence of range and video images acquired during the motion, the robot must incrementally build a model, correct its situation estimate or execute some visual-based motion. The main contribution of this paper concerns the enhancement of our previous modeling methods [10,9,11,8,3,1,4] by including more semantic information. This work has shown through intensive experimentation that scene interpretation is a useful task in mobile robotics because it allows to have information of the environment nature and semantic. In this way, the robot will have the needed information to perform complex tasks. With this approach it becomes possible to command the robot with semantic instead of numerical vectors. For instance the command of going from (x_1, y_1) to (x_2, y_2) can be replaced with *"Go from the tree to the rock"*.

2 The General Approach

This work is related to the context of a Mars rover. The robot must first build some representations of the environment based on sensory data before exploiting

⋆ This research was funded by CONACyT, México
⋆⋆ This research was funded by the PCP program (Colombia -COLCIENCIAS- and France) and by the ECOS Nord project number C00M01.

C.A. Coello Coello et al. (Eds.): MICAI 2002, LNAI 2313, pp. 11–20, 2002.

them. The proposed approach is suitable for environments in which (1) the terrain is mostly flat, but can be made by several surfaces with different orientations (i.e. different areas with a rather horizontal ground, and slopes to connect these areas) and (2) objects (bulges or little depressions) can be distinguished from the ground. Several experimentations on data acquired on such environments have been done. Our approach has been tested partially or totally in the EDEN site of the LAAS-CNRS [8,9], the GEROMS site of the CNES [11], and over data acquired in the Antarctica [16]. These sites have the characteristics for which this approach is suitable. The EDEN site is a prairie and the GEROMS site is a simulation of a Mars terrain. The robot used to carry out these experiments is the LAMA robot (figure 1).

Related work: The construction of a complete model of an outdoor natural environment, suitable for the navigation requirements of a mobile robot, is a quite difficult task. The complexity resides on several factors such as (1) the great variety of scenes that a robot could find in outdoor environments, (2) the fact that the scenes are not structured, then difficult to represent with simple geometric primitives, and (3) the variation of the current conditions in the analyzed scenes, for instance, illumination and sensor motion. Moreover, another strong constraint is the need of fast algorithm execution so that the robot can react appropriately in the real world.

Several types of partial models have been proposed to represent natural environments. Some of them are numerical dense models [5], other are probabilistic and based on grids [7]. There exist also topological models [6]. In general, it is possible to divide the type of information contained in an environment model in three levels (one given model can contain one or several levels) [2]: (1) geometric level: it contains the description of the geometry of the ground surface or some of its parts. (2) topological level: it represents the topological relationships among the areas in the environment. These areas have specific characteristics and are called "places". (3) semantic level: this is the most abstract representation, because it gives to every entity or object found in the scene, a label corresponding to a class where it belongs (tree, rock, grass...). The classification is based on *a priori* knowledge learnt off-line and given to the system. This knowledge consist on (1) a list of possible classes that the robot could identify in the environment, (2) attributes learnt from some samples of each class, (3) the kind of environment to be analyzed, etc.

2.1 The Navigation Modes

We propose here two navigation modes which can make profit of the same landmark-based model: trajectory-based navigation and sensor-based navigation.

The sensor-based navigation mode needs only a topological model of the environment. It is a graph, in which a node (a place) is defined both by the influence area of a set of landmarks and by a rather flat ground surface. Two landmarks are in the same area if the robot can execute a trajectory between

Fig. 1. LAMA robot

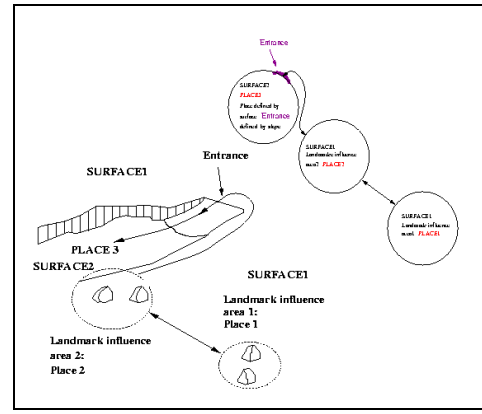

Fig. 2. Topological model

them, with landmarks of the same set always in the stereo-vision field of view (max range = 8m). Two nodes are connected by an edge if their ground surfaces have significantly different slopes, or if sensor-based motions can be executed to reach one place from the other. The boundary between two ground surfaces is included in the environment model by using B-Spline representing the area border [3]. These boundaries can be interpreted as "doors" towards other places. These entrances are characterized by their slope. A tilted surface becomes an entrance if the robot can navigate through it. An arc between two nodes corresponds either to a border line, or to a 2D landmark that the robot must reach in a sensor-based mode. In figure 2 a scheme representing the kind of environment where this approach is suitable and its representation with a graph are shown.

The trajectory-based navigation mode has an input provided by a geometrical planner which is selected inside a given landmark(s) influence area. The landmarks in this navigation mode must be perceived by 3D sensors because they are used to localize the robot (see figure 17). The sensor-based navigation mode can be simpler because it exploits the landmarks as sub-goals where the robot has to go. The landmark position in a 2D image is used to give the robot a direction for the motion (see figure 16). Actually, both of the navigation modes can be switched depending on (1) the environment condition, (2) whether there is 3D or 2D information and (3) the availability of a path planner. In this paper we present overall examples when 3D information is available.

3 Environment Modeling

In order to build this environment model, we developed an approach which consists in steps executed in sequence using different attributes in each one and profiting intensively by contextual information inferences. The steps are environment recognition, image segmentation, region characterization and classification,

contextual information inferences and landmark selection. The steps are strongly connected. A new step corrects the errors that might arise on the previous ones. We take advantage from the cooperation between the segmentation and classification steps so that the result of the first step can be checked by the second one and, if necessary, corrected. For example, over-segmentation is corrected by classification and identification errors are corrected by contextual information inferences.

For some applications, a robot must traverse different types of environment (urban or natural), or must take into account changes in the environment appearance (season influence in natural scenes). All these variations could be given as *a priori* knowledge to the robot. It is possible to solve this problem by a hierarchical approach: a first step can identify the environment type (i.e., whether the image shows a forest, a desert or an urban zone) and the second one the elements in the scene. Global image classification is used as an environment recognition step where a single type of environment is determined (i.e., forest, desert or urban zones). In this way, an appropriate database is found making it easier to label the extracted regions by a reduced number of classes and allowing to make inferences from contextual information. Involving this information helps controlling the complexity of the decision-making process required to correctly identify natural objects and to describe natural scenes. Besides, some objects (such as a river, a hole, or a bridge) cannot be defined or recognized in an image without taking into account contextual information [15]. It also allows to detect incoherences such as a grass surrounded with sky or rocks over trees on a flat ground.

For several reasons, it is better to perform the interpretation of the scene in different steps by using different attributes in each one taking into account the system involved in the image acquisition. The attributes used to characterize environments must be different because they have different discriminative power according to the environment. For instance, in lunar-like environment color is not useful given that the entire environment has almost the same colors, but texture and 3D information are. In terrestrial natural areas the color is important because it changes drastically according to the class the object belongs to.

Now, let us describe the sensors used in our experiments. Our robot is equipped with a stereo-vision system composed by two black and white cameras. Additionally to this stereo-vision system a single color camera has been used to model scenes far away from the robot. We want to associate intensity attributes to an object extracted from the 3D image. This object creates a 2D region in the intensity image acquired at the same time than the 3D one. The 3D image is provided by a stereo-vision algorithm [11]. Image regions corresponding to areas which are closer to the sensors (max range 8m) are analyzed by using 3D and intensity attributes. In these areas, stereo-vision gives valid information. Regions corresponding to areas further from the sensors reliable range will be analyzed by using only color and texture attributes given that 3D information is not available or too noisy. For these areas, since color is a point-wise property of images and texture involves a notion of spatial extent (a single point has no

texture), color segmentation gives a better compromise between precision of region borders and computation speed than texture segmentation, consequently, color is used in the segmentation step.

Environment recognition: Our environment recognition method is based on the metric known as the Earth Mover's Distance [12]. This metric is based on operations research theory and translates the image identification problem into a transportation problem to find the optimal work to move a set of ground piles to a set of holes. The ground piles and holes are represented by clusters on the images which map to a feature space and may be constructed by any attribute on the images (i.e. color spaces, textures, ...). These approaches are not able to identify the elements in the scene, but the whole image as an entity. We construct a 3-dimensional attribute space for the images comparison. Two axes map to I_2I_3, the uncorrelated chrominance attributes obtained from the *Principal Components Analysis*. The other axis correspond to texture entropy feature computed from the sum and difference histograms.

For the environment recognition step we feed our system with six classes of environments: forest (Fig 3), Mars (Fig. 4), Moon (Fig. 5), prairie (Fig. 6), desert (Fig. 7) and a snowed forest (Fig. 8). Every class is constructed with a set of images. Our system finds the environment class where the test image (Fig. 9) belongs. The test image shows a *prairie*. Even thought the classes prairie and forest are similar the system assigns correctly the image test to the prairie class. It is also capable to differentiate *Moon* images from the *snowed forest* images although the colors are similar. In our tests the system was also capable of differentiate Mars from the desert, but the similarity was greater (the work to move a set of clusters to the other was smaller).

Segmentation: Image segmentation for region extraction: this segmentation is based on clustering and unsupervised classification. The image is segmented to obtain the main regions of the scene. This first step can be performed by the use of the color attribute on the 2D image or by the use of geometrical attributes on the 3D image [9,10].

Characterization: Each region of the scene is characterized by using several attributes computed from the color, texture or geometrical informations [14,13].

Classification: Our identification step is based on a supervised learning process, for this reason its good performance depends on the use of a database representative enough of the environment. It is important to remark that prototyping is done to build the learning samples set in order to get a representative enough database. Actually we are making two types of prototyping, one with the images by using image comparison and the other with the learning sampling set. Bayesian classification is used to associate every region in the image with a semantic label. This classification method has some drawbacks. It needs the computation of all the set or the previously defined attributes. Bayesian classification has been criticized arguing that it needs frequently a lot of knowledge

Fig. 3. Forest

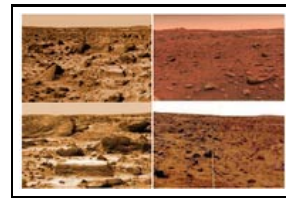

Fig. 4. Mars

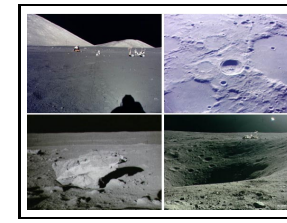

Fig. 5. Moon

Fig. 6. Prairie

Fig. 7. Desert

Fig. 8. Snowed Forest

Fig. 9. Test image

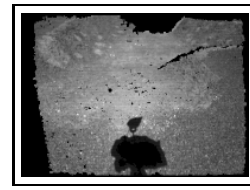

Fig. 10. Original image

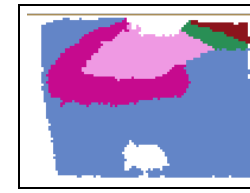

Fig. 11. 3D segmentation

about the problem. It has also been pointed out that this approach has a lot of faults when representing and manipulating knowledge inside a complex inference system. In order to deal with these drawbacks, attribute selection has to be done in a pre-processing step (by using PCA and Fisher criteria) and inferences are added to the system by an independent process such as environment recognition and contextual information.

Contextual information inferences: The specific environment analyzed in this work consist in natural areas where ground is flat or with a smooth slope. By the use of some contextual characteristics of the environment the model consistency can be tested. Possible errors in the identification process could be detected and corrected by using simple contextual rules. A set of rules allow to find eventual errors introduced by the identification step [10]. The probability of belonging to a given class is used to decide whether the region should be re-labeled or not. If this probability is smaller than a given threshold the region is re-labeled.

To show the construction of the representation of the scene based on only 2D information, we present the process in a image. These regions were obtained from the color segmentation phase. Sometimes a real element is over-segmented, consequently a fusion phase becomes necessary. In this step, connected regions belonging to the same class are merged. Figure 12 shows the original image. Figure 13 shows the color image segmentation and the identification of the regions. The defined classes are a function of the environment type. Here, we have chosen 4 classes which correspond to the main elements in our environment: grass, sky, tree and rock. Labels in the images indicate the nature of the regions: (R) rock, (G) grass, (T) tree and (S) sky. The coherence of the model is tested by using the topological characteristics of the environment. The Region at the top right corner of the image was identified as grass, however this region has a relatively low probability (less than a given threshold) of belonging to this class, in this case the system can correct the mistake by using contextual information. This region is then relabeled as tree. Figure 14 shows the final model of this scene. Figure 15 shows the gray levels used to label the classes.

Fig. 12. Original image

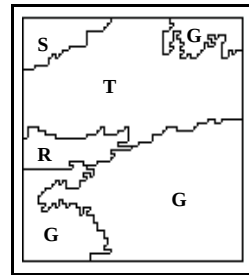

Fig. 13. Segmentation and Identification

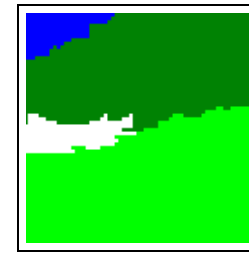

Fig. 14. Final model

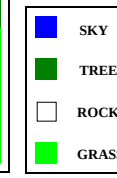

Fig. 15. Classes

3.1 Landmark Selection

Landmarks in indoor environments correspond to structured scene components, such as walls, corners, doors, etc. In outdoor natural scenes, landmarks are less structured. We have proposed several solutions like maxima of curvature on border lines [3], maxima of elevation on the terrain [4] or on extracted objects [1].

Based on our previous works a landmark is defined as a remarkable object [1], which should have some properties that will be exploited in the robot localization or in visual navigation task, but in this work the landmark is associated to a semantic label. The two main properties which we use to define a landmark are: **Discrimination:** A landmark should be easy to differentiate from other surrounding objects. **Accuracy:** A landmark must be accurate enough so that it can allow to reduce the uncertainty on the robot's situation, because it will be used for robot localization.

Depending on the kind of navigation performed (section 2) the landmarks have different meaning. In trajectory-based navigation landmarks are useful to

localize the robot and of course the bigger number of landmarks in the environment the better. For topological navigation a sequence of different landmarks (or targets) is used as sub-goal the robot must successively reach [9]. For this last kind of navigation commutation of landmarks is an important issue. We are dealing with this task, based on the position of the landmark in the image (see section 4, image 16). The landmark change is automatic. It is based on the nature of the landmark and the distance between the robot and the target which represents the current sub-goal. When the robot attains the current target (or, more precisely, when the current target is close to the limit of the camera field of view), another one is dynamically selected in order to control the next motion [9].

4 Robot Navigation Based on the Landmark-Based Model

Robot visual navigation is done by using the proposed model. We illustrate this task with a experiment carried out with the mobile robot LAMA. Figure 16 (a) shows the video image, figure (b) presents the 3-D image and figure (c) shows the 3-D image segmentation, classification and boundary box including the selected landmark. The selection was done taking into account 3-D shape and nature. The second line of figure 16 represent the tracking of a landmark through an image sequence. The landmark is marked on the picture with a little boundary box. The tracking process is performed based on a comparison between a model of the landmark and the image. In [8] the tracking technique used is described in detail. When the landmark position is close to the image edge it becomes necessary to select another landmark. So the figure 16 III presents the new landmark selection based on image segmentation and classification. The next sequence of tracking is shown on the line IV of figure 16 and the next landmark commutation is presented on line V. Finally on the line VI the robot continue navigation task.

4.1 Experiments of Simultaneous Localization and Modeling (SLAM)

We illustrate this task with an experiment carried out in the EDEN site at LAAS-CNRS. In this work SLAM task is based on landmark extraction. The strategy to select the landmarks is the one presented on section 4. Left column of figure 17 shows 2-D images corresponding to left stereo-vision camera. On these images the rocks selected as target and the zone where the target is looking for are shown. The results obtained regarding environment modeling are shown on the second column. The maps of the environment and the localization of the robot are presented on the third column. On the row "I" the robot just takes one landmark as reference in order to localize itself. On the last row the robot uses 3 landmarks to perform localization task, the robot position estimation is shown by using rectangles. The current robot's situation and the numerical attributes of the

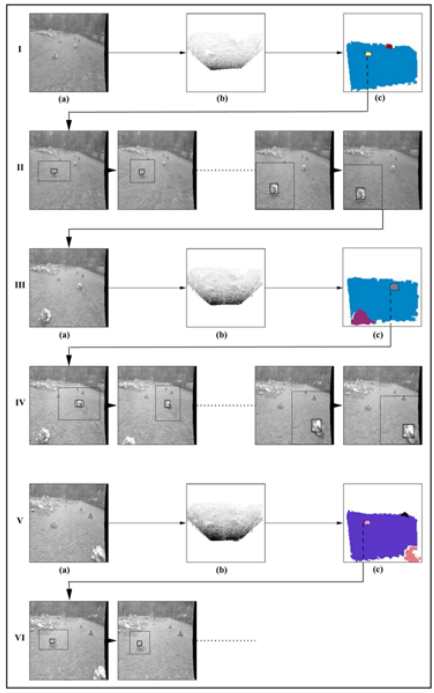

Fig. 16. Visual robot navigation

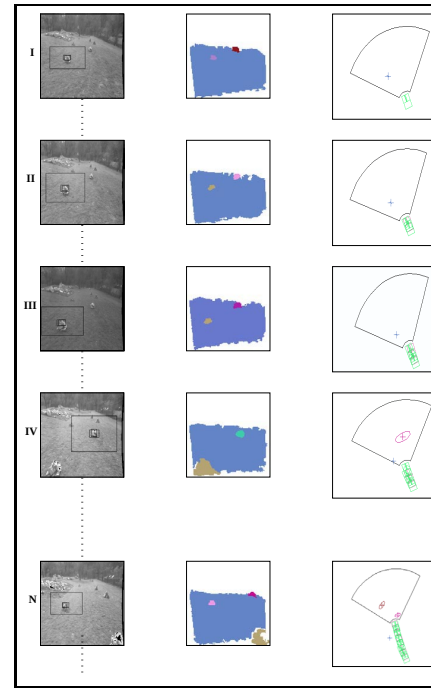

Fig. 17. Simultaneous localization and modeling (SLAM)

landmark features are updated by using an Extended Kalman Filter (EKF). The most important result here is that the robot position uncertainty does not grow thanks to the usage of landmarks. The landmarks allow to stop the incremental growing of the robot position uncertainty.

5 Conclusion

The work presented in this paper concerns the environment representation applied to outdoor mobile robotics. A model of the environment is constructed in several steps: environment recognition, region extraction, object characterization, object identification, landmarks selection. Robot navigation based on the landmark-based model is presented.

References

1. S. Betg-Brezetz, P. Hébert, R. Chatila, and M. Devy. Uncertain Map Making in Natural Environments. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, Minneapolis, USA, April 1996.

2. R. Chatila, J.-P. Laumond, Position Referencing and Consistent World Modeling for Mobile Robots. In *Proc IEEE Int. Conf. on Robotics and Automation (ICRA)*, 1985.
3. M. Devy and C. Parra. 3D Scene Modelling and Curve-based Localization in Natural Environments. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, Leuven, Belgium, 1998.
4. P. Fillatreau, M. Devy, and R. Prajoux. Modelling of Unstructured Terrain and Feature Extraction using B -spline Surfaces. In *Proc. International Conference on Advanced Robotics (ICAR)*,Tokyo, Japan, November 1993.
5. M. Hebert, C. Caillas, E. Krotkov, I. Kweon and T. Kanade. Terrain mapping for a roving planetary explorer. In *Proc. International Conference on Robotics and Automation (ICRA)*, Vol 2, may 1989.
6. I.S. Kweon and T. Kanade. Extracting topological features for outdoor mobile robots. In *Proc. International Conference on Robotics and Automation (ICRA)*, Sacramento, USA, may 1991.
7. S. Lacroix, R. Chatila, S. Fleury, M. Herrb and T. Simeon Autonomous navigation in outdoor environment: adaptive approach and experiment In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, San Diego, USA, may 1994.
8. R. Murrieta-Cid, M. Briot, and N. Vandapel. Landmark identification and tracking in natural environment. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Victoria, Canada, 1998.
9. R. Murrieta-Cid, C. Parra, M. Devy and M. Briot. Scene Modelling from 2D and 3D sensory data acquired from natural environments. In *Proc. IEEE International Conference on Advanced Robotics (ICAR)*, Budapest, Hungary, 2001.
10. R. Murrieta-Cid, C. Parra, M. Devy, B. Tovar and C. Esteves. Building multi-level models: From landscapes to landmarks. Submitted to *the IEEE International Conference on Robotics and Automation (ICRA2002).*
11. C. Parra, R. Murrieta-Cid, M. Devy & M. Briot. 3-D modelling and robot localization from visual and range data in natural scenes.. In *Proc. International Conference on Vision Systems (ICVS)*, Las Palmas, Spain, January 1999.
12. Y. Rubner, C. Tomasi, and L. Guibas. A metric for distributions with applications to image databases. In *IEEE International Conference on Computer Vision*, Bombay, India, 1998.
13. T.S.C Tan and J. Kittler. Colour texture analysis using colour histogram. *I.E.E Proc.-Vis.Image Signal Process.*, 141(6):403–412, December 1994.
14. M. Unser. Sum and difference histograms for texture classification. *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence*, 1986.
15. Strat, T. and Fischler M. Context-Based Vision: Recognizing Objects Using Information from Both 2-D and 3-D Imagery. *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence, 1986 Vol13, Num10, pages 1050-1065, October, 1991.*
16. N. Vandapel, S. Moorehead, W. Whittaker, R. Chatila and R. Murrieta-Cid. Preliminary results on the use of stereo color cameras and laser sensors in Antarctica. In *Proc. 6th International Symposium on Experimental Robotics (ISER)*, Sydney Australia, 1999.