Building Multi-Level Models: From Landscapes to Landmarks

Rafael Murrieta-Cid* Carlos Parra† Michel Devy‡ Benjamín Tovar* Claudia Esteves*

* ITESM Campus Ciudad de México
Calle del puente 222, Tlalpan, México D.F.
{rmurriet,betovar,cesteves}@campus.ccm.itesm.mx

†Pontificia Universidad Javeriana
Cra 7 No 40-62 Bogotá D.C., Colombia
carlos.parra@javeriana.edu.co

‡Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS-CNRS)
7, Avenue du Colonel Roche, 31077 Toulouse Cedex 4, FRANCE
e-mail: michel@laas.fr

Abstract— **In this paper a complete strategy for scene modelling 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 prairie, forest, 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 perform localization and navigation tasks.**

Keywords— **Environment modelling, landmarks, outdoor mobile robotics**

I. Introduction

This paper deals with perception functions required on an autonomous robot to build a multi-level model of the environment. The model here presented combines geometrical, topological and semantic information. The main contribution of this paper concerns the enhancement of our previous modelling method [7], [8], [9] by including more semantic information.

From a sequence of range and video images acquired during the motion, the robot must incrementally build a model and correct its situation estimate. 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 depressions) can be distinguished from the ground. Experimentation over data acquired on these kind of environment has been done. This approach was tested on two suitable sites: a terrestrian site (a prairie located at LAAS-CNRS) [7], [8], and on a simulated planetary terrain [9].

In section II, we describe our global approach to deal with the navigation of a mobile robot in a natural environment, thanks to a multi-level model with geometrical, topological and semantic knowledge. Then, in section III, we present the different perceptual functions used to build such a model from range and color images acquired from the robot itself. These functions provide a landmark-based model. Landmarks will be selected as successive sub-goals along a path the robot must execute. Finally, in section IV, experimental results for a sensor-based navigation task are presented and analyzed. The experimental platform used to carry out these experiments is the robot LAMA (figure 1). LAMA 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.

Fig. 1. The robot LAMA

II. The global approach

In order to build a robust and complete scene model, instead of a single method, this work proposes a system which integrates several functions and tasks. Previous papers [9] focus on the interactions between these functions: image analysis, landmark selection, landmark tracking and Simultaneous Localization and Modelling (SLAM). The system as a whole approach is original and quite functional. This paper points out mainly the image analysis level.

This research was funded by CONACyT, México, We want to thank professor Maurice Briot for his advice.

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

A. Related work

The construction of a complete model for outdoor natural environments applied to mobile robotics is a quite difficult task. The complexity resides on several factors: (1) the great variety of type of scenes to be found in outdoor environments; (2) the fact that the scenes are not structured then difficult to represent with simple geometric primitives; (3) the variation of the current conditions in the analyzed scenes, for instance, illumination and sensor motion, and (4) finally, the need of fast algorithm execution so that the robot can react appropriately in the real world.

Several types of models have been proposed to represent natural environments. Some of them are numerical dense models [4], other are probabilistic and based on grids [6]. There exist also object-based models [1] or topological models [5]. In [2], the information belonging to an environment model, is structured in three levels (one given model can contain one or several levels):

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: the most abstract and knowledge-based representation. This level gives to everyone of their entities the name of a class (tree, rock, ground, etc). The classification is based on *a priori* knowledge: the list of possible classes to be found in the environment, the attributes to measure, the kind of environment to be analyzed, etc.

B. The navigation modes

We have proposed two navigation modes which can profit of the same landmark-based model: trajectory-based navigation or 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 them having always landmarks of the same set in the stereovision field of view (max range = 8m). Two nodes are connected by an edge (1) if their ground surfaces are adjacent, but have significantly different slopes, or (2) if they have the same ground surfaces, but sensor-based motions can be executed to reach one place from the other.

The trajectory-based navigation mode requires a path provided by a geometrical planner (see [6]). This navigation mode is selected inside a given landmark(s) influence area. The landmarks in this type of navigation mode must be perceived by 3D sensors, because they are used to localize the robot. 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 motion direction

Actually, both of the navigation modes can be switched depending on (1) the environment condition, and (2) whether there is 3D or 2D information. When it is available, 3D information make possible a trajectory-based navigation based on robot localization from the 3D landmark positions.

C. Environment modelling

The global model has two main components: the first one describes the topological relationships between the detected ground areas, the second one contains the perceived informations for each area. The global model is a connectivity graph between the detected areas (a node for each area, an edge between two connected areas). In this paper, we focus only on the knowledge extracted for a given area: (1) the list of objects detected on this area, with their positions and classes, and (2) the ground model.

The nodes in the graph (places) are defined as landmark(s) influence areas or ground surfaces with significant different slopes. The boundary between two ground surfaces are included in the environment model by using a B-Spline representing the area border [3]. These boundaries can be interpreted as "doors" towards other places. These entrances towards other places are defined by using their slope; such a tilted surface becomes an entrance if the robot can navigate through it.

In order to build this multi-level model our approach 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, oversegmentation is corrected by classification; 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 knowledges 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 such a 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 [10]. 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. The attributes used to characterize environments must be different because they must have different discriminative capacity according to the environment. For instance, in lunar-like environment color is not useful, 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. Thanks to a stereo-vision system, image regions corresponding to areas which are closer to the sensors (max range 8m), can be analyzed by using 3D and intensity attributes. In these areas, stereo-vision gives valid information. Intensity attributes can be associated to a region extracted from the 3D image. This 3D region corresponds to a 2D region in the intensity image acquired at the same time than the 3D one. For the 2D acquisition, two different sensor configurations have been considered. (1) If we are only interested on the texture information, the stereo images have enough resolution. The left stereo image provides the 2D image on which the texture information will be computed. The indexes between the 3D points and the 2D points are the same, so that the region extracted from the 3D image is directly mapped on the 2D image. (2) If we want to take advantage of a high-resolution color camera, the 2D image is provided by a specific camera, and a calibration procedure must be executed off line in order to estimate the relative position between the 2D and the 3D sensors. The 2D region created by an object extracted from the 3D image is provided by the projection of the 3D border line of the object on the 2D image.

Regions corresponding to areas further from the stereo 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.

III. Perceptual functions

We describe briefly six perceptual functions that are successively executed in order to generate the model of a single area of the environment: (1) the global environment recognition to select the kind of entities that can be found inside, (2) the image segmentation that extracts regions, (3) the region characterization that computes attributes for each region, (4) the region classification that labels these regions, (5) the verification of some contextual constraints in order to improve the region-based representation and at last,(6) the extraction of salient and discriminant landmarks from some labelled regions.

The step (1) is introduced in this paper. For the other steps, more details can be found in [8]. The model is built from the range, color and texture information acquired from the robot. Several color representations have been tested, the best color segmentation was obtained by using the I_1, I_2, I_3 space, defined as [12]: $I_1 = \frac{R+G+B}{3}$, $I_2 = (R-B)$, $I_3 = \frac{2G-R-B}{2}$. These components are uncorrelated, so statistically it is the best way for detecting color variations.

A. *Environment recognition*

Our environment recognition method is based on the metric known as the Earth Mover's Distance [11]. This metric is based on operation research theory and translates the image identification problem into a transportation problem to find the optimal plan 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 attributes (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 $I_2 I_3$, the uncorrelated chrominance attributes obtained from the *Principal Components Analysis*. The other axis corres-

Fig. 2. Forest

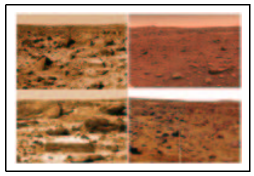

Fig. 3. Mars

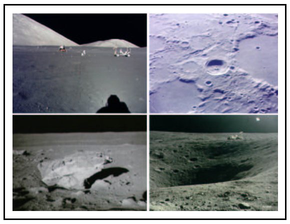

Fig. 4. Moon

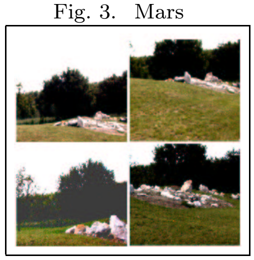

Fig. 5. Prairie

Fig. 6. Dessert

Fig. 7. Snowed Forest

Fig. 8. Test image

ponds to the texture entropy feature computed from the sum and difference histograms [13]. We do not use I_1 to make the system robust against changes in images illumination, neither perform a spatial distribution analysis of the image, which is left to the following steps. Once the environment type or context is known from this first step, a simpler scene interpretation method can be used. In the region identification function, a database organized with respect to the environment type is suitable. It allows to reduce the number of classes, then decreasing the complexity of the problem (i.e. in lunar environment the tree class is not looked for, but the depression class "holes" is).

For the environment recognition step we feed our system with six classes of environments: forest (Fig 2), Mars (Fig. 3), Moon (Fig. 4), prairie (Fig. 5), desert (Fig. 6) and a snowed forest (Fig. 7). Every class is constructed with a set of images. Our system finds the environment class where the test image (Fig. 8) 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 of 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).

B. Image Segmentation

The segmentation algorithm is a combination of two techniques: feature clustering and region growing. The method does the grouping in the spatial domain of square cells, that are associated with the same label defined in an attribute space. The advantage of this hybrid method is that it allows to achieve the process of growing independently of the beginning point and the scanning order of the adjacent square cells.

The division of the image into square cells provides a first arbitrary partition (an attribute vector is computed for each cell). Several classes are defined by the analysis of the attribute histograms, which brings the partition into the attribute space. Thus each square cell in the image is associated with a class. The fusion of the square cells belonging to the same class is done by using an adjacency graph (adjacency-4). Finally, the regions which are smaller than a given threshold are integrated into an adjacent region.

The cell classification is done by using a non supervised classification process, which determines an optimal criterion of class separation by the use of statistical analysis. This approach maximizes a measure of class separability based on standard deviation analysis [8].

This segmentation algorithm can be applied to range images acquired by a stereo-vision algorithm, by the use of 3D attributes (height and normals) computed for each point in the 3D image. The normals (θ and ϕ) are computed in a spherical coordinate system, and are coded in 256 levels.

Image regions corresponding to areas of the environment close to the sensors (in our robot, up to 8 meters) are segmented by using this 3D information. Figure 9 shows a lunar-like environment, figure 10 shows the 3D segmentation. In this example a ground depression in the scene has been successfully segmented. White pixels in segmented images correspond to non correlated points (too distant 3D points, regions with low texture, shadows or occlusions).

Regions corresponding to areas far away from the sensor (beyond 8 meters) will be segmented by using only intensity attributes given that 3D information is not available or too noisy. Color segmentation usually gives a better compromise between the precision of

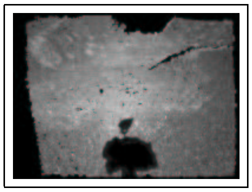

Fig. 9. Original image

Fig. 10. 3D segmentation

region borders and the speed of computation than the texture segmentation; consequently, we decided to use color instead of texture to achieve the segmentation step. The number of no homogeneous regions (sub-segmentation problems) is very small (2%). A good tradeoff between fewer regions and the absence of sub-segmentation has been obtained, even in the case of complex images.

C. Region Characterization

Each object of the scene is characterized by an attribute vector: the object attributes correspond either to 3D features extracted from the 3D image and/or to its texture and its color extracted from the 2D image. The 3D features correspond to the statistical mean and the standard deviation of the distances from the 3-D points of the object with respect to the plane which approximates the ground area from which this object is emerging. We also associate intensity attributes to an object extracted from the 3D image.

Texture and color features are associated globally with the regions provided by the segmentation step on 3D or color images. This strategy generally gives more discriminative information than the one obtained from an arbitrary division of the image.

The color attributes used are I_2I_3. I_1 is not used given that it represents the luminance component which changes drastically with change of illumination. I_2I_3 (chrominance components) are not correlated so information redundancy is not present.

Texture attributes are based on histogram analysis. Histograms change gradually in function of the view point, distance from the sensors to the scene and occlusions. If the acquisition conditions are rather stable (especially constant illumination), the number of data samples required to represent different elements to identify can be reduced [12]. Statistical information can be extracted from these histograms. We have used 6 texture features computed from the sum and difference histograms, these features are [13]: Mean, variance, energy, entropy, contrast and homogeneity.

D. Region 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 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 process provides for each region, a probability of belonging to a given class. This classification method has advantages and drawbacks. It takes into account the different factors in a formal and rigorous frame, it does not need the partition of the attribute space and minimizes the error probability. However, 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 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, the attribute selection has to be done in a pre-processing step (by using PCA and Fisher criteria) and inferences have been added to the system by an independent process using contextual information.

E. Contextual information inferences

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.

The specific environment analyzed in this works consists in terrestrial natural areas where ground is flat or with a smooth slope. If there is sky it has to be in the upper part of the image. It is not intended to cover all the possible configurations, which are too many even for a single and simple environment, but only to detect the most evident errors of contextual consistency. The main rules that have been derived from this context are: (1) A region labeled as *grass* or *rock* cannot be placed between *trees* and *sky*. (2) A region labeled as *rock* cannot be surrounded with a *tree* region. (3) A region labeled as *grass* cannot be surrounded with a *tree* region. (4) A region labeled as *tree* cannot be surrounded with a *grass* region.

Finding these inconsistencies depend only on the knowledge of the *over, below, around*. This relations can be derived from the minimal and maximal vertical coordinates of the regions in an image.

The set of rules allow to find eventual errors introduced by the identification step. If an error is detected it can be corrected using contextual information. For instance, if a region labeled by mistake as *grass* is surrounded by other region labeled as *tree* the region can be re-labeled as *tree*. 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.

At this point of the process, each region in the image has been associated to a class. These regions were obtained from the color or the 3D segmentation phase. The segmentation results in large regions. However, these regions do not always correspond to real objects in the scene. 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.

Fig. 11. Original image

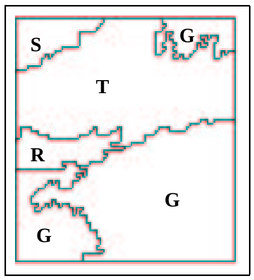

Fig. 12. Segmentation and Identification

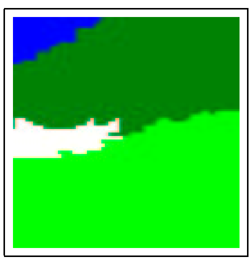

Fig. 13. Final model

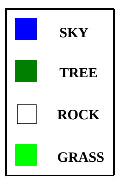

Fig. 14. classes

The construction of the semantic model of the scene based on only 2D information, is illustrated hereafter on the image shown on Figure 11. Figure 12 shows the color image segmentation and the identification of the regions. The defined classes depend on 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 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 13 shows the final model of this scene. Figure 14 shows the gray levels used to label the classes.

F. Landmark selection

The landmark selection phase is composed by two main steps. First, a local model is built from the first robot position in the environment. Then, by using this first local model, a landmark is chosen among the objects detected in this first scene.

A landmark is defined as a remarkable object, which should have some properties that will be exploited in the robot localization or in visual navigation. The two main properties which we use to define a landmark are:
Discrimination. A landmark should be easy to differentiate from other surrounding objects. This property concerns 2D as well as 3D attributes.
Accuracy. A landmark must be accurate enough so that it can allow to reduce the uncertainty on the robot situation, because it will be used to deal with the robot localization. This property is only for the 3D characteristics computed on the 3D regions.

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 or extracted objects [1].

In previous works we have defined a landmark as a little bulge, typically a natural object emerging from a rather flat ground (e.g. a rock). Only the elevation peak of such an object has been considered as a numerical attribute useful for the localization purpose. A realistic uncertainty model has been proposed for these peaks, so that the peak uncertainty is function of the rock sharpness, of the sensor noise and of the distance from the robot [1]. Based on these previous works a landmark is defined as a remarkable object, which should have some properties that will be exploited in the robot localization or in visual navigation, but here the landmark is associated to a semantic label.

In a segmented 3D image, a bulge is selected as candidate landmark if: (1) It is not occluded by another object. If an object is occluded, it will be both difficult to find it in the following images and to have a good estimate on its top. (2) Its topmost point is accurate. This is function of the sensor noise, resolution and object top shape. (3) It must be in "ground contact".

Depending on the kind of navigation performed (section II-B) 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 the landmarks are seen as a sub-goal which the robot has to reach.

Landmark selection based on only 2D is also useful in robotic tasks. The 2D model of the scene can

Fig. 15. Original image

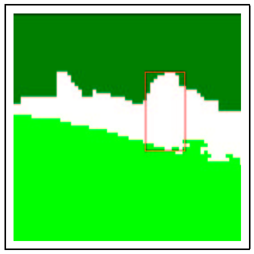

Fig. 16. Landmark selection

be used in order **to give to the robot a goal (direction) corresponding to a landmark of a requested class and 2-D shape**.

Figure 15 shows the original image, figure 16 shows the automatic selection of a landmark based on its nature and shape. In this case the portion of the rock having the largest elongation is selected as the landmark.

Considering all the process our system takes approximately 3 seconds to analyze a scene running on a Linux Pentium III PC Workstation at 800 MHz. This is not yet fast enough for the robot to process the scene during motion however, no all these steps have to be done at the same frequency, for instance environment recognition has to be done with less frequency.

IV. Robot sensor based navigation using the multi-level model

Trajectory based navigation which uses a geometrical planner is done in a given landmark(s) influence area. The landmarks in this type of navigation mode are used to localize the robot [9]. Simultaneous localization and modelling (SLAM) is based on landmark extraction. The navigation mode can be simpler and consisting just on the usage of landmarks as sub-goals where the robot has to go. The landmark position is used to give the robot a motion direction. In our approach the robot sub-goals can be landmarks having a semantic meaning (see [8]). Our final aim is to command the robot with semantic orders instead of numerical ones; 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"*.

For this landmark-based navigation, the commutation of landmarks is an important issue. We are dealing with this task, based on the position of the landmark in the image. In order to navigate during a long robot motion, a sequence of different landmarks is used as sub-goal the robot must successively reach. The landmark change is automatic: it is based on the nature of the landmark and the distance between the robot and the landmark which represents the current sub-goal. When the robot attains the current landmark (or, more precisely, when the current landmark is close to the limit of the camera field of view), another one is dynamically selected in order to control the next motion.

We illustrate this with a experiment carried out with the mobile robot LAMA. Figure 17 (a) shows the video image, figure 17 (b) presents the 3-D image and figure 17 (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 17 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 [7] is described in detail the tracking technique used. When the landmark position is close to the image edge, then it is necessary to select another landmark. So the figure 17 III presents the new landmark selection based on image segmentation and classification. The next sequence of tracking is shows on the line IV of figure 17 and the next landmark commutation is presents on line V. Finally on the line VI the robot continue navigation task.

V. Conclusion

The work presented in this paper deals with the environment representation applied to outdoor mobile robotics. From range, color and texture information, an environment model is constructed in several steps (environment recognition, region extraction, characterization and identification). The multi-level model of the scene is employed in order to select automatically landmarks. A sensor-based navigation mode exploits these landmarks to execute motions.

In this paper, our contribution is the model enhancement by the use of more semantic knowledges during the modelling functions. In future works, we aim to control the robot using mainly the topological and the semantic level of the model.

References

[1] S. Betgé-Brezetz, R. Chatila, and M. Devy. Object-based modelling and localization in natural environments. In *Proc. IEEE International Conference on Robotics and Automation, Osaka (Japon)*, May 1995.

[2] R. Chatila, J.-P. Laumond, Position Referencing and Consistent World Modeling for Mobile Robots. *In Proc IEEE Int. Conf. on Robotics and Automation*, 1985.

[3] M. Devy and C. Parra. 3D Scene Modelling and Curve-based Localization in Natural Environments. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'98)*, Leuven, Belgium, 1998.

[4] M. Hebert, C. Caillas, E. Krotkov, I. Kweon and T. Kanade. Terrain mapping for a roving planetary explorer. In *International Conference on Robotics and Automation ICRA'89*, Vol 2, pages 997-1002, may 1989.

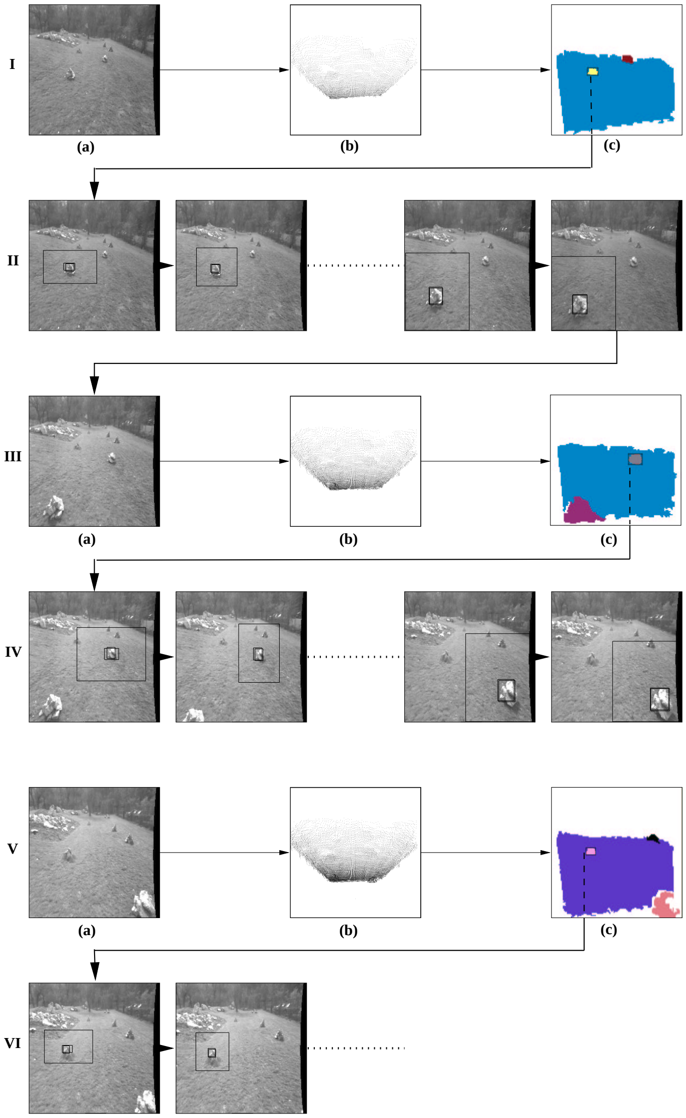

Fig. 17. Visual robot navigation based on landmarks

[5] h I.S. Kweon and T. Kanade. Extracting topological features for outdoor mobile robots. In *International Conference on Robotics and Automation ICRA'91*, Sacramento (USA), 1991.

[6] S. Lacroix, R. Chatila, S. Fleury, M. Herrb and T. Simeon Autonomous navigation in outdoor environment: adaptive approach and experiment In *IEEE International Conference on Robotics and Automation (ICRA94)*, San Diego USA. pp.426-432, 1994.

[7] R. Murrieta-Cid, M. Briot, and N. Vandapel. Landmark identification and tracking in natural environment. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'98)*, Victoria, Canada, 1998.

[8] R. Murrieta-Cid, C. Parra,M. Devy and M. Briot. Scene Modelling from 2D and 3D sensory data acquired from natural environments. In *Proceedings of the IEEE International Conference on Advanced Robotics (ICAR'01)*, Budapest, Hungary, 2001.

[9] C. Parra, R. Murrieta-Cid, M. Devy & M. Briot. *3-D modelling and robot localization from visual and range data in natural scenes.* In Proceedings of International Conference on Vision Systems (ICVS), Las Palmas, España, January 1999.

[10] 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.*

[11] 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.

[12] 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.

[13] M. Unser. Sum and difference histograms for texture classification. *I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence*, 1986.