MODEL-BASED DOOR LOCALIZATION FOR CORRIDOR NAVIGATION

José Sigut
University of La Laguna. Departamento de Física, Electrónica y Sistemas.
38200. La Laguna. Tenerife. Spain
sigut@cyc.ull.es

ABSTRACT

The goal of this work is to provide a mobile robot with visual capabilities to navigate along a corridor and recognize the doors situated on one side of it. A monocular camera and a laser rangefinder were used for this purpose. A model-based vision strategy has been followed. Special emphasis has been put on the perceptual grouping stage and also on the determination of the parameters defining the transformation from the real 3-D scene to the 2-D image acquired by the camera. Several algorithms have been implemented in order to solve these tasks.

1. INTRODUCTION

The problem of indoor robot navigation has been extensively considered in the technical literature [3], [4], [9]. Vision-based approaches rely on optical sensors such as laser rangefinders, cameras or combinations of both. These sensors provide the robot with the information necessary to efficiently explore the surrounding environment. Model-based vision is well adapted for this kind of tasks since many commonly found objects can be modeled by geometric features.

This paper presents several algorithms that exploit an available model of a corridor with offices on one side to allow a mobile robot called Marvin to navigate through it, estimating the pose (position and orientation) of the existing doors. As shown in figure 1, Marvin is equipped with a CCD stereo camera, a Sony EVI-D31 PTZ camera and a laser sensor (only the monocular camera and the laser were used in this work).

The sequence of actions which Marvin is supposed to carry out is the following. As the robot moves along the corridor, it creates, with the help of the laser sensor, a temporary local map of obstacles that is used for short-term movement planning of the platform. If a possible door is detected during this process (any two vertical lines looking like jambs), Marvin stops and moves backward until it finds a position from where a complete image of the door can be acquired. Finally, its relative position and orientation are estimated. It is important to remark that the success of this last operation is strongly dependent on the previous actions, specially the initial door prediction with the laser.

This work is precisely focused on this last stage concerning model-based pose estimation. The general steps to be accomplished in this kind of tasks are outlined in section 2. The perceptual algorithms are described in more detail in section 3 and the algorithms to compute the six translation/rotation parameters which define the relationship between 3-D coordinates and the resultant image coordinates are the subject of section 4. Section 5 is devoted to the conclusions.

Figure 1. A picture from Marvin

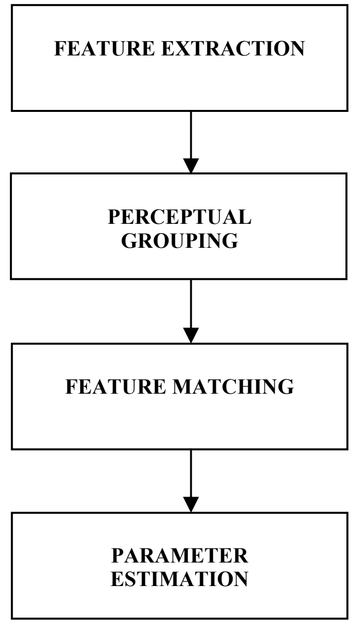

Figure 2. Model-Based Door Recognition Diagram

2. MODEL-BASED DOOR RECOGNITION

Model-based pose estimation consists essentially of finding a match between a model and an image [5], [10]. Most methods perform something like the sequence of steps shown in figure 2.

The aim of the feature extraction stage is the transformation of the original data to give a reduced set of features which appropriately characterize the preprocessed image. For convenience, straight lines were used in this case. Perceptual grouping consists of producing groups of features that are more informative than individual ones and therefore make easier the rest of the recognition process. Matching is one of the central issues of model-based recognition. Given an image and an object model, both represented in terms of their features, the goal is to find the best match between the objects in the image and the available model. In this case, the model of the door was provided by GEM (Generalized Environmental Model), a global data base for world modeling [6]. Finally, once the door has been identified, its parameters of position and orientation are computed.

3. PERCEPTUAL GROUPING

The purpose of perceptual organization is to detect stable image groupings that reflect actual structure of the scene rather than accidental properties [11]. The human vision capabilities seem to include also this ability [2].

There are many different criteria which can be used to group the features. The fundamental image relation of proximity is based on the fact that if two points are close together in the scene, then they will project to points that are close together in the image from all viewpoints. However, it is also possible that points that are widely separated in the scene will project to points arbitrarily close together in the image due to an accident of viewpoint. In particular, the proximity of the endpoints of two line segments may be due to the fact that they are connected or close together in the three-dimensional scene. A similar argumentation is applicable to grouping on the basis of parallelism or on the basis of collinearity.

As it can be seen from the example in figures 3 and 4, things are a little bit easier for this application since most of the extracted segments from the image of interest which are close together (and have been consequently grouped according with the previous criteria) reflect the real structure of the scene. Furthermore before proceeding with the grouping itself, a preprocessing stage was performed in order to eliminate many segments which are too “far” from the predicted door. A threshold measure heuristically set was used to discard these segments. Figure 5 shows the remaining ones. As it becomes clear from figure 6, the objective of obtaining groups of informative features has been met since the original set of 265 segments has been reduced to just a few of them.

4. PARAMETER DETERMINATION

The remaining segments in the example in figure 6 suggest the existence of three possible doors: the one currently detected, another real door on the left, and the piece of wall in the middle which looks like another door due to the wall texture. This represents a typical input to the next stage in the recognition process which must be solved by looking for the best correspondence between the door model supplied by GEM and the three possibilities just mentioned. Again, the initial door prediction plays a fundamental role.

The last step to be accomplished is the computation of the relative position and orientation of the door from the known feature correspondences. There are two main classes of solutions: analytical perspective solutions and numerical perspective solutions. Analytical solutions are relatively easy to be implemented, computationally cheap and work even in scenes with significant perspective effects. However, they work only for a limited number of features, are ambiguous and have very poor error propagation properties. On the other hand, numerical solutions are very general and accurate but they are more complex and computationally demanding. Since the real-time constraint is not very strong in this case, we have chosen this solution.

Three different algorithms were tried: David Lowe's influential and classic algorithm for tracking objects with known geometry [8], a version implemented by Ishii et al which makes different simplifying assumptions [7], and the full projective solution proposed by Araújo et al [1]. The three of them consist of defining a global measure of the discrepancy between the actual image and the image that would be expected given a perspective camera model and an arbitrary estimate for the unknown pose. Then, by replacing the chosen error measure (which is a non-linear function of the pose parameters) with a local linear or low-degree-polinomial approximation on the point corresponding to the current pose estimate, one can compute a correction that in general yields a better pose estimate. This process can be iterated until (ideally) the error function is locally minimized and the current pose estimate converges to the actual pose with the desired precision.

Figure 3. Environment for navigation

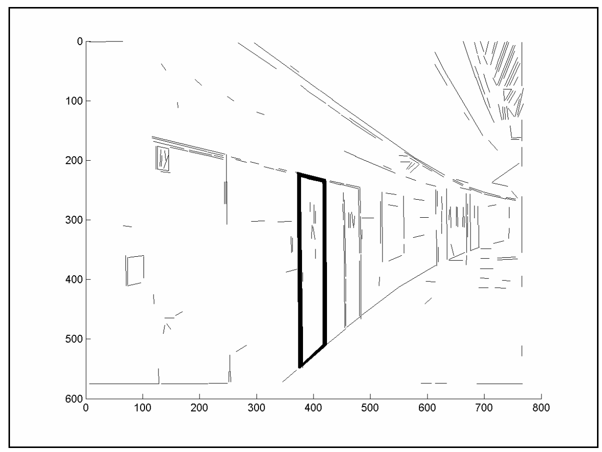

Figure 4. Features initially extracted from the image (the thick line indicates the door currently detected)

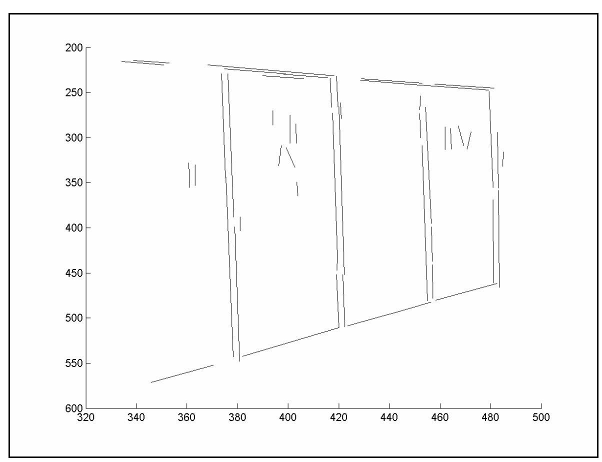

Figure 5. Remaining segments after the discarding operation

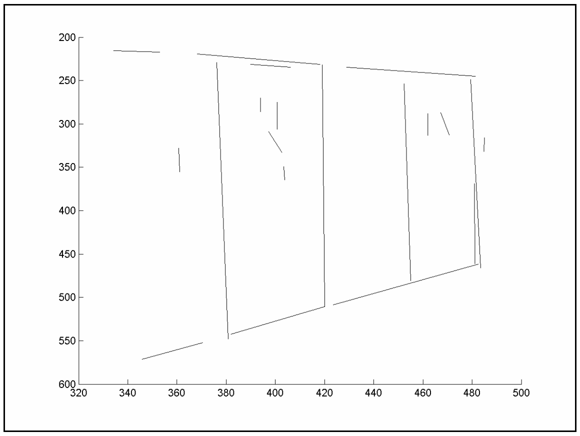

Figure 6. Grouped features

Several experiments were performed to evaluate the performance of these algorithms when applied to this particular problem. In general, the found results were very satisfactory though we have to say that no significant differences among them were observed. By setting the initial parameters to some a priori obtained values, the percentage of convergence was slightly superior to 90% and the average number of iterations was about 5. Figure 7 shows the results corresponding to 1000 executions of the Araújo's method when the initial error is in one of the translational parameters and figure 8 is similar but the initial error is in one of the rotational parameters. The global error measure used is the norm of the vector of distances between the positions of the features in the actual image and the positions of the same features in the reprojected image generated by the estimated pose.

As it was also expected, the simplifying assumptions introduced by Lowe and Ishii et al, resulted in faster execution times but on the other hand this improvement was compensated by the better convergence rate of the algorithm proposed by Araújo et al. The simulations were implemented in MATLAB and the real-time implementation was done in C++. Figure 9 shows Marvin moving towards the recognized door.

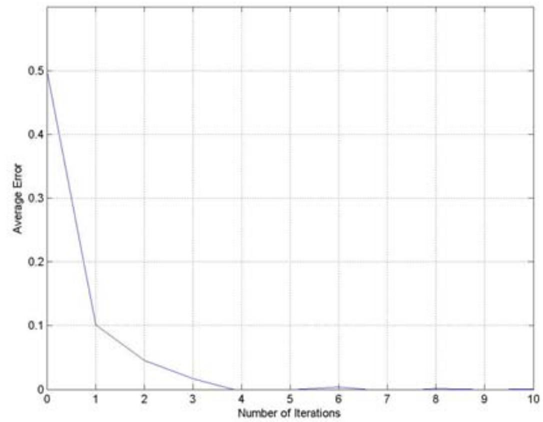

Figure 7. Average error after 1000 executions when the initial error is in one of the translational parameters

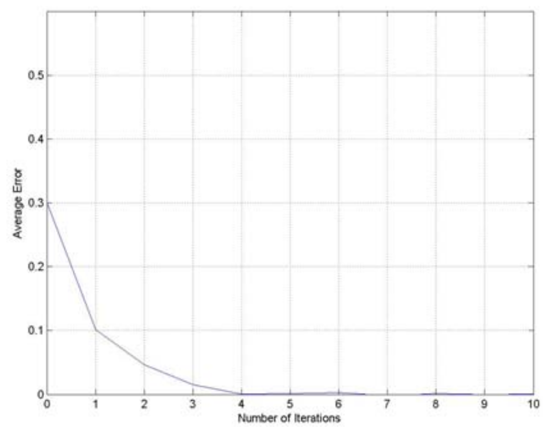

Figure 8. Average error after 1000 executions when the initial error is in one of the rotational parameters

Figure 9. Marving during the recognition process

5. CONCLUSIONS

Different algorithms have been implemented for the purpose of providing a mobile robot with visual capabilities to efficiently recognize the doors situated on one side of a corridor. A model.-based vision approach has been undertaken and the usual steps have been accomplished, putting special emphasis on the perceptual grouping stage and the final step to determine the translation/rotation parameters which define the relationship between 3-D coordinates and the resultant image coordinates.

The obtained results have been in general very satisfactory. As long as the initial door prediction was good enough, the algorithms performed well in most cases. Though some known pose estimation algorithms were tried for comparison purposes, no significant differences were obtained.

ACKNOWLEDGMENTS

The author would like to thank the Lehrstuhl für Realzeit Computersysteme in the Technical University of Munich and specially Dr. Stefan Blum for their support and the supply of Marvin which was used in the experimental part of this work.

REFERENCES

[1] Araújo H. et al, (1996). A Fully Projective Formulation for Lowe's Tracking Algorithm, Technical Report 641, Computer Science Department, University of Rochester.

[2] Beardslee, D., & Wertheimer, M., (1958). Eds. Principles of perceptual organization, *Readings in Perception,* Princeton, N.J.: Van Nostrand, pp. 115-135.

[3] Betke, M., & Gurvits, L., (1997). Mobile Robot Localization Using Landmarks, IEEE Transactions on Robotics and Automation, 13(2), pp. 251-263.

[4] Borenstein, H. et al, (1997). Mobile Robot Positioning: sensors and Techniques, Journal of Robotic Systems, 14(4), pp. 231-249.

[5] Chin, R. & Dyer, (1986). Model-based Recognition in Robot Vision, ACM Comput. Surveys, 18(1), pp. 67-108

[6] Hauck, A. & Stöffler, O., (1995). A Hierarchic World Model Supporting Video-based Localization, Exploration and Object Identification, 2nd Asian Conf. on Computer Vision (ACCV'95).

[7] Ishii, M. et al, (1987). A 3-D Sensor System for Teaching Robot Paths and Environments, International Journal of Robotics Research, 6(2), pp. 45-59.

[8] Lowe, D., (1987). Three-Dimensional Object Recognition from Single Two-Dimensional Images, Artificial Intelligence, 31(3), pp. 355-395.

[9] Ohya, A. et al, (1998). Vision-Based Navigation by a Mobile Robot with Obstacle Avoidance Using Single-Camera Vision and Ultrasonic Sensing, IEEE Transactions on Robotics and Automation, 14(6), pp. 969-978.

[10] Pope, A., (1994). Model-Based Object Recognition. A Survey of Recent Research, Fuzzy sets and fuzzy systems, Technical Report 94-04.

[11] Schlüter, D. & Posch, S., (1998). Combining Contour and Region Information for Perceptual Grouping, Proc. 20. DAGM-Symposium, Informatik Aktuell, Springer, pp. 393-401.