Incremental Map Building using an Occupancy Grid for an Autonomous Monocular Robot

R. Gartshore, A. Aguado, C. Galambos
Center for Vision, Speech and Signal Processing
University of Surrey, Guildford, Surrey, GU2 7XH, UK
Email: R.Gartshore@surrey.ac.uk

Abstract

This paper centres on the problem of estimating the location of features in the environment within a map building framework for a mobile robot with a single camera. Most of the existing approaches to feature location from vision data have been developed by a combination of matching and geometric triangulation. Triangulation is used to obtain equations of reconstruction or structure-from-motion. In contrast, our approach avoids the matching problem by looking for evidence of the location of features on an occupancy grid. Similar to map building via sonar methods, we gather evidence of the trace of potential position of features. Thus, maxima in the occupancy grid define robust feature positions. Preliminary results show that our approach can obtain accurate results in real time. The approach can handle noise in the feature detection and in the robot position. The integration of vision data over time reduces the covariance of the estimates.

1 Introduction

We have developed a map building framework for a mobile robot with a single camera. The framework is divided into three main components. The first component is aimed at estimating the world coordinates of features in the environment. The second component uses features to create a topological map congruent with visibility constraints. The third component uses the topological map with the estimates of features in the environment to control the global and viewpoint planning navigation. This paper centres on the problem of locating and representing features in the environment by using a single camera. Most existing methods of position estimation from vision locate features based on reconstruction. In this approach, the position of the robot or features in the environment are obtained by performing a geometric triangulation for corresponding features [1] [2]. This approach has been shown to be very valuable for calibrated stereo systems and it can be used to compute simultaneously the location of the robot and features [3]. Kalman filters have been included in order to achieve integration of data over time. While in theory this approach has been shown to provide a good solution to the problem of feature location, it remains very difficult to use it in practice. There are three main difficulties with reconstruction: the problem to find

a reliable set of corresponding features, the noise sensitivity of triangulation [4] and the complexity of the cost function during the estimation process [5]. Our approach avoids explicit reconstruction by representing features in the environment on an occupancy grid [6].

Occupancy grids have been successfully used to perform tasks such as obstacle avoidance, map building and localisation [7] [8] [9] [10] [11]. In these methods, evidence of the existence of obstacles is obtained by incrementing elements of a grid representing discrete positions of the environment from sonar data. This approach has proven effective in obtaining a reliable representation of the environment in spite of inaccurate sensor data obtained from wide angle sonar beams and mixed reflections. Additionally, occupancy grids do not rely on the location of correspondences and they integrate information over time maintaining a robust representation of uncertainty.

The motivation of our work is to apply the technique of occupancy grids to map-building methods relying on vision [12] [13]. Images, in an analogous fashion to ultrasonic data, can be used to provide evidence of obstacles in the environment. However, image data can be preferred for two main reasons [14]. First, it has a better angular resolution and thus provides a more detailed definition of the environment. Secondly it is less affected by reflections, thus it has lower noise levels. Additionally, representing visual information in an occupancy grid can be convenient for integrating multiple sensor data.

In contrast to ultrasonic data, images do not provide a direct measure of depth. In order to measure depth, it is necessary to find corresponding features in several images. This problem can be reduced by considering a calibrated stereo rig [12] [13]. However, it can be more difficult to circumvent for a vision system composed of a single camera such as the system described in this paper. In our approach, we avoid the difficulties with the correspondence problem by gathering evidence of features in all the potential locations at any depth. Identification of the most likely cells in the occupancy grid are obtained by considering the confidence in the position of the match and the colour consistency of the features. In our current implementation, we model the world as a projection in a two dimensional plane. That is, we locate features at a fixed height. This simplifies computations and provides a useful representation for a robot where motion is constrained to a plane.

2   Map Building Framework

Our map-building framework which is shown in figure 1 is being implemented on a Pioneer2-DX robot from ActivMedia with a monocular vision system.
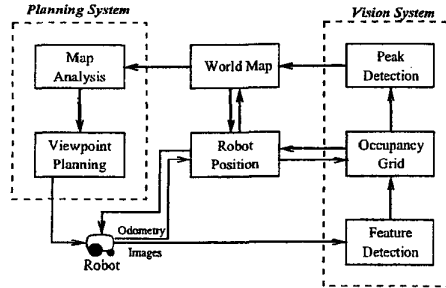

Figure 1: Map-Building System Layout

Input data consists of colour RGB image sequences and odometry information. The system is based on edge detection. Edges are relatively stable image features, they reduce the processing of information and they can be accurately computed from images. The actual topological information concerning the obstacles is obtained during the map analysis stage of processing.

The processing starts by passing images acquired from the camera to the feature detector module. This module uses an edge detector [15] to find vertical lines that define potential boundaries of objects in the world. The vertical edge detector uses a fixed size mask across five scan lines of each image. Vertical features are found in the image by locating maxima in the edge gradient values. The edge detector gives as output the positions of these edges along with their associated RGB gradient value. Since only the central scan line is considered, features are thus projected in two-dimensions (i. e. the horizontal plane at a fixed height).

Detected edge features are used to gather evidence of the object boundaries in the occupancy grid. This requires information of the robot position. The positioning module combines odometry inputs and the features detected in the image to compute the current position of the robot. The problem of position estimation based on occupancy grids has been addressed in [9]. By taking odometry information, an initial estimate of position is obtained. This position is then refined by minimising the discrepancy between the detected features and the projection of world features in the image. Once the position of the robot is known, then evidence of the location of features is reinforced by gathering evidence by back-projecting features from the camera centre.

Using previous frames, all intersections of the back-projected features are added to the occupancy grid using the error model described in section 5.1 along with a consistency criterion of the features gradient values.

As this information is added to the occupancy grid, local areas of the grid are scanned for peaks. The covariance matrices for each of these peaks is calculated from the newly added data and added to the world map.

Most detected maxima in the grid define boundaries of objects, but some are detected on the walls. The aim of the map analysis module is to use the detected maxima to produce a topological representation of the world. This is achieved by maintaining a graph; nodes represent the world elements and links define barriers or obstacles. Without any further information obstacles are formed by all links between the nodes in the map. Using a visibility constraint, openings are formed where features are not occluded.

From the topological map, the covariance of the position of points and obstacles are used as input to a viewpoint planning strategy. Its aim is to direct the robot towards the regions of the world that are not well known. eg where there is a high error covariance associated with a feature. Alternatively when there is a long barrier, more edges are located in between so as to discover new pass ways, as well as avoiding obstacles with high confidence values. In the rest of this paper, we consider in more detail how edges are gathered and detected in the occupancy grid.
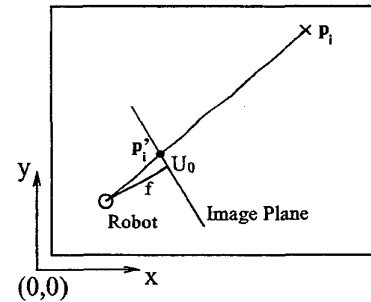
3   Linear Camera Model

Figure 2: Camera Model

As shown in figure 2, a point $\mathbf{p}_i = [X_i, Y_i, 1]^T$ representing an edge in the world is mapped onto the image point $\mathbf{p}'_i = [X'_i, 1]^T$ by projection of the form

$$\mathbf{p}'_i \propto \mathbf{M}\mathbf{p}_i \qquad (1)$$

The projection can be related to a camera model and to the robot pose (ie position and orientation) by,

$$\mathbf{M} = \mathbf{K}[\mathbf{R}|\mathbf{T}] \qquad (2)$$

The robot pose is defined by the rotation $\mathbf{R}$ and the translation $\mathbf{T}$ with respect to the world frame. The intrinsic parameters include the focal length f, the image coordinate of the principal point U_0 and the pixel size K_u.

These parameters can be represented by the matrix

$$K = \begin{bmatrix} K_u & U_0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 \\ 0 & 1 \end{bmatrix} \qquad (3)$$

Intrinsic parameters are computed by an off-line calibration process based on known points on a dual plane calibration chart. Each point in the chart defines a row in the system

$$\begin{bmatrix} 1 & A_1 \\ 1 & A_2 \\ \vdots & \vdots \\ 1 & A_n \end{bmatrix} \begin{bmatrix} U_0 \\ K_u f \end{bmatrix} = \begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_n' \end{bmatrix} \qquad (4)$$

where $\begin{bmatrix} A_i \\ 1 \end{bmatrix} = [R|T]p_i$. In our implementation, the solution of the system of equations in 4 is obtained by combining the closed form solution and robust estimation. First, a collection of potential solutions for the two camera parameters is obtained by solving this equation for pairs of points. After that, the best solution is obtained by the robust mean of all the solutions obtained by the mean shift algorithm [16].

4 Back Projection

Once the internal parameters are known, we can use back projection to produce evidence of the location of world features. This can be achieved by inverting the equations of image formation. According to equation 1, features in the world can be determined from image features by

$$p_i \propto M^{-1}p_i' \qquad (5)$$

In order to obtain a practical realisation of this equation, we consider the back projection of pixels in world coordinates. In this case, the projection matrix only depends on the extrinsic parameters. That is

$$M^{-1} = K[R^T| - T] \qquad (6)$$

From figure 2, we can see that the world coordinates of p_i' are given by $\left[\frac{X'-U_0}{K_u}, f, 1\right]$. Thus, if p_i' is in the homogeneous form $[X_i', 1]^T$, we have that

$$p_i \propto [R^T| - T] \begin{bmatrix} K_u^{-1} \\ f \\ 1 \end{bmatrix} \left(p_i' - \begin{bmatrix} U_0 \\ 1 \end{bmatrix} \right) \qquad (7)$$

We can remove the proportionality by including a constant c. Thus, this equation can be interpreted as a straight line passing through the camera centre and the position of the feature in the image and in the world. In practice, since the calibration only provides the values U_0 and $K_u f$, we cannot determine the world coordinates of p_i'. However, we can determine a point on the line. For example, if we set $f = 1$, then

$$p_i = c[R^T| - T] \begin{bmatrix} K_u^{-1} \\ 1 \\ 1 \end{bmatrix} \left(p_i' - \begin{bmatrix} U_0 \\ 1 \end{bmatrix} \right) \qquad (8)$$

This equation has the same geometric interpretation as equation 7; both define a back projection line.

Back projection equations provide an exact algebraic solution to the problem of feature location. However, a practical realisation cannot assume an exact model. It is necessary to consider probabilistic evidence of the location of obstacles. Additionally, the projection line only defines potential locations of features. Other information such as the colour of features or the gradient obtained from the edge detector can be exploited to increase the confidence of the feature position.

In the remainder of this paper, we consider how we can use an occupancy grid to gather probabilistic evidence defined by the back projection equation.

5 Evidence Gathering and World Feature Detection

Using two images at a time, the features found using the vertical edge detector are projected from the respective camera positions through the features in the image planes towards infinity. At each intersection of these lines, a weight is calculated based on the errors on the robot position and of the image feature positions. A second weight is calculated based on the confidence of the intersection being a matched feature using the edges gradient values.

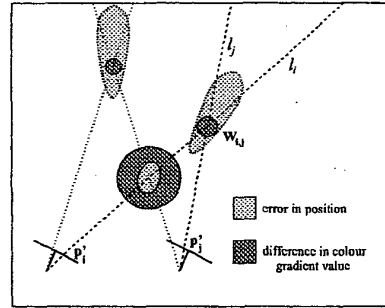Figure 3 shows an example of how two image features

Figure 3: Addition of Features into an Occupancy Grid

are added to the occupancy grid in order to find world feature locations from two camera positions. The feature p_i' with back projection line l_i intersects back projection line l_j of the feature p_j' at $W_{i,j}$. (For clarity, the uncertainty in the robot position and the image feature positions is not shown.) The denser shaded circles centered at the intersections of these back-projected features represent the difference between the two edge gradient values.

The lighter shaded areas depict the error estimate in the position of the world feature, of which the calculation is described below.

5.1 Error Model

The confidence criterion introduced in section 2 for the intersection of features as described above is based upon a Gaussian error model. Figure 4 describes this error distribution for an image feature p'_i and its back projection line l_i with corresponding world point p_i where

σ_P = uncertainty in robot position
σ_θ = uncertainty in image feature position
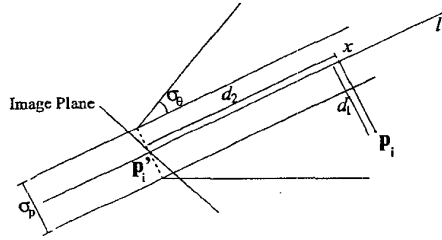
Figure 4: Error Model of Information being added to the Occupational Grid

The line $[x, p_i]$ which is perpendicular to the back-projection line l_i has the error distribution σ_i where

$$\sigma_i^2 = (d_2 \tan(\sigma_\theta))^2 + \sigma_P^2 \qquad (9)$$

Using 9, the probability that the feature is at the given position p_i is determined by

$$P(p_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{d_1}{\sigma_i}\right)^2} \qquad (10)$$

At the intersection of each pair of back-projection lines as shown in figure 3, the probability of $P(p_i)$ and $P(p_j)$ along with the edge gradient colour consistency weighting is added to the occupancy grid to an area of $2.5\sigma_i$ around the intersection.

5.2 Adding World Feature Points

Each peak detected on the feature intersections is added to the occupancy grid as described above is used to update the world map information. By using the Bhattacharyya Bound of equation 11, features in the world are matched to features being added.

Where no match can be found, the new feature is added. Where the feature already exists in the world, the position and its covariance matrix are updated by multiplying the probability density functions of all the feature information.

$$P(error) \leq P(\omega_1) e^{-k(\frac{1}{2})} \qquad (11)$$

where

$$k(\frac{1}{2}) = \frac{1}{8}(\mu_2 - \mu_1)^T \left[\frac{\Sigma_1 + \Sigma_2}{2}\right]^{-1} (\mu_2 - \mu_1)$$
$$+ \frac{1}{2} \ln \frac{|\frac{\Sigma_1 + \Sigma_2}{2}|}{\sqrt{|\Sigma_1||\Sigma_2|}}$$

6 Results

The algorithms described in this paper have been tested with a Pioneer P2-DX robot from ActivMedia. Figure 5 shows the robot in the test office environment used for the results that follow.
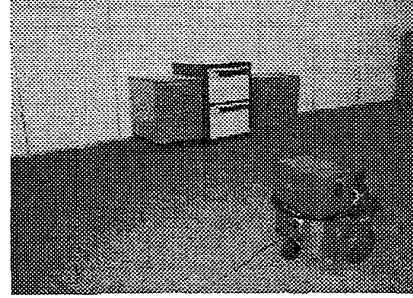

Figure 5: Photograph of Experimental Room Layout

The robot starts with no knowledge of the obstacles in the environment. In this experiment the occupancy grid cell size is set to $25mm \times 25mm$ (whereas [7] uses a $100mm \times 100mm$ cell size). A distance of $100mm$ is travelled between each image.

Figure 6 shows snapshots of the systems knowledge after a number of frames. Highlighted on the images are the edges found by the vertical edge detector. Using the methods described earlier, data is added to the occupancy grid for each of the edges back-projection line intersections using the error model from 5.1. For clarity, the ground truth map of the room's layout is shown under the map being constructed, where only the position error covariances are shown. Subsequent occupancy grids are shown along with the maps created after the peaks found in the grid are added into the map. The world feature positions are displayed with their error covariances drawn to 2.5σ. The gray-level of these ellipses are coded according to the confidence that the feature is in the world. This confidence measure is based upon the number of times the feature has been voted into the grid and the number of times the area containing that feature has been seen.

At the start of the sequence, edge features are added to the occupancy grid. Until enough votes have been cast for any world feature in the grid, a peak is not detected. This is a major strength of the occupancy grid mechanism. It can be seen in each of the grids of figure 6 that data is added to the grid where no features are present in the world. This is to be expected since all possible back-projected trajectory matchings are added into the occupancy grid, but only one match is correct. However, it is not necessary to use computationally expensive matching procedures, since the occupancy grid only builds up strong peaks where features are consistently found.
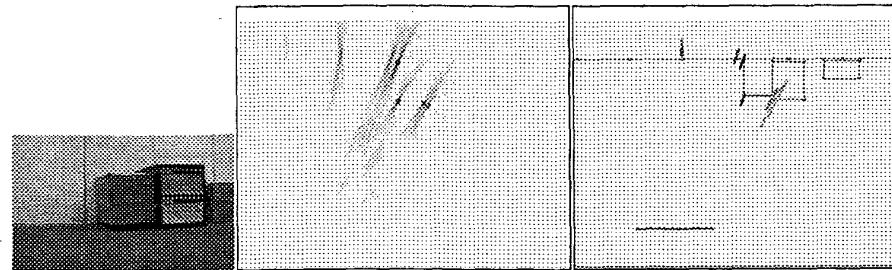
It should be noted that the grid in figure 6(a), contains data relating to many world features. However the evi-
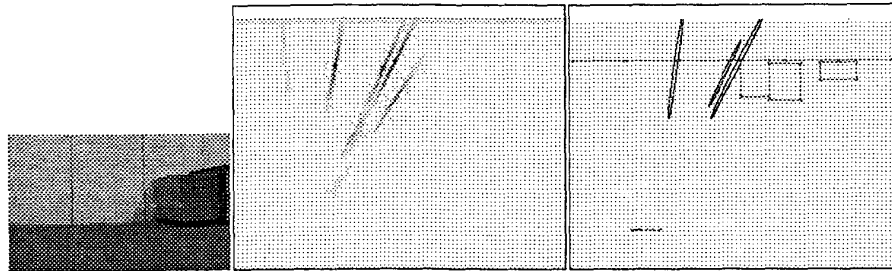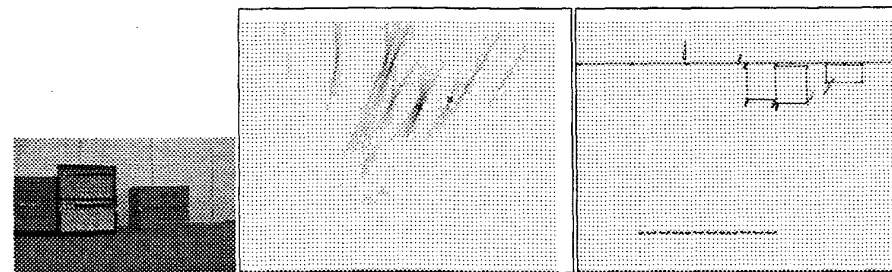
(a) System Snapshot at Frame 6
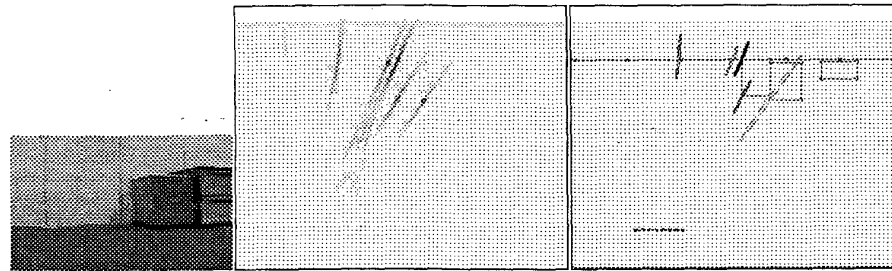
(b) System Snapshot at Frame 9

(c) System Snapshot at Frame 13

(d) System Snapshot at Frame 22

Figure 6: Frames from the Example Robot Sequence: Image with vertical edges detected; Image Evidence Occupancy Grid with current peaks shown; Feature Estimates, Ground Truth Map and Robot Positions

dence of some of these features are not strong enough to be added into the map.
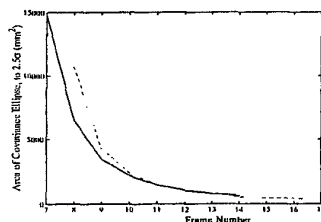

Figure 7: Change in Feature Position Error Covariance Area Over Time for Two Typical Features

As the sequence continues, it can clearly be seen how the robot builds up its world map from image features using the occupancy grid. Figure 7 shows how the area of the error covariance to 2.5σ of the two most frequently seen features, substantially decreases over time.

For the purpose of these results, the robot positioning module is not being used, only globally corrected odometry readings are used to back-project the world features. Therefore the features found at the end of the sequence are further from the ground truth measurements.

By using the edge gradient values, a low confidence weighting is given to world edges seen with very different backgrounds. However, by using the already calculated value, precious computational time is saved by not having to calculate other matching criteria.

7 Conclusions

We have presented a map building framework for a mobile robot with a single camera. The novelty of our approach lies in the combination of an occupancy grid and data from a single camera. Vision information is used to gather evidence of the existence of features in the environment without requiring positive matching of features. Preliminary results are encouraging. Located features are accurate and robust. The occupancy grid effectively reduces the covariance of the estimates by integrating vision data over time. Our current work is aimed at integrating the entire robot system. This will allow us to evaluate the potential of the approach for fully automated environment exploration. Further work on position estimation will be aimed at improving the strategy used to define space partitioning. Additionally, the model of the environment can be enriched by integrating features at several heights via 3D occupancy grids.

References

[1] N. Ayache and O. D. Faugeras, "Maintaining representations of the environment of a mobile robot," *IEEE Transactions on Robotics and Automation*, vol. 5, pp. 804–819, Dec. 1989.

[2] R. Talluri and K. K. Aggarwal, "Mobile robot self-location using model-image feature correspondence," *IEEE Transactions on Robotics and Automation*, vol. 12, pp. 63–77, Feb. 1996.

[3] J. Oliensis, "A linear solution for multiframe structure from motion," in *ARPA94*, pp. II:1225–1231, 1994.

[4] C. B. Madsen, C. S. Andersen, and J. S. Rensen, "A robustness analysis of triangulation-based robot self-positioning," in *Proc of 5th Symposium for Intelligent Robotics Systems*, 1997.

[5] J. Oliensis, "A critique of structure from motion algorithms," 1997.

[6] A. Elfes, "Sonar-based real-world mapping and navigation," *Proc of Journal on Robotics and Automation*, vol. RA-3, pp. 207–209, June 1987.

[7] J. Borenstein and Y. Koren, "Real-time obstacle avoidance for fast mobile robots," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, pp. 1179–1187, Sept. 1989.

[8] H. Yang, J. Borenstein, and D. Wehe, "Sonar-based obstacle avoidance for a large, non-point, omni-directional mobile robot," in *Proc of International Conference in Nuclear and Hazardous Waste Management*, Sept. 2000.

[9] B. Schiele and J. Crowley, "Comparison of position estimation techniques using occupancy grids," in *Proc of International Conference on Robotics and Automation*, pp. 1628–1634, May 1994.

[10] A. Arleo, J. Millan, and D. Floreano, "Learning of variable-resolution cognitive maps for autonomous indoor navigation," *IEEE Transactions on Robotics and Automation*, vol. 15, no. 6, pp. 990–1000, 1999.

[11] W. Burgard, D. Fox, D. Hennig, and T. Schmidt, "Estimating the absolute position of a mobile robot using position probability grids," in *Proc of National Conference on Artificial Intelligence*, 1996.

[12] H. P. Moravec, *Robot Spatial Perception - Stereoscopic Vision Evidence*, 1996.

[13] D. O. Gorodnichy and W. W. Armstrong, "Single camera stereo for mobile robot world exploration," in *Proc of Vision Interface*, pp. 528–535, 1999.

[14] R. A. Brooks, "Visual map making for a mobile robot," *Proc of International Conference on Robotics and Automation*, pp. 824–829, 1985.

[15] C. Harris and M. Stephens, "Combined corner and edge detector," *Proc. Fourth Alvey Vision Conference*, pp. 147–151, 1988.

[16] D. Comaniciu and P. Meer, "Mean shift analysis and applications," in *International Conference on Computer Vision*, pp. 1197–1203, 1999.

Acknowledgments