Occupancy Grid Mapping: An Empirical Evaluation

Thomas Collins*, J.J. Collins*, Conor Ryan*
*Computer Science and Information Systems Department, University of Limerick, Ireland

Abstract—In this paper a quantitative analysis of robotic mapping utilising the fields dominant paradigm, the Occupancy Grid, is presented. The aim of this work is to determine which approach to the robotic mapping problem imbues a mobile robot with the greatest ability to create an accurate representation of its operating environment. We accomplish this by analysing the performance of several established mapping techniques using identical test data. Through evaluating the maps generated by these paradigms using an extensible benchmarking suite that our group has developed we outline which paradigm yields the greatest representational ability.

Keywords - Robotics, Modeling and Simulation, Mapping, Occupancy Grid, Machine Learning

I. INTRODUCTION

The performance of an autonomous mobile robot in acquiring a meaningful spatial model of its operating environment depends greatly on the accuracy of its perceptual capabilities. As it operates in the environment the robot gathers sensory information and subsequently incorporates this into a representation of the environment. The field that is concerned with such issues is known as robotic mapping and is a highly active research field in AI and mobile robotics.

Occupancy Grids have become the dominant paradigm for environmental modeling in mobile robotics [1]. An Occupancy Grid is a tessellated 2D grid in which each cell stores, fine grained, quantitative information regarding which areas of a robot's operating environment are occupied and which are empty [2], [3]. Specifically each individual cell in the grid records a certainty factor relating to the confidence that the particular cell is occupied. Such maps are extremely useful for mobile robotic applications as they facilitate tasks such as navigation, path planning, localisation and collision avoidance [4], [5], [6]. The creation of these Occupancy Grid maps is a non trivial process, as the robot has to interpret the findings of its sensors so as to make deductions regarding the state of its environment. This is facilitated by the use of a sensory model, which is a means of interpreting received sensory measurements.

In mobile robotics the sensors used are less than perfect. One of the most popular sensors is the ultrasonic sonar due to its low cost, its speed of operation and ease of use. These sensors report relative distances between the actual sensory unit and obstacles located within the units perceptual cone. This means that an obstacle, if detected, may be located somewhere within the sonar cone at the distance specified. However these sensors are susceptible to a phenomenon known as specularity. Speculiarity generally occurs when a sensory beam hits a smooth surface and is reflected off that surface at an obtuse angle. This results in either no reading being returned to the emitter unit or an erroneous reading being returned that has bounced off many surfaces. This subsequently creates uncertainty in the mapping process which can reduce the overall quality of the maps created.

II. OCCUPANCY GRID MAPPING TECHNIQUES

For the work outlined herein we utilise five established mapping paradigms which encapsulate the differing paradigms in the domain.

- Matthies and Elfes - 1988, Bayesian framework [8]
- Thrun - 1993, neural network based approach [9]
- Konolige - 1997, enhanced Bayesian framework [10]

A. Moravec and Elfes - 1985:

This technique generates two intermediate models, an empty map $Emp$ and an occupied map $Occ$, which are
subsequently integrated to form a final domain representation. The sensory beam uses a binary classification with cells being in either the free-space or the surface (occupied) area. A two dimensional Gaussian sensor model is used to calculate the probability of the cell being empty if it is in the free-space area and likewise for occupied cells in the occupied area. Specifically a cell in the free-space area has its associated probability of occupancy calculated as follows:

$$P_{E}(xy) = Er(\delta) * Ea(\theta)$$  \hspace{1cm} (1)$$

where $\delta$ represents the distance from a cell of interest to the sensor, $\theta$ represents the angle between the main axis of the sensory beam and the cell and $xy$ are the grid co-ordinates. $Er(\delta)$ is the probability of the cell being empty based on the distance of the cell from the sonar and $Ea(\theta)$ is the probability of the cell being empty based on the angle between the central line of the beam and the line from the sonar to the cell. Likewise a cell in the occupied area of the sensory beam is given by the following:

$$P_{O}(xy) = Or(\delta) * Oa(\theta)$$  \hspace{1cm} (2)$$

The map update used by this technique is heuristic in nature with $P_{E}(xy)$ being integrated into Emp and likewise $P_{O}(xy)$ is integrated into Occ. Finally the empty and occupied maps are combined into a single representation with a thresholding step where the larger value for each cell is chosen for inclusion in the map (equation 3).

$$Map(xy) = Occ(xy) \quad \text{if} \quad Occ(xy) \geq Emp(xy)$$

$$1 - Emp(xy) \quad \text{if} \quad Emp(xy) > Occ(xy)$$  \hspace{1cm} (3)$$

The first limitation with this technique comes from the fact that specular reflection is not considered. Specular reflection occurs when the sensory beam (in this context sonar) reflects off multiple surfaces and then either returns or does not return to the emitter, causing an erroneous reading in either case. This causes cells to be incorrectly labelled as unoccupied, or an unoccupied cell to be incorrectly labelled as occupied. The second issue comes from the fact that when dealing with cells in the Occ and Emp maps once the probability value associated with a cell converges to a certainty of 1, the probability value associated with that cell cannot be altered by any future evidence.

B. Mattheyes and Elfes - 1988:

This technique used the same sensory model as Moravec and Elfes for the purposes of sensor interpretation. However they did develop a more rigorous Bayesian based map updating formula. This updating technique makes the assumption that $Occ(xy) = 1 - Emp(xy) = Map(xy)$. Mattheyes and Elfes used a single map during operation rather than the empty, occupied and integrated maps utilised by Moravec and Elfes previously. Cells in the free-space areas of the beam are updated using equation 4:

$$Map(xy) = P_{E}(xy) * Map(xy)$$

$$Map(xy) = P_{E}(xy) * Map(xy) + (1 - P_{E}(xy)) * (1 - Map(xy))$$  \hspace{1cm} (4)$$

and cells in the occupied sections are updated in an analogous manner.

There are two main disadvantages to using the Bayesian update formula introduced in this paradigm. Firstly a single update can change the occupancy value of a cell drastically which means that cell values can fluctuate. The second disadvantage is that once a cell has converged to certainty i.e. either 0 or 1 the occupancy value cannot be changed.

C. Thrun - 1993:

This method outlines an occupancy grid mapping paradigm which utilises neural networks. The sensory interpretation aspect of this algorithm is implicitly defined in the sensor interpretation network which reports values in the range $<0...1>$ for the sensory readings that it is presented with. Once an occupancy value has been determined it is integrated using equation 5

$$Map(xy) = 1 - \left\{ 1 + \frac{prob(Occ_{xy})}{1 - prob(Occ_{xy})} \right\}^{-1} $$

$$Prob_{t}^{T} \left\{ 1 - prob(Occ_{xy}|s^{t}) \right\}^{-1} \frac{1 - prob(Occ_{xy})}{1 - prob(Occ_{xy}|s^{t})} - prob(Occ_{xy})$$  \hspace{1cm} (5)$$

where $prob(Occ_{xy})$ is the prior probability of a cell’s occupancy, $s$ represents a sensor reading and $t$ represents the time-step.

The limitation of this paradigm arises form the nature of neural networks. When dealing with neural networks it is desirable to train the network until convergence. However it is not practical to train the sensor interpretation network to convergence in this context. This is because to do so would encode environmental characteristics in addition to sensory characteristics in the network which degrades generalisation.

D. Konolige - 1997:

This method also separates the sensory model into occupied and empty sections. However in this case an identical formula is used for both. On receipt of a sensory reading all cells in the beam are updated as follows:

$$P(xy) = \frac{P(r = D|C_{i}) * P(r \geq D|C_{i})}{(1 - (F * r_{i})) * F}$$  \hspace{1cm} (6)$$

where $D$ is the sensor reading, $C_{i}$ states that the cell is occupied, $P(r = d|C_{i})$ is the probability of getting the reading $D$ given that the cell is occupied and $F$ is a small constant representing the probability that an obstacle could exist at a point in the beam other than the range measured. Konolige’s method also introduced ‘Pose Buckets’ as a means of dealing with redundant information and addressed the problem of specular readings through probabilistic inference. For purposes of updating the map the method developed by Konolige combines the probability of the cell being occupied and empty using logarithms and addition

$$Map(xy) = 10log_{10}(Emp(xy)) + 10log_{10}(Occ(xy))$$  \hspace{1cm} (7)$$
where

\[ E_{mp}(xy) = E_{mp}(xy) \cdot P_E(xy) \]

\[ O_{cc}(xy) = O_{cc}(xy) \cdot P_O(xy) \]

and \( P_E(xy) \) and \( P_O(xy) \) are determined by equation 6 and \( E_{mp}(xy), O_{cc}(xy) \) are determined from the intermediate representations that the technique maintains.

An issue with this paradigm is the way in which specularity estimation is applied to individual cells. Specifically if a cell is very confident of the specularity of a sonar reading then this should be propagated to all cells in the sonar beam not just the cell itself as there is a great deal of inter-dependence between cells. Also when a sonar reading is given a high probability of specularity the effect of that reading on the map should be reduced. However the technique deals with the free-space segment of the sonar beam. It would be desirable if both the free and occupied segments of the beam were considered. Finally the assumption is made that environmental surfaces are distributed randomly and that the probability density of a reflection from such a surface is constant. However while this may be true in the simplified case no justification of this assumption is presented for the general case.

E. Thrun - 2001

This mapping technique is quiet different from the four paradigms outlined previously. It takes a view of a sensor model as describing the characteristics of the particular operating environment from the causes (occupancy) to the effects (measurements). This is unlike the sensory model utilised in the previous occupancy grid paradigms which model the effects to the causes. To differentiate Thrun labelled his sensory model the ‘Forward Model’ the general form of which is shown in equation 8.

\[ p(s_i, C_i|m) = p(s_i|m, C_i)p(C_i|m) \]

In equation 8 \( s \) again represents a sensor reading, \( m \) represents the map and \( C_i \) is the correspondence variable relating to the particular reading \( i \). The correspondence variable are binary constructs which account for the different causes of sensor measurements.

This paradigm uses the Expectation Maximisation (EM) algorithm [12] as its basis. Specifically sensor interpretation operates in the E-step of the algorithm with the M-step using the expectations calculated in the E-Step as the basis for the updating of the map. The iteration of both steps continues until the algorithm has converged i.e. a sequence of maps is generated that performs hill climbing in the expected likelihood space with the final map being the optimal solution to the problem at hand.

III. DEVELOPING A BENCHMARKING SUITE FOR OCCUPANCY GRID MAPPING

In Controlling Software Projects: Management, Measurement and Estimation Tom DeMarco stated:

"You cannot control what you cannot measure"

The field of robotic mapping is no different with metrics required which allow for empirical and quantitative evaluation. Historically, in the field, there has been an over reliance on visual inspection of the results obtained form differing mapping techniques which is a highly subjective approach to evaluation. Therefore a clearly defined set of metrics are required if we are to make objective statements about the quality of the maps created by the differing paradigms we have encountered.

A. Evaluating Occupancy Grid Maps

For practical evaluation of the paradigms outlined herein a measure of map quality is required. The measure must be specific to the field we are dealing with but general enough to allow comparison between the maps produced by the differing paradigms. Lee outlined a number of properties that a potential metric should have[13]. These are:

- The metric must be clearly defined.
- The metric must be multi-valued.
- The metric must reflect the purpose of the map.
- The metric must balance coverage and detail.
- The metric must be applicable during the construction of the map.

Of the researchers that have specified benchmarking techniques, Cho[14], Elfes[15] and Lee[13], a benchmarking technique that would allow evaluation of maps based on both their use and structure was not found. Rather a benchmarking suite, originally specified in [16], was refined in light of the properties specified above.

B. A Benchmarking Suite

The benchmarking suite encompasses the following elements:

- An image comparison algorithm based on correlation [17].
- A direct comparison method called Map Scoring specifically designed for probabilistic maps [18], a modified version of the Map Scoring method that only tests the correctness of the obstacles in the map, ignoring the free-space areas.
- A path analysis technique which tests the usefulness of a map as a means of navigation rather than treating it as if it were a picture.

1) Correlation: By conceptualising a map as an image we can use a technique from image analysis known as Baron’s cross correlation coefficient, equation 9, as a basis for evaluating the map.

\[ C_N(y) = \frac{\langle I_T \rangle - \langle I_i \rangle}{\sigma(I_T)\sigma(I_i)} \]

In equation 9, \( I_i \) is the map to be matched, \( T \) is the original map, \( \langle \rangle \) is the average operator and \( \sigma \) is the standard deviation over the area being matched. The generated map is compared to an ideal map of the operating environment, which was constructed from either architectural blue prints or explicit measurements. The result is a percentage value that specifies the similarity of the two maps.

where

\[ E_{mp}(xy) = E_{mp}(xy) \cdot P_E(xy) \]

\[ O_{cc}(xy) = O_{cc}(xy) \cdot P_O(xy) \]
2) Explicit Cell Based Map Analysis with Map Score:
In [18] Martin and Moravec developed a map comparison measure called Map Score to facilitate the automatic learning of sensor models. Map score explicitly compares two maps on a cell by cell basis and bases the result directly on the differences. This means that the maps must have exactly the same orientation, position and coverage for the result to be meaningful. The score for two maps $M$ and $N$ is calculated as the sum of the squared differences between corresponding cells and can be specified as follows:

$$\text{MapScore} = \sum_{m_{x,y} \in M, n_{x,y} \in N} (m_{x,y} - n_{x,y})^2$$  \hspace{1cm} (10)$$

where $m_{x,y}$ is the value of the cell at position $(x,y)$ in $M$ and $n_{x,y}$ is the value of the cell at position $(x,y)$ in $N$. Map score gives a positive value representing the difference between two maps (generally the ideal map of the environment and the generated map that we are evaluating), so the lower the number, the more alike the two maps are. This value is then used to calculate a percentage difference between the two maps.

For a single set of maps this works well. However, it is not a normalised score so no comparison can be made between the results from differing maps. To normalise the score, we compute the worst possible map that could be compared to the ideal map, based on where the robot travelled in an environment. This is done by setting the empty cells in the ideal map to 1, and only change the value of an unknown or occupied cell if it is in within detection proximity of an empty cell i.e. within the sensory field of the robot if the robot happened to be at that empty location. This then allows comparison to be made between the map scores achieved by paradigms over different maps. Figure III-B.2 illustrates an ideal map of an environment and the associated normalisation map.

3) Analysing a map based on its usefulness to a robot:
The previous analysis techniques are focused on comparing ideal to generated maps. However, to fully evaluate a generated map its usefulness to a mobile robot must be considered, as the main context in which such maps are used is in facilitating mobile robots to complete tasks such as navigation. Simply evaluating a map against a perfect snapshot of the operating environment is unrealistic as a map might still be usable to a robot without being a completely faithful representation of that environment. Specifically if the map provides an abstraction of the environment with which a path planning algorithm can specify navigable real world paths then the map can be used. Therefore it is the quality of these paths that indicates the value of the map and the subsequent map evaluation is based on testing two elements:

- The degree to which the paths created in the generated map would cause the robot to collide with a structural obstacle in the real world, and are therefore invalid. These are known as false positives. The False Positives are calculated as follows:
  1) Calculate all the possible paths in the ideal map again using a Voronoi diagram.
  2) Count the number of paths that could not be completed in the generated map due to the existence of obstacles in the map where there are actually none in the ideal map. These are false negatives.

- The degree to which the robot should be able to plan a path from one position to another using the generated map, but cannot because such paths are invalid in the ideal map. These are known as false negatives. The False Negatives are calculate as follows:
  1) Calculate all the possible paths in the generated map using a Voronoi diagram.
  2) Overlay each path (edge in the Voronoi diagram) from the generated map onto the ideal map and count the percentage of the edges that pass through occupied spaces, which would cause the robot crash if it actually followed the path.

These metrics form the basis of the path based analysis utilised to evaluate the maps generated by the various systems.
IV. EXPERIMENTATION AND RESULTS

Experimentation consisted of testing the mapping paradigms with identical data obtained from a number of runs in five test environments. The experimental data was gathered by operating within the environments and recording the sensory information. In total there were three runs completed in each environment which led to the generation of a total of 30 individual maps, each of which was evaluated using the benchmarks outlined. Figure 2 presents some illustrative maps generated by the various mapping paradigms over a single run in one of the environments. Table I presents the results for the evaluation of the five mapping paradigms, correct to two decimal places. For brevity the mapping paradigms are referred to using the first letter(s) of their author(s) name and the year of publication. For example Konolige’s 1997 paradigm is referred to as K97. The benchmark titles are also abbreviated, for example Corr refers to the correlation metric, False Pos refers to the false positive path analysis etc.

A. Analysis of Results

1) Correlation: As can be seen in table I (A), Thrun’s forward model based paradigm achieved the highest correlation, followed by Konolige’s paradigm. This is keeping with the illustrative maps outlined in figure 2. Thrun’s 1993 paradigm having the lowest correlation of the systems tested can be attributed to two causes. Firstly it has a tendency to overestimate free space as can be clearly seen from figure 2 and also it has a tendency to model the extremities of the sensors as being occupied. Therefore when calculating the correlation these inconsistencies are picked up and reflected in the result obtained. The result obtained by Matthies and Elfes 1988 paradigm is similar to that of Thrun 1993. This result can be explained by the fact that Matthie’s and Elfes’ 1988 technique also has a tendency to overestimate free space although in this case the bias toward registration of the range extremities as being occupied is not as great. This is because the implicit sonar model developed through the use of neural networks for the purposes of reading interpretation and subsequent map extraction in Thrun 1993 is functionally similar to the explicit model specified by Matthie’s and Elfes in their 1988 paradigm. An interesting observation is that the ad-hoc updating mechanism of Moravec and Elfes’ 1985 paradigm has out performed the Bayesian update of its successor Matthies and Elfes 1988. This is because the Bayesian paradigm makes the paradigm susceptible to fluctuations in occupancy values when used in conjunction with an approach that has a penchant for over estimation of occupied/empty space.

2) Map Score Metrics: When dealing with the map score all metric, a lower score means that there is less of a difference between the two maps. The overall trends apparent in the previous results are also in existence here. Table I (B) presents the results for the map score all metric. As can be seen Thrun’s 2001 paradigm had the best performance achieving the lowest overall percentage difference between the generated map and the ideal map, followed by Konolige’s 1997 paradigm. Again Thrun’s 1993 neural network based paradigm had the worst performance. The reasons for these performances are the same as those outlined previously in relation to the correlation. The map score all metric was designed to compare a generated map with the complete ideal map of that environment. The map score occupied cells metric on the other hand deals solely with the occupied cells in the map. Table I (C) depicts the results for this metric where it can be seen that the results follow the trends set forth in the previous benchmark.

3) False Positives: Mapping systems that perform poorly in this test are those that have a tendency to
update free-space too strongly. As can be seen in table I (D) Thrun’s 2001 paradigm again had the best performance. This is because the structure of the occupied space in maps generated by the 1993, 1988 and 1985 paradigms causes a number of inconsistent paths to be created due to the tendency to render areas that reach past the occupied cells to the extremity of the sensor beam as being unoccupied. This subsequently causes the creation of possible paths on either side of the correctly identified environmental obstacles, which are not possible in the actual environment. Therefore a higher number of the created paths will be unusable and this is reflected in the metric score. Such paths do not arise to such an extent in the maps created by Konolige’s 1997 paradigm however it is prone to over estimation of occupied space which affects its overall performance. Although Thrun’s 2001 paradigm also has a slight tendency to overestimate unoccupied space at the sensor extremities this space is accurately bounded by occupied areas meaning reducing the overall effect of non-viable paths.

4) False Negatives: Whereas the previous benchmark was concerned with determining the usability of the map as a basis for safe robot navigation this metric is concerned with the usability of the map as a basis for planning a path in the real world environment. This benchmark presents the percentage of false negative paths in the map. These are paths that cannot be completed in the generated map but can be completed in the ideal map. Therefore this test is essentially the inverse of the previous benchmark. Table I (E) illustrates that Thruns 2001 and Konolige’s tied for the best overall result under this test. However, the interesting observation is that the marked contrast evident in the results achieved by the various paradigm under the previous tests is not in existence in this case. This can be attributed to the tendency of the 1985, 1988 and 1993 paradigms to over estimate free space meaning that the number of paths from the ideal environment that cannot be completed in the generated map are fewer and hence the overall score is closer to the better performing paradigms that in the previous cases.

V. CONCLUSION

This article has outlined a comprehensive empirical study of five standard occupancy grid paradigms to determine which approach provides a mobile robot with the greatest ability to represent its operating environment. From the overall results presented derived using the benchmarks outlined previously, it can be seen that Thrun’s 2001 paradigm has the best performance with Konolige’s paradigm coming second. Moravec and Elfes’ 1985 paradigm came third followed by Matthies and Elfes’ 1998 paradigm and Thrun’s 1993 paradigm. These results serve to clearly demonstrate the performance advantage provided by the forward modelling approach to the robotic mapping problem. However the results must be tempered by the fact that there is a sphere of application issue with Thrun’s paradigm as it cannot be applied on line. This is a factor of it’s use of the EM algorithm. Within the paradigm the EM algorithm has to pass through the data set a number of times before it converges with the data having to be gathered before the technique can actually be applied which is in contrast to inverse model based approaches which can construct a map as the data is being gathered.

Therefore, based on the empirical results outlined herein from a purely map creation point of view Thrun’s 2001 paradigm is the superior paradigm as it facilitates the creation of an extremely usable and accurate map of a robots operating environment, provided that appropriate operational data is available for the environment. However if on-line operation is the over riding concern then Konolige’s paradigm is superior as it is a real time paradigm which creates a functional representation of a robots operating environment.

REFERENCES