Experiments in Outdoor Operation of RatSLAM

David Prasser, Gordon Wyeth, Michael Milford,
School of Information Technology and Electrical Engineering
University of Queensland
St. Lucia, Queensland 4072
Australia
{prasserd, wyeth, milford}@itee.uq.edu.au

Jonathan Roberts, Kane Usher
CSIRO ICT Centre
P.O. Box 883, Kenmore, Queensland 4069
Australia
{Jonathan.Roberts, Kane.Usher }@csiro.au

Abstract

This paper shows initial results in deploying the biologically inspired Simultaneous Localisation and Mapping system, RatSLAM, in an outdoor environment. RatSLAM has been widely tested in indoor environments on the task of producing topologically coherent maps based on a fusion of odometric and visual information. This paper details the changes required to deploy RatSLAM on a small tractor equipped with odometry and an omnidirectional camera. The principal changes relate to the vision system, with others required for RatSLAM to use omnidirectional visual data. The initial results from mapping around a 500 m loop are promising, with many improvements still to be made.

1 Introduction

RatSLAM is a biologically inspired system for simultaneous localisation and mapping (SLAM)[1]. RatSLAM has been deployed successfully on Pioneer 2 mobile robots [Milford and Wyeth, 2003; Milford *et al.*, 2004; Prasser *et al.*, 2004] using a camera as an external sensor for mapping the environment. To date all work with RatSLAM has been conducted with forward facing cameras with narrow fields of view (< 50°). In this paper the initial steps towards migrating RatSLAM to an outdoor robot equipped with an omnidirectional camera are discussed. The majority of this work is occupied with changes to the vision system and the visual learning process in order to benefit from the increased field of view.

1.1 RatSLAM

RatSLAM is based on techniques used in computational models of the rodent hippocampus. The system contains three main modules: Pose Cells (PC); Local View (LV); and Path Integration (PI). The Pose Cell module is the heart of the system; it is an array of neural units that maintains the robot's pose estimate. Each unit in the PC can be considered to correspond to a region in (x, y, θ) space. The more activated a unit becomes, the more RatSLAM believes it is near that unit's position. The Path Integration module is responsible for adjusting the activated Pose Cells based on the robot's sense of self motion. Finally, the Local View represents the external sensors of the robot, in the form of another array of neural units. RatSLAM uses Hebbian learning to associate active LV units with active PC units to build a map in the form a set of vision – pose correspondences, which can later be used to adjust PC activity. The LV units respond to some form of cue in the robot's camera images. In the past, this has been the presence of particularly coloured objects or significant edges in the environment.

1.2 Appearance Based Visual Learning

RatSLAM associates the visual scene with a position. RatSLAM does not need a complicated analysis of the environment, provided that the Local View representation is the same or similar each time the robot visits a particular location. This can be accomplished by making LV units respond to particular visual appearances. The most recent RatSLAM vision system memorised views as it travelled through the environment [Prasser *et al.*, 2004]. Each LV unit corresponded to one of the learnt views and was activated when the camera image matched the unit's learnt view. This system worked quite well and the new outdoor system should benefit from functioning in the same way.

The notion of learning the robot's view and associating it with place has been used before in biologically inspired robots [Arleo *et al.*, 2001], and in other robot localisation systems, in particular systems using panoramic or omnidirectional cameras [Gonzalez-Barbosa and Lacroix, Ulrich and Nourbakhsh]. These systems for learning

1 This research is sponsored in part by an Australian Research Council grant.

panoramic views have used histogram matching. In a histogram matching scheme images are represented by a histogram of some set of attributes, for example histograms of greyscale brightness, colour, or edge direction and intensity. The primary advantage of histogram matching is that when the histogram is computed the information about the position of features in the image is discarded. Since rotation of an omnidirectional camera about its optical axis results in a rotation of the image about its centre, the histogram description of the image should not change as the camera rotates. Given that, on mobile robots, the camera's main axis is parallel to the robot's rotational axis, then a histogram description of an image should be invariant to the rotation of the robot. Aside from the invariance to robot rotation, histograms have the other useful properties:

- compact representation of images;
- a small level of translation invariance; and
- a number of convenient matching metrics, including χ^2 matching [Gonzalez-Barbosa and Lacroix, Ulrich and Nourbakhsh].

These properties make the histogram matching of visual appearance an attractive option for the deployment of RatSLAM on a robot with an omnidirectional camera.

1.3 Summary of Paper

This paper is primarily concerned with the steps needed to make RatSLAM work with outdoor data. The most significant changes were made to the vision system which is outlined in the next section. Other changes to RatSLAM are detailed in section 3 while the outdoor robot and other aspects of the experimental setup are described in section 4. RatSLAM trajectories and other results are in section 5 and conclusions and a summary of further work are in section 6.

2 Vision System

The new platform, an autonomous tractor, is equipped with an omnidirectional camera, providing approximately 320° of visual coverage around the robot with a small blind spot facing towards the rear of the robot (Figure 1).

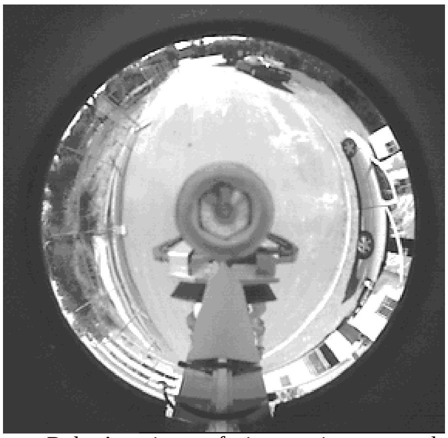

Figure 1: Robot's view of its environment through an omnidirectional camera, the sensor that RatSLAM uses to maintain its localisation. The robot is facing towards the top of the image. The support for the mirror obscures part of the robot's field of the view towards its rear.

The large level of visual coverage provides an opportunity to generalise from one robot pose to another. When RatSLAM was used with 50° field of view cameras, it was unable to extrapolate from one view to another. This forced the robot to travel to the new position to determine its visual appearance. This is particularly troublesome for robot rotation, for example while traversing a corridor in one direction RatSLAM is unable to gather the information needed to re-localise when travelling in the opposite direction (except by using a turn-back-and-look behaviour).

2.1 System overview

The vision system has four basic steps. Firstly, the image is converted to a panoramic representation and normalised in intensity (Figure 2). A histogram of the panoramic image's intensity is constructed and used to represent the visual scene to a matching mechanism. This histogram is compared to a learnt library of reference histograms to produce a shortlist of candidate views. Histograms that are sufficiently distinct from those already learnt are added to the library, as is the panoramic image that produced the histogram. Finally using the learnt panoramic images the orientation of the robot relative to when each of the candidate views was learnt can be recovered.

Figure 2: A panoramic image constructed from Figure 1. The area near the centre of Figure 1 has been discarded as it contains little information useful for re-localisation. The area of the panoramic image containing the camera blind spot or dead zone has been also cropped. The image shows a view of about 320° around the robot.

Histogram Matching

By using a rotation invariant representation of the camera input, it is possible to build a visual scene matching system similar to that already used successfully with RatSLAM. A well demonstrated view matching method for omnidirectional cameras is Histogram Matching.

While many different image attributes can be included in the histogram matching scheme, for the time being RatSLAM uses only normalised greyscale, which is matched using the Jeffrey divergence. The χ^2 statistic was originally used but it behaves badly when a histogram bin has a zero entries. Jeffrey divergence has previously been found to be a good metric for histogram based localisation [Ulrich and Nourbakhsh, 2000]. For two histograms, H and K, with entries h_i and k_i, the divergence is:

$$d_j(H,K) = \sum_i h_i \log\left(\frac{2h_i}{h_i + k_i}\right) + k_i \log\left(\frac{2k_i}{h_i + k_i}\right) \quad (1)$$

The divergence is calculated between the histogram of the current camera image and all of the learnt histograms. Learnt histograms whose divergence is below a threshold are considered to be reasonable matches and are sent as candidates to the next stage of the vision system. When there are no matches against the learnt histograms then the current histogram is memorised. In this way the robot will traverse its environment learning places that are visually distinct from each other. Each histogram effectively represents a small area of the environment. However since histogram matching uses visual appearance it is possible that one histogram may match many separate locations in the environment.

Recovery of Orientation

If only the histograms are stored then by using a rotation invariant recognition scheme we have lost the ability to recover the robots orientation during re-localisation. The solution is to store the image that generated the histogram. After recognising a scene by using histogram matching the reference image can be matched against the current camera image for various hypothetical rotations of the robot. The best fit will indicate the difference in robot orientation from when the reference image was acquired.

Determining rotation can be best accomplished from the panoramic image. In the unwarped image each column of pixels corresponds to a particular bearing, so simulating a change in orientation of the robot can be accomplished simply by shifting the image in the appropriate direction. Additionally the panoramic image does not retain the centre part of the omnidirectional image. This area is not thought to be useful for localisation as it usually contains only the road near the robot.

In the current system the panoramic images are reduced in resolution to lower the computational load of the matching process. The learnt image in memory is rotated through 360° in 36 discrete steps and compared to the current camera panoramic image using a Sum of Absolute Differences (SAD) metric. The number of rotational steps taken to find the best fit is reported by the vision system to the rest of RatSLAM, which uses it as a coarse measure of bearing.

A portion of the full panoramic image is a dead zone created by the support for the mirror. This area rotates with the robot and appears static in all camera images so it cannot be used in the matching process. The position of the dead zone must be tracked when rotating images to ensure that the matching metric is not computed for this area in either the camera image or the learnt template image (Figure 3). This will introduce a bias towards making matches where the two dead zones are not overlapping, since in these circumstances the SAD is calculated over a smaller area. Normalisation by the number of pixels that contributed to the matching solves this problem.

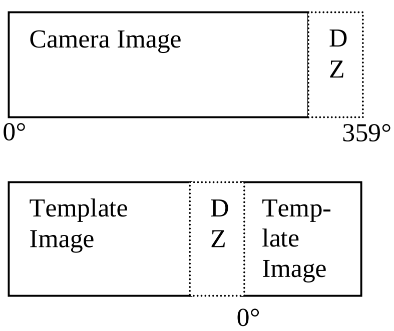

Figure 3: When rotating the template image to determine at what orientation it best matches the current camera image the camera's dead zone must be taken into account. The robot has no knowledge of what lies in the dead zone in either the current or template images so these areas cannot be compared.

3 Alterations to RatSLAM

Apart from the new vision system there were some other changes that needed to be made to RatSLAM. The most significant of these were made to the visual association process in order to again to take advantage of the omnidirectional camera. The spatial resolution that RatSLAM operates at was also reduced.

3.1 Local View

In the RatSLAM architecture the Local View (LV) cells represent the robot's external sensors to the rest of the network. As in recent indoor RatSLAM work each LV cell corresponds to one learnt template. The activation of the cell, V_i, increases as the difference between the camera image and the template, d_i, decreases:

$$V_i = \begin{cases} 1/(d_i + \varepsilon) & d_i \le d_{\max} \\ 0 & d_i > d_{\max} \end{cases} \tag{2}$$

Where d_{max} is a sensitivity parameter and ε prevents numerical blow out. Finally the activation of all of the LV cells are normalised to unity. The orientation of the robot relative to the template also needs to be expressed to the rest of RatSLAM; this is accomplished by a set of offset indices that accompany the LV cells. These offsets are the number of steps that the template image was rotated by to achieve the best match.

3.2 Visual Association

RatSLAM associates Local View units with activated units in the three dimensional Pose Cell array (Figure 4). The orientation of the image is also stored during the visual association process by determining which pose cell would be activated after undoing the rotation indicated by the vision system. In other words, the activated LV units are associated with the pose cells that should be active if the image was detected at no offset. If an active pose cell j located at orientation θ_j is to be associated with LV unit i, and offset, γ_i, then the LV unit is actually associated with a pose cell unit located at the same position but at orientation θ'_{ij}:

$$\theta'_{ij} = \theta_j - \gamma_i \tag{3}$$

Similarly, when re-localising, the LV injects energy to the pose cells that have been associated with the LV but shifted in the θ direction by an amount controlled by the LV unit's offset:

$$\theta_j = \theta'_{ij} + \gamma_i \tag{4}$$

The vision system reports offsets in the integer range of zero to 35, which is the same as the number of cells in the RatSLAM system's θ axis. This drastically simplifies the process of computing the array index of the pose cell after rotation.

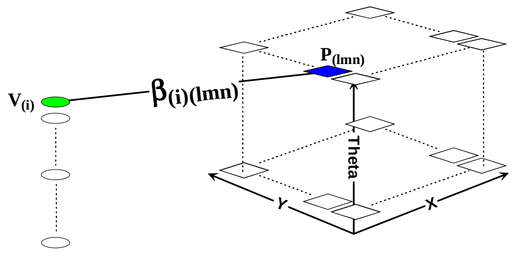

Figure 4: The local view network and pose cell network. Activated units in the local view, V_i, become associated with activated units in the pose cells, P, through learnt weighted connections between the two networks. The relative orientation of the robot reported by the vision system is used to dynamically change the Pose Cells each LV-PC link is connected to.

4 Experimental Setup

The experiments described in this paper were performed offline on data acquired by the robot during several traversals through an outdoor environment. The offline processing of data was a matter of experimental convenience; the vision and RatSLAM algorithms can be made to run in real-time. The outdoor robot is a tractor developed by the CSIRO [Usher *et al*, 2003]. Aside from the omnidirectional camera and odometry sensors the tractor is also equipped with a laser range finder and compass, neither of which are used by RatSLAM.

Figure 5: The autonomous tractor. The omnidirectional camera and mirror assembly is mounted on an elevated point above the front wheels.

4.1 Data Sets

Two data sets were acquired from the robot as it traversed a large out door looped road approximately half a kilometre in length. In the first experiment the robot completed two loops in a clockwise direction, while in the second the robot completed one traversal in each direction. There was a time delay of about 45 minutes between the first and second datasets. Camera images were acquired at 1 Hz and robot odometry at about 15 Hz. The robot's path takes it through both built up areas and unstructured natural areas. Aside from changes in lighting caused by clouds obscuring the sun, there are also humans moving around from time to time in the robot's field of view. These changes make the visual environment dynamic rather than static.

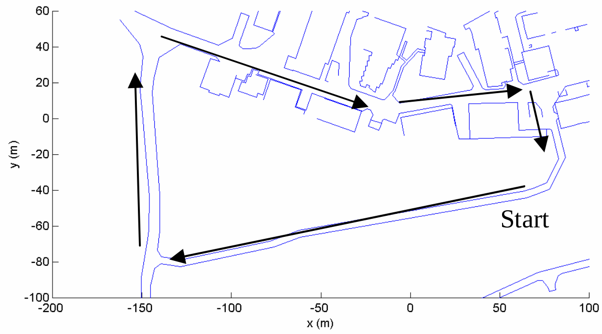

Figure 6: Approximate trajectory of the robot during data acquisition. The arrows indicate the path of the robot during the first lap of the environment.

4.2 System Parameters

The panoramic images are constructed with a resolution of 0.01 rad/pixel and occupy a solid angle of 5.6 rad horizontally by .52 rad vertically. The histograms have 64 bins and a Jeffrey divergence of 300 is considered to be the maximum divergence before learning a new histogram. The image matching is conducted at a lower horizontal resolution, where each pixel corresponds to 10° horizontally. If there was no camera dead zone then each reference image would 36 × 28 pixels in size. This resolution is chosen to simplify the image rotation process. The Sum of Absolute Differences matcher uses a difference of less than 150 as a match. Most vision parameters were selected by estimate and the only parameter that was extensively tuned was the histogram divergence. The RatSLAM visual association process had to be made weaker to reflect the increased ambiguity in the vision system. Finally RatSLAM pose cell resolution was increased from 250 mm to 2000 mm, reflecting the change in scale from previous indoor environments to the new larger setting.

5 Results

Two main results are acquired from the experiments: the Local View activity and the Pose Cell trajectory. RatSLAM maintains an approximately Cartesian mapping of the environment with an emphasis on consistent and repeatable results rather than physical accuracy. A good result for RatSLAM is when repeated robot trajectories in the real world correspond to repeated paths of active Pose Cells.

5.1 Vision Results

The activated Local View units during the experiment are shown in Figure 7. The system begins with no learnt LV units and numbers them in the order that they are learnt, causing the first rising line in Figure 7. At 300 seconds into the experiment the robot has completed one circuit of the loop and begins a second one, at this point previously learnt histograms are found again and the LV units begin to be activated again in the same pattern as on the first lap. The system continues to learn new histograms and recruit more LV units on the second lap, although at a reduced rate. This is partly because the some parts of the environment have changed visually between laps but also because the system has been set to a low level of generalisation. The continual increase in the number of learnt histograms will eventually finish once all of the likely views of the environment are learnt.

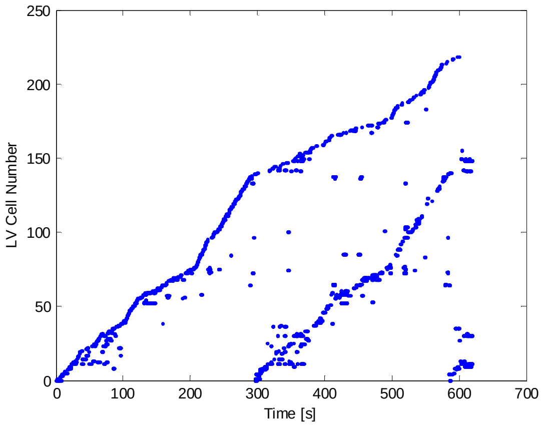

Figure 7: Activated Local View (LV) units versus time as the robot traverses the environment. For any given input zero or more templates may be found. Beginning at 300 seconds the robot begins to retrace its path and re-encounter previously seen views, resulting in repeated patterns of Local View activity.

5.2 Trajectory Results

The RatSLAM trajectory results for the forward path are shown in Figure 8. This figure shows the trajectory of both laps superimposed upon each other. There is a very good overlap of the second lap's trajectory upon the first indicating the network is using visual information to continuously correct its believed position in the world. RatSLAM is also able to recognise the looped nature of the path and adjust its position to account for the odometry errors that accumulate. This can be seen when after completing the first lap RatSLAM changes its believed position from point *a* to point *b*.

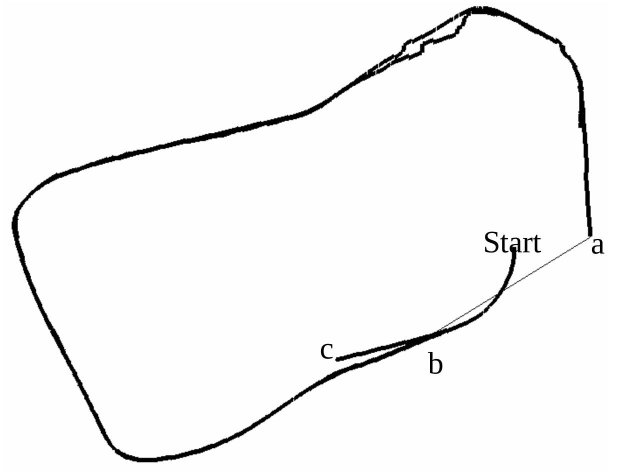

Figure 8: RatSLAM trajectory while completing two forward traversals of the route shown in Figure 6. During the second lap the RatSLAM recognizes that it is repeating a learnt path and adjusts its perceived position from a to b. Re-localisations are shown as thin lines.

The system does make one error early in the first lap where it incorrectly re-localises from point c to point b. Figure 9 shows camera images taken at both of these points and it is easy to see why these two points could be misclassified, especially when the images are compared in such a coarse way. Usually RatSLAM ignores short term vision errors like this, however this error persisted long enough for RatSLAM to decide that there was sufficient evidence to re-localise. This is more an error on the part of the vision system than of the SLAM component. In order to tell the images in Figure 9 apart a more detailed description will be needed than just a histogram of grey scale intensity.

Figure 9: Panoramic images acquired at point b (top) and point c (bottom) in Figure 8. A consistent misclassification of the bottom image causes RatSLAM to re-localise from c to b.

5.3 Travel in Both Directions

As a final test both data sets were combined into one large run in which the robot travelled in both directions along the loop. This is a particularly difficult data set as the end of the first run does not coincide with the start of the second run effectively causing a global kidnap of the robot. At the same time a large time window has elapsed unbeknownst to the robot during which the visual environment may have changed. The trajectory of the combined run is shown in Figure 10. The main purpose of this experiment is to test the ability of RatSLAM to localise the robot while travelling in a novel direction along a route that has already been learnt.

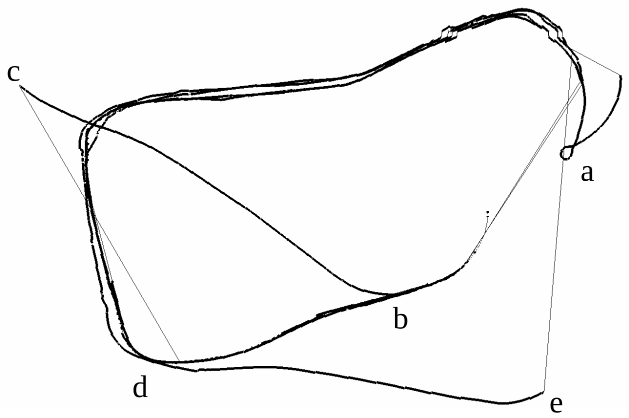

Figure 10: RatSLAM trajectory while completing three forward and one backwards traversals of the route shown in Figure 6. The robot performs a 180° turn at the point a and maintains a consistent path up to point d, despite travelling in the opposite direction to the one it has experienced. On the third and fourth circuits the robot is unable to maintain its localisation all of the time resulting in the incorrect path sections b-c and d-e.

The robot maintains good localisation on the reversed path from point *a* to point *d*, although it loses track of its position after point *d* until point *e*. This demonstrates that by using an omnidirectional vision sensor RatSLAM can generalise about its orientation. Previously RatSLAM Pose Cell trajectories would not necessarily have overlying forward and reverse paths.

On the third and fourth laps RatSLAM fails to remain localised between *b* and *c* and between *d* and *e*. This region of failure appears to have had a large variation in lighting between the first two and the last two laps. The area can be roughly characterised being a natural environment while the areas where localisation was well maintained could be described as built up areas.

If the experiment had been performed on a continuous data set and over a shorter interval of time the results would have been better. The result highlights the fact that the current brightness normalisation process is inadequate for dealing with long term lighting variations. Correcting this would be necessary to give the system the capability to operate outdoors for extended periods of time.

6 Conclusion

We have shown that it is at least feasible to transfer the existing RatSLAM system to an outdoor environment. The major changes were to the vision system to take advantage of the omnidirectional camera. The vision system still operates under an appearance based scene learning scheme albeit using a different set of primitives to previous RatSLAM vision systems. There are points of failure in the system that could be ironed out by adopting a richer set of feature detectors. As expected histogram matching provides a quick method of visual appearance learning that is useful for robot localisation and mapping.

6.1 Future Work

The system could be expanded to use other image attributes such as colour in the histogram recognition stage. Other histogram matching techniques such as Earth Mover's Distance should be investigated as well as more sophisticated methods for ensuring illumination invariance. Also the system can be easily blended with the edge detection scheme previously used with RatSLAM [Prasser *et al.*, 2004] at both the histogram stage and the orientation recognition stage. This will be the immediate next stage of development.

RatSLAM's capabilities could also be tested on less topologically structured outdoor environments, for example, large open areas without obvious paths or routes. Finally, RatSLAM could be fully deployed to operate on the robot in real time.

References

[Arleo *et al.*, 2001] Angelo Arleo, Fabrizio Smeraldi, Stéphane Hug, and Wulfram Gerstner, "Place Cells and Spatial Navigation based on Vision, Path Integration, and Reinforcement Learning." *Advances in Neural Information Processing Systems*, 2001.

[Gonzalez-Barbosa and Lacroix, 2002] J.-J. Gonzalez-Barbosa and S. Lacroix. "Rover localization in natural environments by indexing panoramic images." *Proceedings of the International Conference on Robotics and Automation 2002,* IEEE, 2002.

[Milford and Wyeth, 2003] Michael Milford and Gordon Wyeth. "Hippocampal Models for Simultaneous Localisation and Mapping on an Autonomous Robot." *Proceedings of the Australasian Conference on Robotics and Automation (ACRA)*, 2003.

[Milford *et al.*, 2004] Michael Milford, Gordon Wyeth, David Prasser "RatSLAM: A Hippocampal Model for Simultaneous Localization and Mapping." *Proceedings of the International Conference on Robotics and Automation, in press*, IEEE, 2004.

[Prasser *et al.*, 2004] David Prasser, Gordon Wyeth and Michael Milford. "Biologically Inspired Visual Landmark Processing for Simultaneous Localization and Mapping." *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (accepted)*, 2004.

[Ulrich and Nourbakhsh, 2000] I. Ulrich and I. Nourbakhsh. "Appearance-based place recognition for topological localization." *Proceedings of the International Conference on Robotics and Automation 2000,* IEEE, 2000.

[Usher *et al*, 2003] Kane Usher, Peter Corke and Peter Ridley, Visual servoing of a car-like vehicle -- an application of omnidirectional vision, *Proceedings of the International Conference on Robotics and Automation 2003,* IEEE, 2003.