Automatic Landmark Selection for Navigation with Panoramic Vision

Simon Thompson, Alex Zelinsky and **Mandyam Srinivasan**
Robotic Systems Laboratory
Department of Systems Engineering, RSISE
The Australian National University
Canberra, ACT 0200 Australia
[simon|alex]@syseng.anu.edu.au, srini@rsbs.anu.edu.au

Abstract

The use of visual landmarks for robot navigation is a promising field. It is apparent that the success of navigating by visual landmarks depends on the landmarks chosen. This paper reviews a monocular camera system proposed by [Bianco and Zelinsky, 1999] which automatically selects landmarks and uses them for localisation and navigation tasks. The monocular system's landmark selection policy results in a limited area in which the robot can successfully localise itself. A new landmark navigation system which uses a panoramic vision system to overcome this restriction is proposed and the first steps taken in the development of the new system are reported.

1 Introduction

Visual navigation is one of the key problems in making successful autonomous robots. Vision as a sensor is the richest source of information about a mobile agents enviroment and as such contains information vital to solving navigation problems. One limit to visual navigation is the narrow field of view offered by normal monocular camera systems. Obviously the greater the field of view of the camera the more information about the environment can be extracted through visual processing.

Another problem is that vision provides us with too much information. In fact one of the main tasks of vision science is to work out which parts of the information is necessary for the given job and to discard or put aside that which is not.

1.1 Panoramic Imaging for Navigation

Panoramic imaging can solve the first problem mentioned above [Yagi *et al.*, 1994]. By the use of convex mirror shapes greater fields of view can be achieved. Panoramic mirrors give the full 360 degrees of horizontal visual field and can get over 140 degrees in the vertical field as well as shown in figure 1. This increase in the visual field comes at the cost of resolution in some parts of the picture however, as it must be captured with a camera with a normal field of view.

Figure 1: A panoramic image

With the full 360 degree horizontal field of view there is more information from which to achieve the navigation tasks of localisation, path planning and path traversal. [Matsumoto *et al.*, 1997] use panoramic pictures to navigate a robot down corridors using a view sequence approach. The robot memorises a sequence of panoramic images along a route, acquiring new images once the current view drops below a set correlation threshold with the last image. This works well in the corridor environment, with the environment only changing sufficiently to warrant a new view being memorised every 1-2 meters. In an open room however the rate of acquisition would increase dramatically as the panoramic image would change sharpely over small distances.

1.2 Natural Landmarks for Navigation

An attractive solution to this problem is the use of landmarks for the navigation tasks. Using landmarks in navigation requires only parts of an environment to be stored and can contribute to accurate localisation with the aid of triangulation techniques. The use of landmarks would overcome the acquisition rate problem by not requiring the entire visual field to remain constant, therefore maximising the area in which one set of landmarks can be used for localisation. Landmarks however, rely on the template matching of the stored landmarks and the landmarks in the current image. Using normal correlation techniques this matching decreases in reliability for smaller landmarks as smaller patterns are more likely to undergo change due to changes in the environment and distortion effects due to movement of the robot. Matsumoto et al's [1997] view sequence approach can be thought of as using one big landmark for the navigation tasks, this avoids the problems of small templates but suffers from high image acquisition rates and high storage cost.

Obviously the success of navigation using small landmarks depends on the choice of landmarks [Thrun, 1996]. [Bianco and Zelinsky, 1999] describe a monocular system where landmarks are chosen on the basis of their reliability. To be selected, landmarks must display uniqueness in the immediate surroundings and the ability to remain reliable as the robot moves through the environent. To this end the system monitors the reliablity of landmarks through a series of small arcs. This 'turn back and look' behaviour is inspired from observations of honey bee flight patterns [Collet and Zeil, 1996] [Lehrer, 1993].

This paper reviews Bianco et al's monocular system and looks at the applicability of panoramic camera systems to visual navigation using landmarks. In light of the increased field of view available to the robot, both the landmark selection strategy and the localisation procedures need to be adjusted. The first steps taken in the developement of a panoramic system are presented and plans for further work are detailed.

2 Bianco's Monocular System

Bianco et al's [1999] system uses a monocular camera which captures NTSC format video images. These images are input into a Fujitsu Tracking Card which contains specialised hardware for template matching between regions of the video image. The two main stages of the system are the landmark selection phase and the localisation phase. Landmarks in this case refer to small unprocessed regions, or templates, of the image captured by the camera. The same hardware is used in both sytems except for introduction of the panoramic sensor in the new system. Details of the Fujitsu Tracking Card can be found in [Bianco, 1998] [Heinzmann and Zelinsky, 1998].

2.1 Landmark Selection

The process of landmark selection is aimed at increasing the localisation and navigation abilities of a mobile robot in subsequent exposure to the environment. This means that landmarks must be reliable, strongly identifiable, they must be distributed throughout the image to minimise the error in navigation calculations, and they must be able to withstand distortions due to temporal and translational distortions. To this end Bianco et al [1999] selects a number of landmarks from different sections of the static environment based on their reliability. This reliability is then tested in a dynamic environment by moving the robot in a series of short arcs.

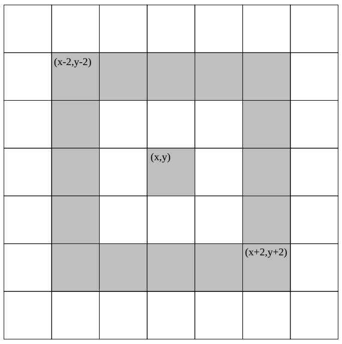

Figure 2: The Valley Method for automatic landmark selection

Static Reliability of Landmarks

The static reliability of a landmark in Bianco et al's [1999] model, is determined by the uniqueness of the landmark within its local region of the image. Local uniqueness is defined as the degree to which the landmark template differs from the area of the image immediately surrounding the landmark. This approach is based on the 'The Valley Method' proposed by Mori et al [Mori *et al.*, 1995] to generate attention tokens in a scene. Bianco et al adapts this method for the present task of automatic selection of landmarks as shown in figure 2. The candidate landmark (having its top left corner in pixel (x,y)) is correlated with the 16 possible templates 2 pixels removed from the landmark template (ie those with top left corners (x-2,y-2) to (x+2,y+2)). Comparing the minimum matching distortion resulting from the match of the landmark and the sixteen surrounding templates with the distortion of the landmark correlated with itself, the local uniqueness of the landmark is found. More formally:

$$r = 1 - g/g'$$

where r is the reliability of the landmark, g is the distortion of the landmark matched with itself, and g' is the minimum matching distortion from the surrounding circle of pixels. Given that g should only result in distortion due to noise, then the higher the distortion of the minimum of the surrounding templates, the steeper the valley in the distortion matrix and subsequently the more unique the local template should be (see figure 3).

The most reliable landmarks are then subjected to the dynamic testing phase.

Note that the size of the landmarks selected might also play a role in their reliability. As mentioned in the in-

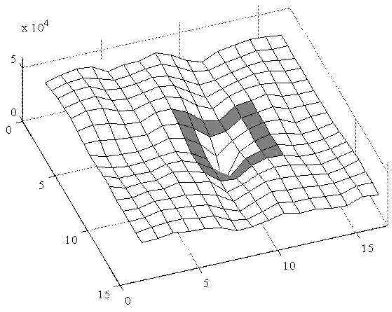

Figure 3: An example of the valley in the distortion matrix caused by a reliable landmark. Figure from Bianco et al 1999

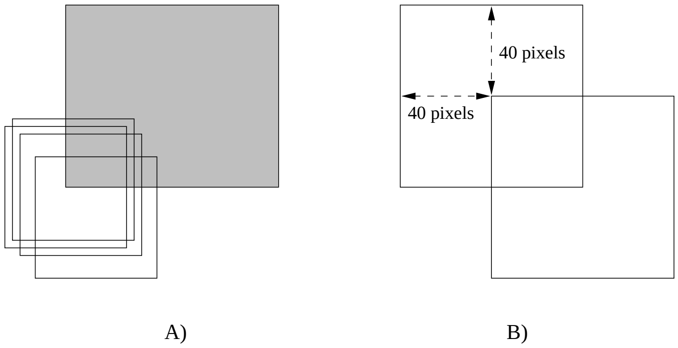

Figure 4: Landmark grouping within sectors. A) Landmarks grouped about an interesting feature. B) Possible landmarks are evaluated every 40 pixels.

troduction, [Matsumoto *et al.*, 1997] effectively use the entire image as a landmark and this proves to be very reliable. Problems arise however in a dynamic environment where parts of the scene might become occluded. A number of smaller distributed landmarks would seem to solve this problem. What size landmarks should then be chosen? Computing the reliability of different sized landmarks in each possible image location would dramatically increase the computation time required for automatic landmark selection. Bianco et al [Bianco and Zelinsky, 1999], for reasons explained in the Localisation section below, chose landmarks of 80x80 pixels (actually stored as 16x16 templates with a magnification factor of 5).

Landmarks in Image Sectors

Bianco et al [1999] divides the visual field up into a number of sectors. An equal number of landmarks is chosen from each sector. This ensures a good distribution of landmarks for triangulation purposes and also maximises the chance of a large proportion of visible landmarks when landmarks are being tracked or located in dynamic scenes. If a moving object in the environment results in the landmarks from one sector being occluded in the visual field then the equal distribution of landmarks throughout sectors maximises the possibility that other landmarks will be visible. Landmarks can still become grouped within a sector as shown in figure 4 part A). An interesting feature, in this case the corner of the grey square, can result in landmarks that are very close together being selected. To avoid this Bianco et al only evaluate landmarks in the image every 40 pixels (half the size of the landmarks) seen in figure 4 part B). This also further reduces the amount of computation time needed for the search.

Dynamic Reliability of Landmarks

To be good navigational cues, landmarks need to be reliable in a changing environment. They need to remain strongly identifiable under small changes in lighting and shifts in perspective. To this end Bianco et al test the reliability of the statically selected landmarks throughout a series of small arcs. This Turn Back and Look (TBL) phase is inspired by observations of wasps and bees on flights away from the hive [Collet and Zeil, 1996] [Lehrer, 1993]. Figure 5 shows the path taken by Bianco et al's [1999] robot on the TBL phase. The robot moves further away from the original, or goal, position while keeping the camera oriented towards it. Potential landmarks are tracked along this path and their reliability measures are evaluated at approximately 400 steps along the way. Each landmark's dynamic reliability is given by the average of the reliability measures taken over the entire path. If this averaged value is greater than a set threshold the landmark is kept, if it is less it is discarded.

The question arises why Bianco et al use the local uniqueness reliability measure in the dynamic phase of landmark selection. It would be intuitively more sensible to chose landmarks which had the minimal matching distortion between the current and original landmarks during the TBL phase. In fact Bianco et al's system monitors both the reliability and the distortion measures. This strange choice will be discussed further in the Localisation section below.

2.2 Localisation

Localisation is essentially locating the selected landmarks in the current image. Knowing the positions of the landmarks from both the current and the original scene, navigation vectors pointing towards the goal position can be calculated. Note that in these experiments it is assumed the robot is in an area which the landmarks are visible. Localisation in the monocular model is done by a brute force global search of the visual field, with

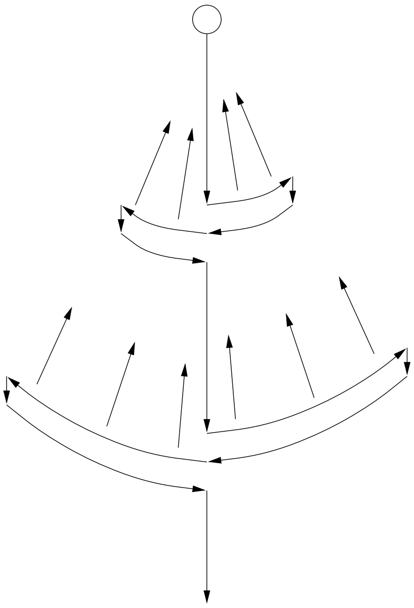

Figure 5: Bianco's TBL phase for dynamic landmark selection

the position of a landmark being determined by the area of the visual field which most closely matches the stored landmark.

By itself this strategy would only result in successful localisation in a small area surrounding the goal position. If the robot is positioned further away from the goal position the landmarks will have a decreased size in the current view and thus matching will deteriorate. Similarly positioning the robot closer to the view will affect matching by increasing the size of landmarks. Moving the robot further to the left or right of the goal position will cause perspective shifts in the landmarks, again leading to poor matching results.

Figure 6: An example of the landmarks Bianco's model selected - figure from Bianco et al 1999

Bianco et al [1999] addresses the change of size problem by conducting matching of landmarks on various magnifications of the stored landmarks. It was mentioned above that the selected static landmarks were of size 80x80 pixels but were actually stored as 16x16 templates with an associated magnification of five. To search for landmarks which have grown or shrunk due to motion towards or away from the scene Bianco et al simply conduct the search with landmarks of magnifications ranging from 1 to 8.

The change in the landmarks due to perpective shift, however, is not so simple to solve. Indeed the inspiration behind this approach seems to side-step this issue. The homing flights of bees or wasps are likely to lie along straight lines and have more need to resolve magnification differences than those due to perspective shift. No attempt at specifically addressing this problem in the monocular system apart from the dynamic relaibility checking. This is evidenced in the conical shape of the area in which the robot can correctly localise.

By basing the path of the TBL phase on the flight of bees and wasps Bianco et al incorporate this limitation into his system. Admittedly the field of view offered by the monocular camera also contributes to this limitation. This observation also provides insight as to the choice of local uniqueness as the dynamic relaibility measure discussed in the previous section. If a landmark is not going to vary significantly in perspective shift it makes sense that if it retains its strongly identifiable nature over time, then simple magnification differences will not alter its basic structure that much.

Once the landmarks have been located the navigation vector for reaching the goal can be calculated. For each landmark a displacement vector is found which is simply the difference between the current position of the landmark from the original position. This vector is then weighted by the difference in magnification of the current landmark compared to the original. Landmarks which have a large differnece in magnification are given more weight than those with small differneces. This reflects the varying distances from the goal position inferred by diferent magnifications.

These vectors then need to be *fused* together to give an overall navigation vector. For reasons of simplicity Bianco et al do this by a weighted average. The weighting is based on the reliability of the individual landmarks. Fusing the landmarks in this way means that individual landmark matching errors can be corrected for by averaging while more trust can be put in those landmarks which show a high reliability.

The resulting navigation vector has to be converted in to some sort of motor commands that the robot can understand. This of course depends on the robot used and as such will not be described here. More details on this and an indepth description of the navigation vector calculations can be found in [Bianco and Zelinsky, 1999] [Bianco, 1998].

3 Panoramic System: First Steps

The goal in using a panoramic vision system in landmark navigation is to increase the area in which a mobile robot can successfully localise with a given set of landmarks. Thus increasing its ability to learn and navigate through new environments. To this effect, a panoramic system is being developed based on Bianco et al's [1999] monocular model, the first steps of which are reported here. At this point the system automatically selects static landmarks in a panoramic view and by using them to localise can cancel out the effects of rotation of the camera.

The image from the panoramic camera is dewarped and the resulting image is a full colour 320x120 pixel panoramic view.

3.1 Landmark Selection

It is assumed in the landmark selection phase that time is not a crucial factor as the robot is in an 'exploratory' phase. In particular during the static phase the robot is not moving at all and time restraints can be relaxed. It is the intention, however, to maintain a high enough frame rate to ensure that basic optic flow and obstacle avoidance behaviours can be maintained during this stage.

Static Reliability of Landmarks

Landmarks in the panoramic system are selected in much the same way as in the monocular system. Local uniqueness is used to identify potential landmarks. The selected landmarks are also subjected to a test of global uniqueness. Global uniqueness refers to the landmarks uniqueness throughout the entire image. Bianco et al's [1999] system does not check a landmark's global uniqueness. Instead relying on the assumption that nothing within the field of view will be similar enough to warrant the computation time required. This assumption is reasonable with the visual field from a normal camera, but begins to break down with a panoramic field of view, the highly structured and repetitive office environment and the loss of resolution due to the panoramic system. Also given that time is not an important constraint at this stage of the robots behavioural cycle, it makes sense to make a full search of the image to determine the global uniqueness of the landmarks.

The local and global reliability measures are combined to give the static reliability measures for the potential landmarks.

Landmarks in Image Sectors

Like Bianco et al's [1999] model the image is divided into sectors and the landmarks are selected in equal numbers from those sectors. The panoramic model however can be divided up into vertical sectors which offers the prospect of powerful triangulation properties for localisation. The distribution of landmarks throughout 360 degrees of visual field also decreases the chance of localisation errors due to objects moving through the environment.

Dynamic Reliability of Landmarks

The discussion of Bianco et al's [1999] monocular system showed that the TBL phase was crucial in determining the area in which correct localisation could occur. With an increase in the field of view this phase becomes even more important. To extend the conical localisation space of the bee inspired TBL phase it is obvious that we simply need to alter the path to incorporate all possible directions. Figure 7 part (A shows the arcs of the old TBL extended into circles. It may even be enough just to simply move through a series of line segments such as in part B) of the figure.

The TBL phase now causes landmarks in the image to undergo changes due to perspective shift which were absent in the monocular system. Does this mean that the dynamic reliability measure of local uniqueness will no longer work? It seems there is even more reason to use a distortion measure in assessing dynamic reliability. It is interesting to note that for any point along the TBL phase, there will be at least two sectors (depending on how many are defined) whose landamrks are relatively unaffected by perspective shifts. Also local uniqueness and distortion measures applied throughout the TBL phase should pick out landmarks with the smallest perspective shifts; i.e. the landmarks the furthest away from the camera.

An alternative to these measures is to directly test sets of landmarks localisation performance at steps along the path (the nodes shown in figure 7 part A) although this would prove computationally expensive). A full investigation into all these considerations is needed.

3.2 Localisation

Landmarks are located using a brute force search. Because the TBL phase of landmark selection has not yet been implemented, localisation experiments have been limited to resolving rotation. This is a trivial task but an important one none the less. While acting as a useful testing stage for landmark location, it also allows a robot to orient itself with respect to its environment. Suppose a robot selected a group of landmarks at a location and then headed off in a given direction. To repeat this manouver the robot would need to orient itself as it was initially before being able to steer in the correct

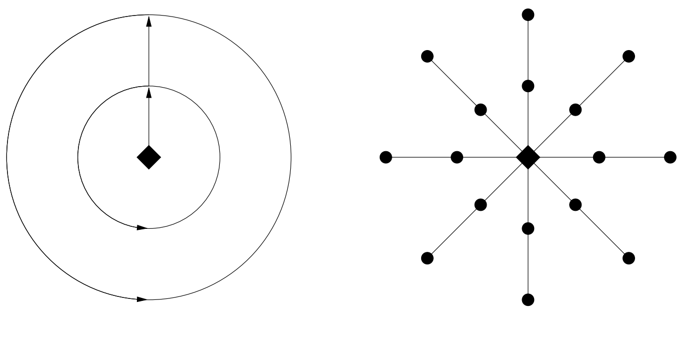

Figure 7: TBL phase for panoramic system A) Arcs extended to circles. B) Proposed TBL phase path.

direction. The current system simply determines the displacement vectors between the original and current landmarks along the x-axis of the image, averages these values and produces a navigation vector accordingly.

Localisation and navigation in the panoramic system can be approached with more sophistication than the monocular model. While a brute force search approach is currently used to locate landmarks, it is possible to use found landmarks to decrease the search space for remaining landmarks. The geometrical relationships between the vertical sectors of the panoramic image now allow the consideration of *sets* of landmarks rather than single ones. This is espescially important in the calculation of navigation vector as vectors from landmarks which lie along the direction to the goal, the goal line, should be more trustworthy than those whose landmark lie to either side. This reflects the degree of perspective shift in the landmarks from different sectors and would not be evident in the weightings given by the reliability measure. Of course an intial goal vector must first be found in order to know which landmarks lay along the goal line. Assuming the first navigation vector is not totally wrong, a second vector can be calculated which gives stronger weighting to the goal line landmarks. The panoramic system should result in a localisation space which surrounds the goal and extends equally on all sides.

4 Conclusion

Bianco et al's [1999] monocular system for navigation using visual landmarks is based on the concept of using reliable landmarks. By using landmarks which are both statically and dynamically reliable the system navigates back to the position at which the landmarks were selected. The area in which this task is preformed successfully is restricted by the monocular camera system and the path of the TBL phase. A panoramic system is being developed which will overcome these restrictions by increasing the field of view and extending the TBL phase. The system is only in its first stages but intial results suggest that the panoramic approach is well suited to the task.

References

[Bianco and Zelinsky, 1999] G. Bianco and A. Zelinsky. Biologically inspired visual landmark learning and navigation for mobile robots. Submitted to IROS 99, 1999.

[Bianco, 1998] G. Bianco. *Biologically-Inspired Visual Landmark Learning and Navigation for Mobile Robots.* PhD thesis, Dipartimento di Electtronica per l'Automazione, Universita Degli Studi Di Brescia, 1998.

[Collet and Zeil, 1996] T.S. Collet and J. Zeil. Flights of learning. *Journal of the American Psychological Society*, pages 149–155, 1996.

[Heinzmann and Zelinsky, 1998] J. Heinzmann and A. Zelinsky. A visual interface for human-robot interaction. *Field and Service Robotics*, pages 540–547, 1998.

[Lehrer, 1993] Lehrer. Why do bees turn back and look? *Journal of Comparative Physiology*, 172(A):549–563, 1993.

[Matsumoto *et al.*, 1997] Y. Matsumoto, M. Inaba, and H. Inoue. Memory-based navigation using omni-view sequence. In *Int. Conf. of Field and Service Robotics*, pages 184–191, 1997.

[Mori *et al.*, 1995] T. Mori, Y.Matsumoto, T. Shibata, M. Inaba, and H. Inoue. Trackable attention point generate based on classification of correlation value distribution. In *JSME Annual Conf. on Robotics and Mechatronics (ROBOMEC 95).*, pages 1076–1079, 1995.

[Thrun, 1996] S. Thrun. A bayesian approach to landmark discovery and active perception in mobile robot navigation. Technical report, School of Computer Science Carnegie Mellon University, 1996.

[Yagi *et al.*, 1994] Y. Yagi, S. Kawato, and S. Tsuji. Real-time omnidirectional image sensor (copis) for vision-guided navigation. *IEEE Transactions on Robotics and Automation*, 10(1):11–21, 1994.