Omni-directional Visual Navigation

Niall Winters

Computer Vision and Robotics Group,
Department of Computer Science,
University of Dublin, Trinity College,
Dublin 2 - Ireland.

Niall.Winters@cs.tcd.ie

José Santos-Victor

Instituto de Sistemas e Robótica,
Instituto Superior Técnico,
Av. Rovisco Pais, 1,
1049-001 Lisboa - Portugal.

jasv@isr.ist.utl.pt

Abstract. This paper considers a method for the visual-based navigation of a mobile robot in an indoor environment. The sole sensor used is an omni-directional camera. Navigation is achieved by combining appearance based methods and visual servoing. A database of omni-directional images is obtained during a training stage. This database maps the environment in a *topological* sense. Since the image set is relatively large, it is compressed using Principal Component Analysis. A low dimensional eigenspace is constructed, from which the robot can determine its *global* topological position along a path. Visual servoing is applied to ground plane dewarped images in order to control *locally* the pose of the robot. The method of obtaining these images is outlined, as is the closed loop control scheme used to keep the robot moving in the centre of a corridor. Finally, real world experimental results for both static and relatively dynamic environments are presented and proposals made for the extension of the system.

1 Introduction

1.1 Motivation

The overall aim of our work is to confer on a mobile robot goal-driven motion within an indoor environment. We desire that the robot navigate autonomously to a specified destination using only a single camera for sensing. In order to achieve this task, a number of sub-destinations may also, but not necessarily have to be visited. Therefore, we require a relevant representation of the environment. In addition, a robust method of controlling the robot's pose is required. This paper presents what we believe to be an appropriate solution to this problem: the combination of appearance-based methods and visual servoing.

Early research into the navigation of a mobile robot using vision centered on the 3D reconstruction of a scene. While this approach is accurate, it is time consuming and has been shown to be unreliable for almost all real world situations.

Other non-visual sensors have also been used to build up geometric representations of the environment. These are highly complex, with the input coming from a number of sensors (sonar, laser, etc.) which is then combined, using a suitable filter, into an environmental map. The main criticism of this approach is that algorithms using the map assume an accurate metric representation in a global coordinate system. In practice this is problematic to achieve as the map usually contains irrelevant information, thus adding to the computational load. In addition, these systems are slow and not very suitable for real-time navigation.

To overcome the need for multiple sensors, active vision [1] takes the approach of altering the robot's viewpoint rather than passively observing it. The structure of a specific scene can also be taken into account so as to build *customized* algorithms. These work well at the local level but since they have no global map they encounter difficulties with global tasks.

Our approach is somewhat different from those detailed above. Our motivation is to take the best elements of each and merge them, thus maximizing the advantages of both, while minimizing the disadvantages. The use of appearance based methods [9] [13] affords us the opportunity to obtain a *qualitative* measure of the robot's *global* position within its environment. Visual servoing is then applied to a remapping of the current image to control *locally* the pose of the robot relative to image features.

Appearance based methods have been an active area of research in the recent past. Localization [8] has been achieved by matching images using the frequency domain of the Fourier Transform. Horswill [7] used the actual views as landmarks. Zheng [16] took the approach of using dynamic programming to determine which path to take from a number of choices, even though this had a very high matching cost. Neilson [12] concentrated on the translation of database images into motion commands. The work most closely related to ours is that of Vassallo [14]. She too combined appearance based methods and visual servoing, although her images, matching scheme and method of servoing were different from those detailed in this paper.

This work focuses on using omni-directional images as an implicit topological representation of the environment, i.e. our global map. Progression towards the goal can simply be obtained by comparing the current view with images acquired *a priori*. For example, a single corridor can be represented by a sequence of images. This sequence is a *link* in a graph structure. A number, $N \geq 2$, of such sequences are connected via *nodes*. Nodes then represent specific places in the real world where actions (such as turn right, for example) can be taken. Thus nodes include corridor intersections, entrances to offices, exits etc. Whenever the robot reaches a nodal point the required action can easily be launched. Topological maps are a parsimonious representation of the environment in that they are not complex and scale easily, one subgraph can always be connected to another. Our topological map is represented as a manifold in a low dimensional eigenspace obtained from Principal Component Analysis (PCA).

Once the topological representation of the environment is available, we then resort to visual servoing on local features to control the robot trajectory. It is often the case that the topological map can easily be associated with some qualitative structural elements of the environment. These can be used to control locally the robot trajectory. In this paper, we control the robot heading and position by servoing on the corridor guidelines. These are extracted from ground plane dewarped omni-directional images. Thus the robot maintains its central position, correcting both orientation and position errors. Other examples of this local servoing scheme could include centering the robot relative to a door frame, orienting the robot relative to a specific object to grasp, etc. We show that the methods of extracting the image signals do provide meaningful information for closed loop control purposes.

1.2 Omni-directional Vision

Omni-directional vision is the name given to the generation of panoramic images from a single camera. A system provides a 360° view of the robot's environment around the vertical in addition to $\sim 30°$ above the horizontal. They have becoming popular over the last number of years and are now used in many applications including autonomous navigation and video conferencing. Their many advantages include:

- Landmarks are always in the field of view, except for occasional occlusion. This is advantageous when utilizing topological maps in that the more information the image contains the more stable it is.
- Panoramic views are acquired in a *single frame*.
- They have no moving parts, thus consume less power, vis-*à*-vis a pan-tilt mechanism and therefore have increased reliability.
- Circles perpendicular to the axis of rotation are invariant under rotation.

Their main downfall is that images acquired are of lower resolution than standard. Therefore, they are not suitable in situations where high resolution is a prerequisite, personal identification for example.

This paper is outlined as follows: Section 2 presents a brief overview of our system. Section 3 details eigenspace construction and position estimation, while Section 4 explains how ground plane dewarped images are recovered. In Section 5 we present our preliminary experimental results. Finally, in Section 6 we draw our conclusions and outline the future direction of our research.

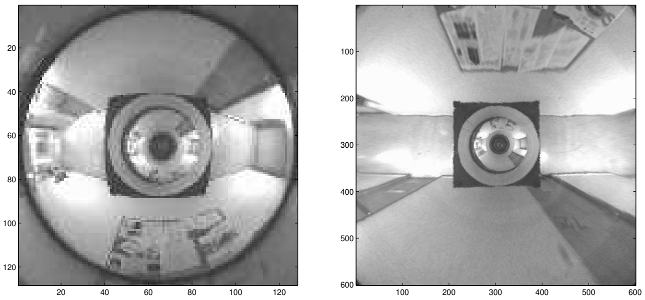
2 System Overview

As shown in Figure 1(a) our system is a TRC Labmate from HelpMate Robotics Inc. Mounted upon this is an omni-directional vision system (Figure 1(b)) built in-house. It contains a Cohu CCD camera pointed upwards, looking at a spherical mirror. Grayscale images were captured using a Captivator frame grabber at full resolution, 768x576 pixels. These were sub-sampled to 128x128 images for PCA and 600x600 for visual servoing. The images which can be obtained from the system are of three types: warped panoramic, ground plane dewarped and dewarped panoramic. These three types are shown in Figure 2. Hicks [5] has demonstrated how ground plane dewarped images can be obtained *directly* from a specialised mirror.

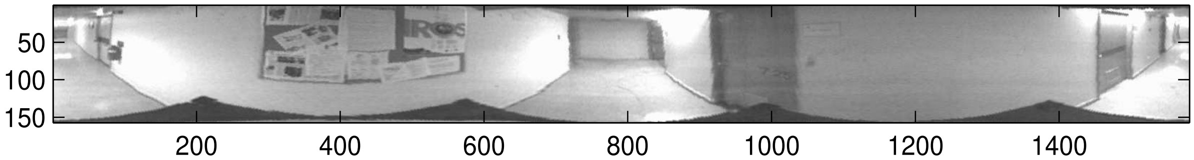
As the robot moves down the corridor it captures two images, one warped panoramic and one ground plane dewarped. The warped panoramic image is passed to the appearance based system, so as the robot

(a) (b)

Fig. 1.: (a) Labmate mobile robot with attached omni-directional camera. (b) Close-up of the camera.



(a) (b)



(c)

Fig. 2.: The three types of images that can be obtained from the omni-directional vision system: (a) warped panoramic, (b) ground plane dewarped and (c) dewarped panoramic.

can determine how far it has progressed along the path to its goal. The ground plane image is passed to the visual servoing system. The sequence is as follows:

– Determine the topological position by projecting the current warped omni-directional image into the eigenspace.
– Edge detect within bounding boxes of the ground plane image to find corridor guidelines. These bounding boxes are positioned in the next image using prediction.
– Calculate the robot's orientation and local position, relative to the centre of the corridor.
– Generate a reference angular velocity command in order to minimize the robot's orientation and position errors.
– Relying on the appearance based system, make decisions at specific locations, e.g. turn left at the end of the corridor.

3 Eigenspace Construction & Position Estimation

3.1 Overview of our Method

Our approach to constructing an eigenspace representation of the environment is based on the method by [11]. A large image set is acquired and compressed using Principal Component Analysis. The eigenvectors of the covariance matrix of the set of images forms an orthonormal basis. These are the principal components. There are as many of them as there are images. For example, if we have 100 images, then we are dealing with a 100D space. All the eigenvectors contain components from all of the images and so are sometimes referred to as *eigenimages*. If exact classification is required, all eigenvectors need to be utilized. In fact, for almost all situations this is not the case. Hence, we need only use a small number (typically ten to twelve) of these eigenvectors as an orthonormal basis for a *lower* dimensional space. Which eigenvectors to keep is determined by their associated eigenvalues. For our experiments the first ten to twelve eigenvectors are enough. Position is estimated by projecting the most recently captured (warped omni-directional) image into the low dimensional eigenspace. The closest point, on the manifold, to this projected point is the optimal position estimate.

3.2 Eigenspace Construction

This subsection outlines the method used to construct the eigenspace. N images are read in a line-scan manner into an array $\mathbf{X}$, normalized and converted from 128 x 128 matrices to 16384 x 1 column vectors. The average of all the images, $\mathbf{a}$ is determined and subtracted from all N images thus giving a matrix $\mathbf{B}$. The eigenvectors, $\mathbf{e}$ of the covariance matrix, $\mathbf{R} = \mathbf{B}\mathbf{B}^{\mathbf{T}}$ form the orthonormal basis we require. Unfortunately, computing them is highly computationally intensive. Fortunately, a solution [10] to the problem lies in computing the singular value decomposition (SVD) of $\tilde{\mathbf{R}} = \mathbf{B}^{\mathbf{T}}\mathbf{B}$. Thus, we obtain singular vectors $\tilde{\mathbf{e}}$ and values $\tilde{\lambda}$ which are then converted to the required eigenvectors and eigenvalues by the equations:

$$\lambda_i = \tilde{\lambda}_i$$

$$\mathbf{e}_i = \tilde{\lambda}_i^{-1/2}\mathbf{B}\tilde{\mathbf{e}}_i.$$

In the next stage a lower dimensional subspace is built using only the first M, ($M \ll N$) eigenvectors. This lower dimensional eigenspace has the advantage of accounting for most of the variance in the images. Each image, minus the average, is projected into this space by the following dot product to obtain a set $\mathbf{T}$ of projected points:

$$\mathbf{T}_i = \mathbf{e}_i^T.(\mathbf{X}_i - \mathbf{a})$$

Each of the successively projected points are connected by a linear interpolation forming a manifold in the eigenspace as shown in Figure 3. Points in the eigenspace which are closely related correspond to similar images in terms of the $l^2 norm$. It should be noted that a 3D eigenspace is shown for illustrative purposes only and represents one corridor.

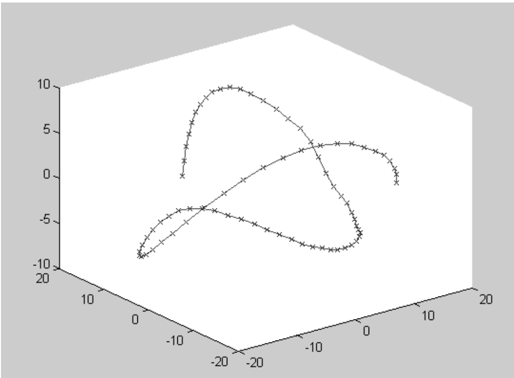


Fig. 3.: The projected points connected together forming a manifold in the M=3 eigenspace.

3.3 Position Estimation

An unknown image representing the robot's current position is projected into the eigenspace forming a point, $\mathbf{c}$. Since it is similar to previously acquired images it will lie close to the manifold but not on it. Which image it is closest to, i.e. the robot's current topological position is calculated as follows:

$$\mathbf{c} = \frac{\sum_{i=1}^{m} (\mathbf{P}_i^2 \lambda_i)}{\sum_{i=1}^{m} \lambda_i}$$

where $\mathbf{P}$ holds the distances, along each dimension in the eigenspace, between the projected point (representing the newly acquired image) and each point (representing the *a priori* data) on the manifold. λ_i is the associated eigenvalue.

4 Ground Plane Dewarped Images

In order to be able to control locally the position and orientation of the robot visual servoing is used. We wish to servo upon the corridor guidelines but since they are warped in the original omni-directional image this task is complex. We take the alternative approach of dewarping the images to a ground plane view *before* servoing. Thus the servoing task is made much simpler as now the world-to-image Jacobian is an exact linear transform. In other words, the 2D projections of 3D straight lines on the floor are themselves straight.

To achieve the required mapping it is necessary to minimize a cost function $\mathcal{F}$ which compares the projection of the current image with known projection data. This known data is obtained from the equations detailed in [4]. Once the mapping has been calculated, all images can then be remapped using a look up table, thus increasing the speed of dewarping.

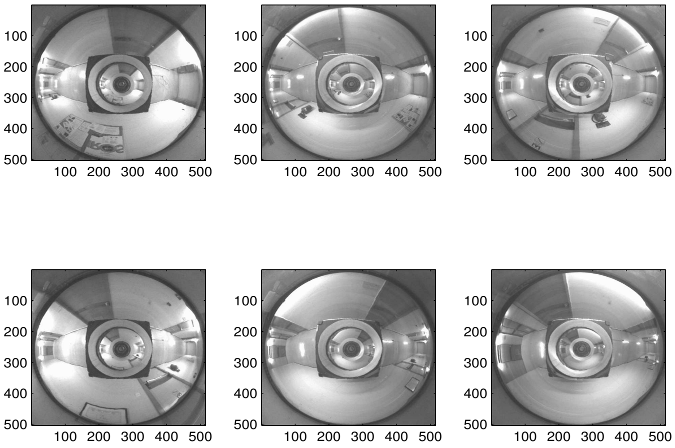


Fig. 4.: A sequence of omni-directional images.

5 Experimental Results

5.1 Experimental Environment

Experiments were undertaken along the corridors of Instituto de Sistemas e Robótica (ISR). This represents a simple indoor office environment. Processing was carried out on-board the Labmate mobile platform by a Pentium II 350MHz PC. The reference positions at which images were acquired were ordered according to the direction of motion, thus maintaining the causality constraint. Images, both warped omni-directional and ground plane dewarped were acquired every 50 cm. A sequence of both types of images is shown in Figures 4 and 5.

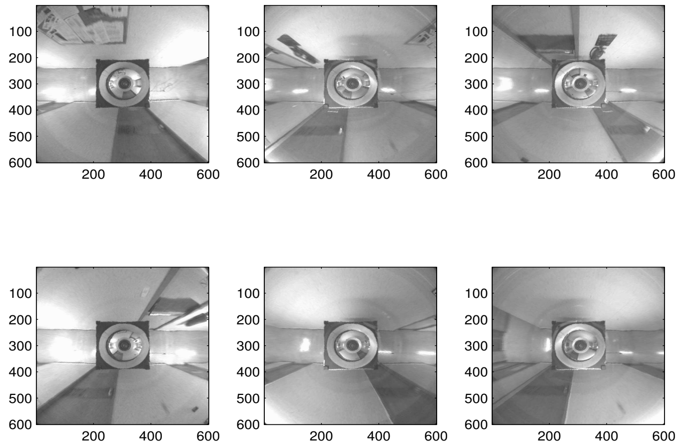


Fig. 5.: A sequence of ground plane dewarped images.

5.2 Eigenspace Results

Figure 6 shows the first 9 eigenimages computed from the image set of 50 omni-directional images, representing one corridor. They are shown in descending order in accordance with their eigenvalues. As an example of

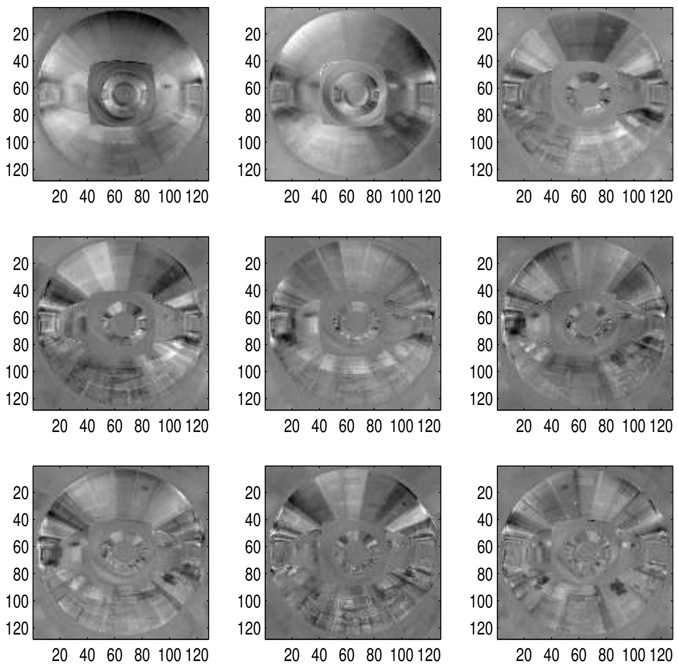
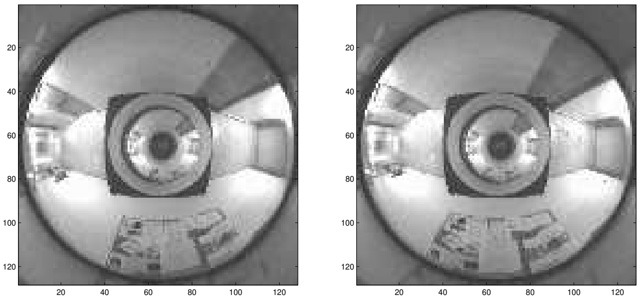


Fig. 6.: The first 9 eigenimages obtained with the omni-directional vision system.

the effectiveness of the appearance based approach, Figure 7(a) shows the unknown image that was projected into the eigenspace and its closest match (Figure 7(b)) obtained from the set of previously acquired images. This retrieved image represents the qualitative topological position of the robot. Ten to twelve basis vectors were found to be enough to correctly determine the nearest image to the newly projected one. With twelve dimensions, misclassification did not occur. Figure 8(a) shows how the first eigenvalues accounted for most of the variance in the images.

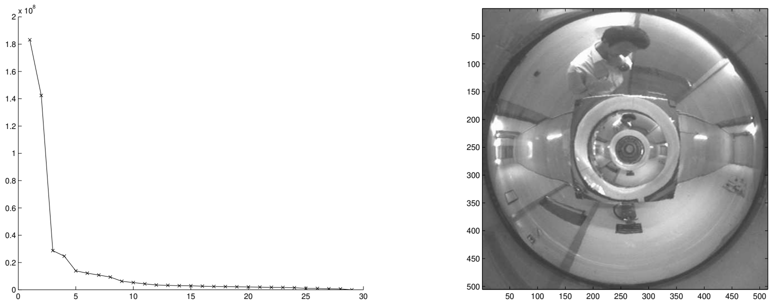
A major advantage of using omni-directional images is that the robot encounters no problems in determining its topological position in a relatively *dynamic* environment. We define a relatively dynamic environment



(a) (b)

Fig. 7.: (a) Unknown and (b) retrieved omni-directional images.

as one where a person can walk and occlude the robot's view, while the robot is moving. When a person is as close as possible to the robot (see Figure 8(b)) they do not create a great enough occlusion so as to cause the robot to misinterprete its topological position. Using PCA the closest image is still correctly determined. An alternative approach is taken by [2]. They used a number of small image windows (9 x 9 pixels, for example) to determine position and deal with occlusion. The success of their method depended on the number and size of windows used.



(a) (b)

Fig. 8.: (a) Eigenvalue drop off. (b) A person close to the robot as seen in the omni-directional image. This did not effect the topological determination of position.

A 3D plot of the images acquired at run time, R versus those acquires a priori, P is shown in Figure 9. It can be seen that the *global* minimum at each time instant is the estimate of the robot's current topological position. If due to some disparity, a local minimum is chosen as the estimate, this incorrect evaluation can be nullified by restricting the search space to images close to a reasonable choice. This approach is valid since images are captured sequentially according to the direction of motion.

5.3 Closed Loop Control Results

Once the position of the robot in the topological map has been determined the next stage involves the calculation of the robot's heading and orientation. This was achieved by dewarping the image to ground plane view (see Section 4) and then tracking the extracted corridor guidelines. Tracking was accomplished by predicting where the central point of the extremities of the bounding boxes would lie in the next image. Differential odometric information was used to increase the speed of tracking. Future improvements to our vision processing algorithm will eliminate the need for this odometric information altogether. Further details of the tracking procedure are presented in [15] and are omitted here due to space considerations. Extracting the corridor guidelines was achieved by using the well known Sobel [6] edge detection method within the predefined bounding boxes. Once points were obtained, they were robustly matched using the RANSAC [3] fitting procedure. An alternative approach is to use first order derivatives to determine the heading direction.

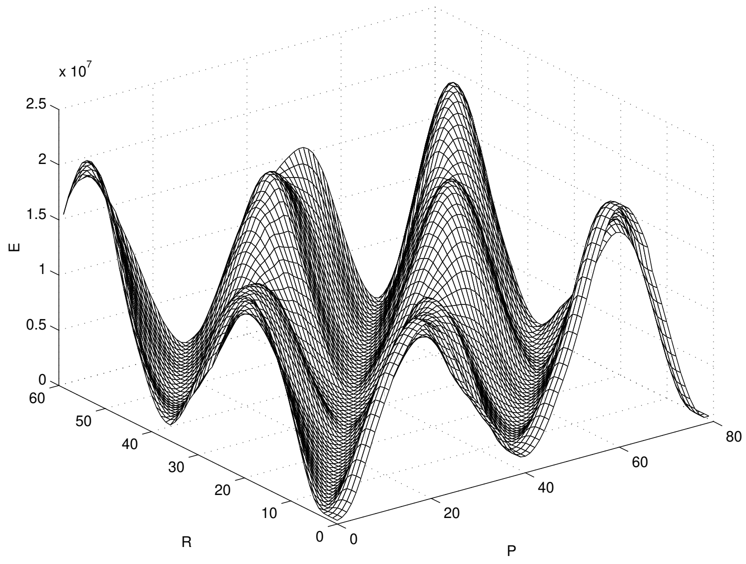
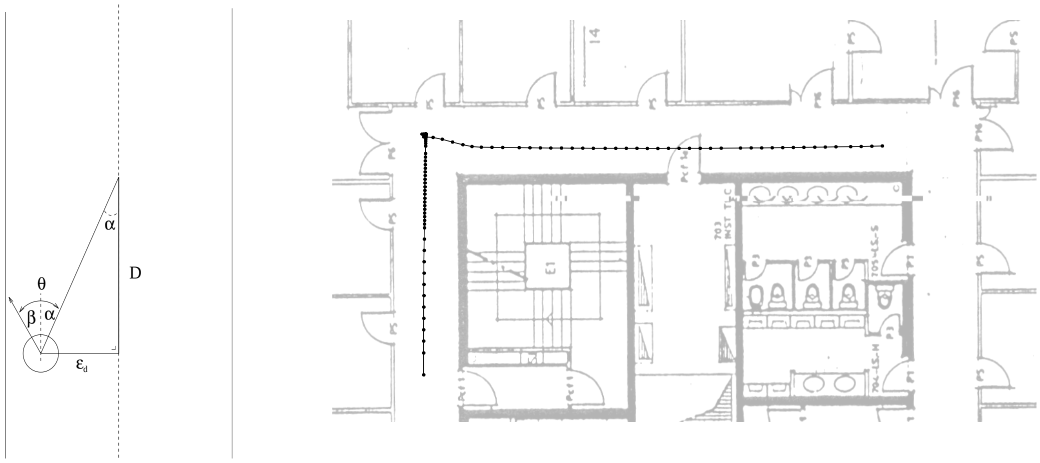


Fig. 9.: A 3D plot of images acquired at run time, R versus those acquired a priori, P. This plot represents the traversal of a single corridor. The global minimum is the estimate of the robot's topological position.



(a) (b)

Fig. 10.: (a) Angles used in the control law. We desire that $\theta = 0$. (b) The path travelled by the robot at our research institute. The total distance travelled was about 21 metres.

In our experience, this method is not robust due to the fact that ground plane dewarped images are relatively noisy.

Once the guidelines were found the robot determined its translation from the centre of the corridor and its orientation. This information was used in a closed loop control scheme which was designed to keep the robot moving in a straight line trajectory down the centre of the corridor. The use of the ground plane dewarped images significantly simplifies this task since the image of the ground plane is just a uniformly scaled version of the ground plane coordinates (no perspective effects). In other words, the world-to-image Jacobian is an exact linear transformation. The control law used is as follows:

$$\theta_{i+1} = \theta_i - \mathcal{K}(\beta_i + \alpha_i)$$

To see how this works refer to Figure 10(a). We desire that θ be zero, in other words that the robot is facing forwards and moving along a straight path. β is the robot's heading. α depends on ϵ_d, the robot's displacement from the centre of the corridor, according to the following equation:

$$\alpha = \arctan\left(\frac{\epsilon_d}{D}\right)$$

This can be seen as a simple kinematic planner to generate reference signals which are followed by the robot. D can be increased or decreased to control the influence of the displacement on the overall control scheme. For our purposes it was set to 6.5% of the image width, while $\mathcal{K}$ was set to 0.45, thus ensuring a smooth trajectory back to the centre of the corridor. It should be noted that control was not captured at the training stage, training images were obtained open loop. Figure 10 represents the results from robot navigation along a section of the 7th floor at ISR. The distance travelled was approximately 21 metres. This traversal was accomplished using the omni-directional vision system as the sole sensor. In order to display the path graphically, odometry was used.

Figure 11 shows a sequence of ground plane dewarped images. The dashes represent the predicted positions of the central point of the bounding box extremities. As can clearly be seen, this is accurate and vastly improves the probability of extracting the corridor guidelines rather than erroneous data such as door frames.

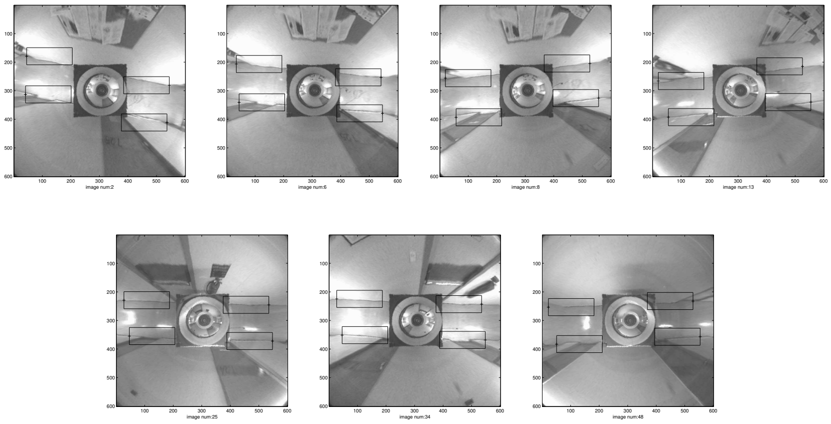


Fig. 11.: Ground plane view of the robot's orientation and translation over time. The dash represents the predicted position of the bounding box extremities.

6 Conclusions & Future Work

In this paper we have demonstrated how to obtain the qualitative position of a mobile robot in a simple indoor environment using a single omni-directional vision sensor. Neither odometric information nor artificial landmarks were required. This was achieved by merging two distinct paradigms, namely appearance based methods and visual servoing. Since we had a relatively large image set, PCA allowed us to build up a low dimensional eigenspace representation and project images obtained at unknown positions into it. The manifold in the eigenspace acted as a topological map of the environment. A method for obtaining ground plane dewarped images was outlined. It was shown that they provide a robust means for the robot to control its motion in a simple indoor environment. Results were presented showing that omni-directional vision provides a suitable means for mobile robot localization. Omni-directional vision allowed for the correct determination of topological position 100% of the time. Our method proved both reliable and robust even when faced with occlusions that a standard CCD camera set up would be unable to handle.

The next stage of this research will concentrate on a number of areas. Improvements to the RANSAC fitting procedure, to allow for a more robust and faster determination of the corridor guidelines, will be investigated. Using the power spectrum density, instead of simple images, as data to build the eigenspace representation of the environment is a near-term goal.

Future work shall be concerned with providing a mechanism whereby the robot can control its motion through doors and around corners. As an initial step to achieving this goal, the topological map representation will have to be extended. Multiple eigenspaces could be connected together to allow for qualitative positioning within a more complex indoor environment. Further investigation of the control scheme to be utilised in a dynamic environment shall be undertaken.

Acknowledgements

This work has been partially funded by the project PRAXIS 2/2.1/TPAR/2074/95, SIVA and the EU TMR network SMART II.

References

1. J. Aloimonos, I.Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1(4):333–356, January 1988.
2. V. Colin de Verdiere and J.L. Crowley. Local appearance space for recognition of navigation landmarks. In R.B. Fisher and G.M. Hayes, editors, *6th International Symposium on Intelligent Robotic Systems (SIRS'98)*, pages 261–269, Edinburgh, Unitied Kingdom, July 1998.
3. M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981.
4. J. Gaspar and J. Santos-Victor. Visual path following with a catadioptric panoramic camera. In *7th International Symposium on Intelligent Robotic Systems (SIRS'99)*, Coimbra, Portugal, July 1999.
5. A. Hicks and R. Bajcsy. Reflective surfaces as computational sensors. In *Workshop on Perception for Mobile Agents at CVPR'99*, Colorado, USA, June 1999.
6. B. Horn. *Robot Vision*. MIT Press, 1986.
7. I. Horswill. Polly: A vision-based artifical agent. In *Proceedings of the AAAI-93*, Washington, DC, USA, 1993.
8. H. Ishiguro and S. Tsuji. Image-based memory of environment. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'96)*, pages 634–639, 1996.
9. S.D. Jones, C. Andresen, and J.L. Crowley. Appearance based processes for visual navigation. In *Processings of the 5th International Symposium on Intelligent Robotic Systems (SIRS'97)*, pages 551–557, July 1997.
10. H. Murakami and V. Kumar. Efficient calculation of primary images from a set of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(5):551–515, 1982.
11. H. Murase and S. K. Nayar. Visual learning and recognition of 3d objects from appearance. *International Journal of Computer Vision*, 14(1):5–24, January 1995.
12. J. Neilsen and G. Sandini. Learning mobile robot navigation: A behaviour based approach. In *IEEE International Conference on Systems, Man and Cybernetics*, San Antonio, Texas, October 1994.
13. M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991.
14. R. F. Vassallo, H.J. Schneebeli, and J. Santos-Victor. Visual navigation: Combining visual servoing and appearance based methods. In R.B. Fisher and G. M. Hayes, editors, *6th International Symposium on Intelligent Robotic Systems (SIRS'98)*, pages 137–146, Edinburgh, Unitied Kingdom, July 1998.
15. N. Winters and J. Santos-Victor. Mobile robot navigation using omni-directional vision. In *Irish Machine Vision and Image Processing Conference (IMVIP'99)*, Dublin, Ireland, September 1999.
16. J. Y. Zheng and S. Tsuji. Panoramic representation for route recognition by a mobile robot. *International Journal of Computer Vision*, 9(1):55–76, January 1992.

This article was processed using the TEX macro package with SIRS99 style