# A Holistic Approach to Mobile Robot Navigation using Omnidirectional Vision

**Niall Winters, B.Sc.(D.Hons)**

A thesis submitted to the

Department of Computer Science,

University of Dublin, Trinity College

in fulfillment of the requirements for the Degree of

**Doctor of Philosophy**

**University of Dublin, Trinity College**

October 2001

# Declarations

The work described in this thesis has not previously been submitted for a degree at this or any other University. Unless otherwise stated, it is entirely the author's own work.

Trinity College Library may lend or copy this work upon request.

Niall Winters

Dated: $26^{th}$ October, 2001

For Mom, Dad and Eimear.

For Becky.

# Acknowledgements

This dissertation represents the culmination of an eight year academic journey from life as an undergraduate to that of a postgraduate researcher. There are many people to whom I owe the deepest sense of gratitude for helping me along the way.

At Trinity College, Dublin, I would like to acknowledge the support of Prof. John Byrne. My thanks to Dr. Gerard Lacey for introducing me to the field of Omnidirectional Vision and supervising the genesis of this research.

In the course of undertaking my Ph.D., I had the opportunity to study at the Instituto de Sistemas e Robótica, Instituto Superior Técnico, Lisbon, Portugal. I would like to extend my sincerest thanks to Prof. José Santos-Victor for his distinguished guidance, help and support. His input, the open discussion of ideas and the tireless giving of his time were very much appreciated. Thanks also to Prof. João Sentieiro.

I would like to acknowledge the support of the EU TMR network, SMART II, Enterprise Ireland, the EU IST project Omniviews (IST-1999-29017) and the Fundação para a Ciência e Tecnologia (FCT).

One's fellow researchers are an important part of everyday life as a postgraduate student. I would like to thank all the members of the Computer Vision and Robotics Group at Trinity College, Dublin, particularly Mark Dennehy, Ulrike Dicke, Nikla Duffy, Ivan Fox, Damian Gordon, Richard Greenanne, Shane MacNamara, Robert McGrath, and Fergal O'Hart.

The members of the Computer Vision Laboratory - VisLab at ISR were an extraor-

dinary eclectic group of people with whom to work. I made many wonderful friends during my stay. My thanks to António Bastos, Alexandre Bernardino, Carlos Carreira, Vitor Costa, Eval Bacca Cortes, Claudia Deccó, Nuno Gracias, Jonas Hornstein, Klaudia Jankowska, Luis Jordão, João Maciel, Javier Minguez, César Silva, Freek Stulp, Raquel Vassallo and Sjoerd van der Zwaan. It was fun sharing an apartment, at various times, with fellow members Etienne Grossmann, Matteo Perrone, Diego Ortin Trasobares, Carlo Favali and Marco Zucchelli. Special thanks are due to my fellow Omnidirectional Vision researcher, José Gaspar. His perspicacious knowledge of the subject proved an inspiration. Thanks also to Sajjad Fekri Asl.

My university education began at the National University of Ireland, Maynooth. Many of the people I met there have remained friends. My thanks to Eamon Gaffney, Mark Johnston, Ger Murphy, David O'Connor and Brian O'Halloran. They reminded me that it is never too late to leave the Ivory Tower! Thanks also to my good friend Peter McNulty.

I would especially like to thank Mom, Dad and Eimear for their unending love, encouragement and support. My gratitude to them knows no bounds.

Finally, heartfelt thanks are due to Becky for all she has done over the past number of years. Her love, kindness and support make life as pleasurable as it is.

**Niall Winters**

*University of Dublin, Trinity College*

*October 2001*

# Abstract

This dissertation presents a novel methodology for vision-based robot navigation. One of the key observations is that navigation systems should be designed through a *holistic* approach, encompassing aspects of sensor design, choice of adequate spatial representations with associated global localisation and local control schemes.

We tackle a number of design issues. Taking inspiration from biology, where wide field-of-views are common, we use an omnidirectional camera. This gives us a 360° horizontal view of the environment.

An *appropriate* environmental representation is a key element for successful navigation. We argue that emphasis should be placed on building the appropriate representation rather than relying upon highly accurate information about the environment. Since our robot is designed to travel long distances, we choose a *topological* environmental representation. The topological map is encoded by a low-dimensional eigenspace obtained via Principal Component Analysis. We detail a local control scheme which allows our robot to effectively use the environmental representation for qualitative localisation.

Finally, we present a method termed, *Information Sampling* which calculates the most discriminating information within the environment traversed by the mobile robot. By developing a method which allows the robot to focus its attention on this data, it is better able to make effective use of its (limited) computational resources. This enables it to more efficiently handle the complexity of the perception process.

# Publications Related to this Ph.D.

[1] **Journal Publications**

(i) Niall Winters and José Santos-Victor, Information Sampling for Vision-based Robot Navigation, In *Journal of Robotics and Autonomous Systems*, to appear.

(ii) José Gaspar, Niall Winters and José Santos-Victor, Vision-based Navigation and Environmental Representations with an Omni-directional Camera, In *IEEE Transactions on Robotics and Automation, Volume 16 Number 6*, pages 890-898, December 2000.

[2] **Conference Publications**

(iii) Niall Winters and José Santos-Victor, Visual Attention-based Robot Navigation using Information Sampling, In *Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Hawaii, USA, October 2001.

(iv) Niall Winters and José Santos-Victor, Information Sampling for Appearance based 3D Object Recognition and Pose Estimation, In *Proceedings of the 2001 Irish Machine Vision and Image Processing Conference*, Maynooth, Ireland, September 2001.

(v) Niall Winters and José Santos-Victor, Information Sampling for Optimal Image Data Selection, In *Proceedings of the 9th International Symposium on Intelligent Robotic Systems*, Toulouse, France, July 2001.

(vi) Niall Winters and José Santos-Victor, Omni-directional Visual Navigation, In *Proceedings of the 7th International Symposium on Intelligent Robotic Systems*, Coimbra, Portugal, July 1999.

(vii) Niall Winters and José Santos-Victor, Mobile Robot Navigation using Omnidirectional Vision, In *Proceedings of the 3rd Irish Machine Vision and Image Processing Conference*, Dublin, Ireland, September 1999.

[3] **International Workshop Publications**

(viii) Niall Winters, José Gaspar, Etienne Grossmann and José Santos-Victor, Experiments in Visual-based Navigation with an Omnidirectional Camera, *Proceedings of the IEEE ICAR 2001 Workshop: Omnidirectional Vision Applied to Robotic Orientation and Nondestructive Testing*, Budapest, Hungary, August 2001. Invited Talk.

(ix) Niall Winters, José Gaspar, Gerard Lacey, and José Santos-Victor, Omnidirectional Vision for Robot Navigation, In *Proceedings of the 1st International IEEE Workshop on Omnidirectional Vision at CVPR 2000*, Hilton Head Island, USA, June 2000.

(x) Niall Winters and Gerard Lacey, Overview of Omni-directional Vision for use with a Teleoperated Mobile Robot, In *Proceedings of the Joint VIRGO-SMART-MobiNet Workshop on Computer Vision and Mobile Robotics*, Santorini, Greece, September 1998.

[4] **Reports**

(xi) Niall Winters, José Gaspar, Alexandre Bernardino and José Santos-Victor, Vision Algorithms for Omniviews Cameras, EU IST Project: Omniviews - Deliverable DI-2, September 2001.

(xii) Claudia Deccó, José Gaspar, Niall Winters and José Santos-Victor, Omniviews Mirror Design and Software Tools, EU IST Project: Omniviews - Deliverable DI-3, September 2001.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

*This chapter introduces the research undertaken for this dissertation, on autonomous, vision-based robot navigation. The motivation for the work is outlined and the proposed method is placed in the context of previous approaches to the problem. An extensive and detailed literature review of the state-of-the-art is provided. The main contributions to the literature are listed and a chapter outline is given.*

This dissertation addresses the problem of autonomous, vision-based mobile robot navigation in structured environments. This topic of research is far from new; a significant amount of work has been done in the past and a vast bulk of literature exists on the variety of works and approaches taken to solve the problem. However, as we will discuss throughout the course of this dissertation, many questions remain unsolved.

We rely upon vision to sense the environment. The reasons for this choice are multifold. First of all, vision provides high resolution information about the environment, which is successfully used by many biological vision systems, to solve a large number of different tasks. Another motivation arises from the challenge of understanding visual perception and testing solutions on artificial systems. How can images, which convey only 2D information, be used in a robust and efficient way to drive the actions of an autonomous system, that operates in a three-dimensional space?

Vision allows us to build a representation of the world which is functional. By this we mean that the goal of vision (in the context of navigation) may not necessarily be that of providing an accurate (metric) reconstruction of the world but rather, an abstraction of it. Such an abstraction is used as the robot's internal model of the environment. This is a key difference between vision and for example, sonar or laser. Using either of the latter sensors, an accurate metric representation (either 2D or 3D) of the world can easily be created[1]. Thus, there is a one-to-one relationship between the structure of the environment and the robot's perception of it. Consequently, navigation algorithms which rely on such sensors often focus on using both the measurements and metric map to constantly localise the robot, rather than using the available resources to actually *drive* the robot towards the desired configuration.

Even though working navigation systems exist, they often rely upon modifications of the environment to facilitate the navigation task. Many fundamental questions, which may have a strong impact on the way we design such systems remain, to a large extent, unanswered. This dissertation addresses some of these questions, in a *holistic* approach towards the design of autonomous navigation systems:

- What **camera/image geometry** is the most adequate for a given navigation problem?
- What **environmental representations** should be used? How can they facilitate global localisation of the vehicle? How can local visual control be applied to the robot, whenever the map is no longer necessary? What are the localisation accuracy requirements and how do we combine global and local navigation?
- What kind of **attention mechanisms** should be used to concentrate the (limited) system resources on the most relevant, in terms of position estimation, sensory input?

[1]Reconstruction can also be achieved using vision. However, such an approach is hardly suitable for real-time navigation, both for complexity and robustness reasons.

Looking at biology again, studies of animal navigation [23, 124] suggest that most species utilise a very parsimonious combination of perceptual, action and representational strategies that lead to much more efficient solutions when compared to those of today's robots. Numerous insects, in spite of having limited sensory and computational resources, manage to solve complex navigation problems in real-time [149]. In various ways, all of these aspects are fundamental to the navigation process. On the one hand, only by answering a number of these questions can we really understand the reasons for the success of biological navigation systems. On the other hand, progress in addressing such fundamental questions may have a dramatic impact on the simplicity, robustness and performance of future autonomous navigation systems.

**Camera Geometry: Omnidirectional Vision** One striking observation in biological vision systems is the diversity of "ocular" geometries. Many animals' eyes point laterally, which may be more suitable for navigation purposes. The majority of insects and arthropods benefit from a wide field-of-view and their eyes have a space-variant resolution. To some extent, the performance of these animals can be explained by their specially adapted eye-geometries.

One possible idea is to design a camera for the specific purpose of autonomous navigation. Extending the field-of-view is a step in this direction. For that purpose, in our work, we use an *omnidirectional camera*.

An omnidirectional camera captures a 360° view horizontally and approximately 110° vertically. In terms of navigation, it offers a number of attractive properties including its wide field-of-view, rich information content, simplicity and increased robustness to occlusion. Two camera designs are presented: (*i*) a conventional camera, pointed upwards, viewing a spherical mirror and (*ii*) a log-polar camera viewing a constant vertical resolution mirror.

**Environmental Representations and Navigation** The complexity of the navigation problem has warranted in-depth and active research over the last three decades. While a large body of worthwhile results were obtained, research often focused upon building metric representations of the world rather than on *tailoring* the representations to the navigation task.

Our robot must travel long distances and so to maintain a precise estimate of position is not only computationally intensive but is not required for successful completion of the task. Instead, in line with the navigation scheme used by both humans and animals [87], we utilise a *topological* estimate of position, where the world is represented qualitatively by a set of images. As we do not require complex systems to capture precise (metric) information, problems of drift and slippage are easily overcome. Furthermore, topological maps deal only with proximity and order and so global errors do not accumulate.

A disadvantage of building topological maps using conventional, narrow field-of-view images is that visually similar places are often indistinguishable. We believe that the increased amount of environmental information provided by an omnidirectional camera alleviates this problem and thus it is particularly suited to capturing topology.

Our solution for determining the qualitative position of the robot (i.e. global localisation) is *appearance-based*. Each reference image is associated with a qualitative robot position. As detailed in Section 3.4.1 (p. 69), localisation is achieved by directly computing the distance, between the current view and the reference images. The closest reference image is the best estimate of position. This matching can be efficiently achieved in a low-dimensional eigenspace [102], obtained via Principal Component Analysis.

Besides using an appropriate environmental representation, there is the need to control the robot's local pose. As an example, when driving, humans make effective use of demarcations along the road for guidance. A key point here is that a minimal

amount of quality information is all that is required to accomplish the task at hand.

Naturally, in indoor environments one can always use simple knowledge about the scene geometry to locally control the robot pose. In our case, we remap the omnidirectional images to *bird's-eye views*[2] of the ground plane and servo upon corridor guidelines. Since we use omnidirectional images, these guidelines are always within the field-of-view.

**Attention Mechanisms: Handling Complexity** A related issue to the navigation problem is that of determining the most informative data within the environment traversed by the mobile robot. Finding this data allows the robot to maximise the use of its limited computation resources. In a way, we are talking about the development of attention mechanisms which, by focusing the system's resources on a subset of the sensory data, allows it to handle the dramatic complexity, intrinsic to most perceptual systems.

Traditionally, "good" information was either provided *a priori*, in the form of artificial landmarks [85, 111], or could only be determined in highly textured environments [130, 131, 166]. More generally, we detail a method, termed *Information Sampling*, for selecting the most discriminating information from an *a priori* set of images. This discriminating information is defined as data which changes significantly from image to image. This method is not restricted to the type of images used and is applied on a pixel-by-pixel basis. Unlike most previous research in this area, the method is non-feature based, and advantageously, can be used with low textured images. We applied the method to robot navigation and object recognition.

[2]That is, scaled orthographic views

## 1.1 Camera Geometry: Omnidirectional Vision

While the optical properties of conic mirrors have been known since the times of Ancient Greece [141] and the idea of the panorama in art became popular in the $18^{th}$ century [8], it was not until 1843 that Joseph Puchberger of Retz, Austria was arguable the first to patent a panoramic camera. Throughout the $19^{th}$, $20^{th}$ and $21^{st}$ centuries a large number of camera designs followed. Today, the spectrum of application has broadened to include such diverse areas as tele-operation [154], video conferencing [114], virtual reality [89], surveillance [133], 3D reconstruction [46, 134], structure from motion [20] and autonomous robot navigation [18, 32, 48, 152, 153, 157, 164, 168]. For a survey of previous work, the reader is directed to [162]. A relevant collection of papers, related to omnidirectional vision, can be found in [36] and [37].

Omnidirectional and panoramic images can be generated by a number of differing systems. These can be classified into four distinct design groupings:

1. Camera-Only Systems
2. Multi-Camera – Multi-Mirror Systems
3. Single Camera – Multi-Mirror Systems
4. Single Camera – Single Mirror Systems

### 1.1.1 Camera-Only Systems

A popular method used to generate omnidirectional images is to rotate a standard CCD camera about its vertical axis, as shown in Figure 1.1. The captured data, i.e. perspective images, are then stitched together so as to obtain panoramic 360° views[3]. High resolution is the primary benefit of this approach, as it does not depend on the camera resolution but on the angular resolution of rotation. As explained in [68], range

[3]Vertical line scans [6] have the advantage of not requiring stitching but do exhibit time constraints.

data can be acquired if the focal point of the camera is kept a given distance away from the axis of rotation. Unfortunately, in terms of robot navigation, the rotating camera

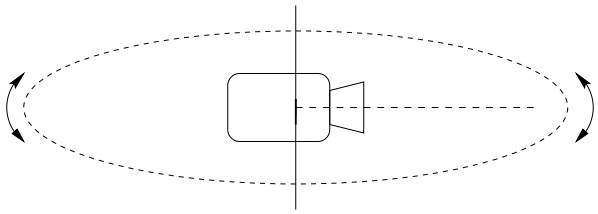


**Fig. 1.1**: Schematic of a rotating camera.

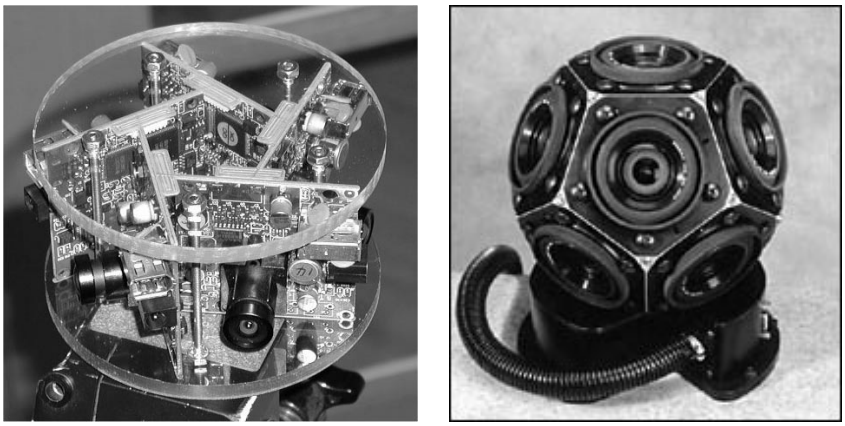
approach has many disadvantages. Its slow capture speed and inability to view actions omnidirectionally and simultaneously make it unsuitable for real-time applications or for navigating through dynamic scenes; a moving object shall be viewed multiple times in *different* positions. In addition, its moving parts mean that it is not the most robust method detailed here.

Instead of relying upon a single rotating camera, a second camera-only design combines cameras pointing in different directions. A number of systems have been built including the *FlyCam* from Xerox Research [43] and the *RingCam* (see Figure 1.2(a)) from Microsoft Research[4]. Here, images are acquired using inexpensive board cameras and are again stitched together to form panoramas. In [43], piecewise perspective warping, of quadrilateral regions, was used to correct high lens distortion and to map images, obtained from each camera, onto a common image plane, before stitching occurred.

Naturally, one can increase the number of cameras used in order to obtain very high resolution images. Such a system, termed the *Dodeca* camera, is commercially available for applications including teleimmersion, simulation and entertainment. Given the nature of the images obtained, specialised viewing equipment is required. Multiple cameras are arranged in a dodecahedron, as shown in Figure 1.2(b) so as to image 360° in the horizontal direction and 290° in the vertical direction (or 91.7% of the

[4]The researcher was contacted for any publications regarding this design. Unfortunately, none were available at the time of writing.

surrounding environment).



(a) (b)

**Fig. 1.2**: Camera-only omnidirectional systems: (a) The RingCam uses board cameras mounted in pentagonal fashion. (b) Very large resolution images are obtained using the Dodeca camera.

One of the oldest methods utilised to capture a wide field-of-view is to equip a camera-only system with a specialised lens. Cao *et al.* [16] describe a system that uses a fish-eye lens [95, 161] for such applications as line following [39] and beacon recognition [15]. Fish-eye lenses, because of their short focal length, can view up to a hemisphere but the acquired images exhibit large radial distortion which requires modelling [98]. Lenses with the advantage of no radial distortion currently cost $\sim \$12,000$.

Finally, Greguss [51] developed a lens, which he termed the Panoramic Annular Lens, to capture a panoramic view of the environment. It consisted of a glass block with two parabolic mirrors and two refracting elements. Unfortunately, its vertical viewing angle was rather limited and so the ground plane near the camera could not be viewed, thus affecting its application to vision-based mobile robot navigation.

### 1.1.2 Multi-Camera – Multi-Mirror Systems

This approach consists of arranging a cluster of cameras in a certain manner along with an equal number of mirrors. Nalwa [103] achieved this by placing four triangular *planar*

mirrors side by side, in the shape of a pyramid, as shown in Figure 1.3(a). A camera was then placed under each mirror and the images obtained from all four cameras were combined to give a 360° panoramic view of the environment. This approach gives high resolution images: 2880 horizontal × 432 vertical lines, with a good depth of field and so has found application in multimedia technologies. A six-camera version has also been designed.

One significant problem which needs to be addressed when using multi-camera – multi-mirror systems is that of geometric registering and intensity blending together the images. Creating *seamless* panoramic views has long been an area of research, see for example [14]. This is a difficult problem to solve given that, even with careful alignment, unwanted visible artefacts are often found at image boundaries. More recently, Majumder *et al.* [89] have addressed this problem when using multi-camera – multi-mirror systems. Achieving real-time results required an SGI 02 with an R10000 CPU. Figure 1.3(b) shows the system composed of 12 cameras with trapezoidal planar mirrors, arranged in a two-tier structure. The field-of-view was 360° in the horizontal direction and 90° in the vertical direction. The primary application for the device was immersive teleconferencing.


(a) (b)

**Fig. 1.3**: Multi-camera – multi-mirror systems from: (a) FullView Inc. and (b) the University of North Carolina at Chapel Hill.

Somewhat the same approach has been applied by Kawanishi *et al.* [73] using six cameras and a hexagonal mirror for the real-time generation of omnidirectional stereo.

Currently, multi-camera – multi-mirror systems are not the most suitable for robot navigation due to their inherit complexity, weight and high power consumption. In addition, if bandwidth is a constraint, other systems offer a better solution. As an historical aside, Nalwa's camera [103] was the first system to provide a panoramic live view of a scene from a single viewpoint.

### 1.1.3 Single Camera – Multi-Mirror Systems

The main goal behind the design of single camera – multi-mirror systems (also known as Folded Catadioptric Cameras [106]) is compactness. A simple example of such a system is that of a planar mirror placed between a light ray travelling from a curved mirror to a camera, thus "folding" the ray. A schematic of single camera – multi-mirror systems are shown in Figure 1.4(a) and (b).

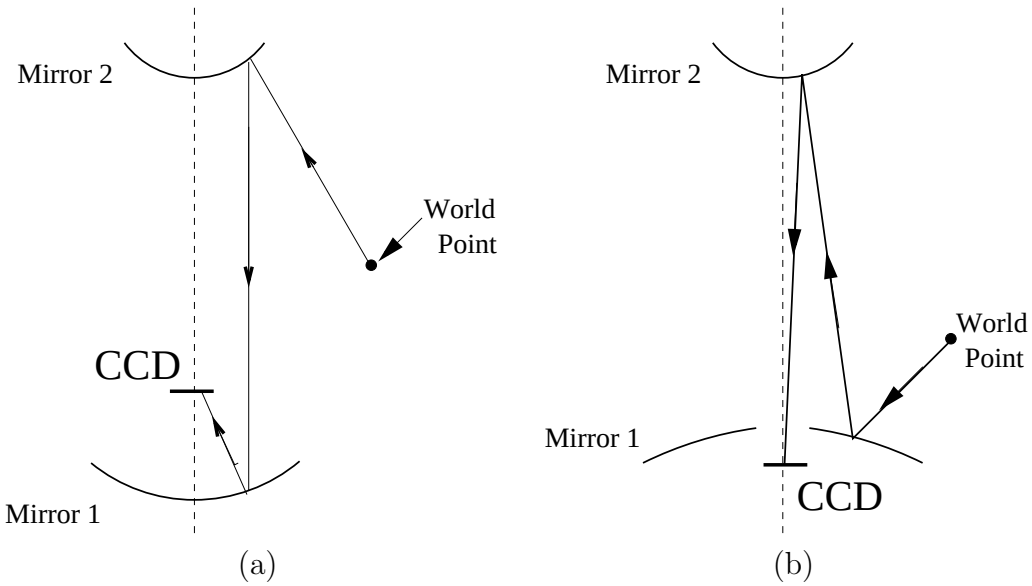



**Fig. 1.4**: Schematic of single camera - multi-mirror systems from (a) Bruckstein and Richardson and (b) Nayar and Peri.

Bruckstein and Richardson [12] presented a design, shown in Figure 1.4(a), that used two parabolic mirrors, one convex and the other concave. Figure 1.4(b) shows the more general design by Nayar and Peri [106], consisting of any two mirrors with a conic profile. They noted that such a system reduces the level of field curvature because the field curvature introduced by one mirror is compensated for by the other. In practice, given that the CCD used is small, the resolution of the final image is greatly reduced.

### 1.1.4 Single Camera – Single Mirror Systems

In recent years, this system design has become very popular and is the approach we chose for application to vision-based robot navigation. The basic method is to point a CCD camera vertically up, towards a mirror, as shown in Figure 1.5. Significantly, in this category there are a number of mirror profiles that can be used to project light rays to the camera.

The first, and by far the most popular design, uses a **standard mirror profile**: planar, conical, elliptical, parabolic, hyperbolic or spherical. All of the former, with the exception of the planar mirror, can image a 360° view of the environment horizontally and, depending on the type of mirror used approximately 70° to 120°, vertically.

The second design involves specifying a **specialised mirror profile** in order to obtain a particular, possibly *task-specific*, view of the environment. In both cases, to image the greatest field-of-view, the camera's optical axis is aligned with that of the mirrors'. A detailed analysis of both the standard and specialised mirror designs are given in Section 2.4 (p. 36). From an historical perspective, the focal properties of mirrors with a conic profile were discovered by the Greek geometer Diocles [141]. The designs for the omnidirectional systems used in this work are described in Section 2.8 (p. 45)

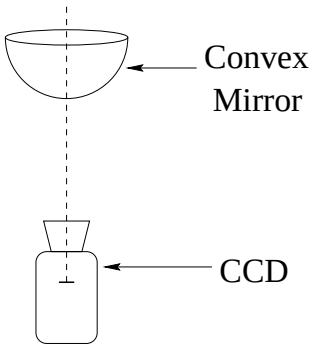



**Fig. 1.5**: Schematic of an omnidirectional system with a standard convex mirror.

## 1.2 Environmental Representations and Navigation

As previously mentioned, the choice of a suitable environmental representation is crucial to the design of a navigation system. For instance, the process of determining 3D distances or structure using vision alone is somewhat complex and often sensitive to noise. Hence, a metric representation may not be the best choice, if vision is the main sensor available. Instead, in our approach we have use a topological representation of the environment, where relevant places are directly represented by omnidirectional images.

Once a suitable environmental representation was chosen, the next (intimately) related question, is that of determining a robust manner of localising the robot with respect to the global representation/map. Here we adopt *appearance-based methods* for *qualitative* localisation.

The last question is how can we locally control the robot in order to complete the mission? In our case we visually servo upon corridor guidelines, which are always within the field-of-view of our omnidirectional camera.

We shall now provide a literature review of the state-of-the-art. A number of related works are relevant to our methodology. It should be emphasised that most are concerned with only some of the issues addressed in this dissertation. For example, works on topological mapping may use a number of sensors to determine their position.

Alternatively, works on appearance-based matching may not have considered real-world navigation. They simply define it as a matching problem, *without* recourse to a local control strategy or appropriate sensor usage.

### 1.2.1 Topological Navigation

Over the past few years the use of topological maps for navigation has grown in popularity. Notably, some of the research detailed below was inspired by theories about how humans navigate through their environment (as detailed in Section 3.2 (p. 59)).

Perhaps the first attempts to use topological maps with a physical mobile robot were implemented, separately, by Chatila and Laumond [21] and Crowley [26] in 1985. Chatila and Laumond's thesis was that three world models were necessary: a geometric model, a topological model and a semantic model. The key point to note in their work was that the topological level was built *after* acquiring geometric data. Thus, they had to overcome the difficulties of geometric modelling when constructing and maintaining an accurate world model. Crowley developed a navigation system equipped with a rotating ultrasonic range sensor, where a network of pre-learned places allowed for global path planning.

The first research work which attempted to go directly to the topological level, using a *physical* mobile robot, was undertaken by Sarachik [123]. While no topological map was actually constructed, she did implement a "room finder". This was achieved by determining the dimensions of the room in which the mobile robot was located by using two vertically placed cameras. The long-term goal was to combine the room finding module with a door finding module and so gain the ability to navigate along a topological map.

Mataric developed a robot named "Toto" [91] to act as a testbed for integrating a topological map architecture into a subsumption-based [11] mobile robot. Landmarks, detected using sonar and compass readings, served a dual purpose: to encode topo-

logical structure and to act as communicators. When a detected landmark matched a node on the topological map, the robot was localised to the position of that node. This information was subsequently communicated to all other nodes (landmarks) on the map for path planning.

Kosaka and Pan [78] implemented a system which used a neural network to process the incoming images from a conventional camera and then applied fuzzy logic to deal with the uncertainty in the inferences drawn from the visual data. They termed their architecture *FUZZY-NAV* and used a topological model of corridors for navigation. Hallway navigation was possible by transforming images to Hough space and using the output from a neural network to keep the robot centred in the corridor. Landmark detection, based on recognising door frames from different perspectives, was partially implemented. It is not clear how the system would cope with the disappearance of corridor guidelines from the narrow field-of-view images.

Košecká [79] presented a successful navigation system, where the mobile robot was modelled as a point in a 2D configuration space. The environment was represented by a place graph and travelling from one node (i.e. landmark) to another was achieved by visual servoing. A pan-and-tilt camera was used for tasks including wall following and door servoing. It was assumed that landmarks and their feature co-ordinates in the global system were known *a priori*.

As noted in Section 3.3.2 (p. 64), topological maps often have problems differentiating between similar locations, especially when using sonar sensors which are limited in range and angular resolution. Kortenkamp and Weymouth [77] addressed this problem by characterising *distinctive places* from sonar and conventional camera data, combined using a Bayesian network. Nodes in the environment were termed gateways and were classified as such by using sonar sensors. It was only after classification that visual information was added to differentiate between gateways. The visual information consisted of simple vertical edges, and successful recognition of a visual cue relied

upon its location, direction, distance and length. The visual cues were stored in an abstract scene representation (ASR) and eight ASRs were needed at each location in order to account for orientation change. Scale changes were not considered. Chapter 5 describes our approach to the problem of finding discriminating information within the environment.

Franz *et al.* [44] implemented "view graph" based navigation. A view graph is simply defined as a collection of views that describe a relevant path. Inspired by insect navigation, their system travelled between nodes on the graph using a *homing* strategy. Homing is the ability to find views which are connected to a start view. They did not use views of every spatial position along a route but instead chose to take omnidirectional *snapshots*[5] of the environment at particular time intervals determined by the distance between neighbouring views. Spatial closeness was defined by the degree of similarity between views: problems of occlusion were not addressed. Importantly, snapshots did not represent distinctive places in the environment and were not labelled. Moreover, experiments were for the most part carried out in simulation. A small experiment was implemented on a mobile robot platform, where the robot was able to home over a distance of about 1 metre but real-world experimental evaluation was left for future work.

In [110], Owen and Nehmzow presented a landmark-based mobile robot navigation system. A topological map was constructed by self-organising sonar and compass data using a Restricted Coulomb Energy classifier. Essentially, only landmarks which were sensed by the robot, over a distance of 10cm, were added to the map, thus attempting to alleviate the problem of missing landmarks as navigation proceeded. Since perception was affected by the orientation of the robot, the authors chose to mount the sonar sensor on a turret and use the on-board compass to ensure that it faced north at all

[5]Cartwright and Collett [17] developed the snapshot theory of navigation in order to explain how bees locate sources of food. As the bee flies, it takes snapshots of the environment, so that, on subsequent visits to the same area, it can easily relocate these important sources.

times. Unfortunately, this made the system invariant to rotation.

Vassallo *et al.* [122, 144] implemented a topological navigation system. In terms of the overall approach to the navigation problem, this research is closely related to ours but the imaging geometry, matching scheme and servoing strategy were different from those presented in this work.

Ulrich and Nourbakhsh [143] presented a system quite similar to the one detailed here, although published three years after this work began. They too utilised an omnidirectional camera but instead of using eigenspaces as topological maps (see Section 3.4 (p. 66)), their approach used colour histograms and nearest neighbour learning. It was not implemented on a real robot and no local control was available. Additionally, since histograms are invariant to rotation, they cannot be used for such tasks as turning at a corner, thus limiting the applicability of this approach.

### 1.2.2 Appearance-based Methods for Global Localisation

Perhaps the first use of appearance-based methods for navigation was developed by Hong *et al.* [60, 61] in 1990. Although no large-scale navigation was undertaken, the idea presented was that large-scale navigation tasks could, in principal, be divided into a sequence of small-scale tasks. These would then be solved by local image-based homing. As implemented, homing was defined as the robot's ability to find its way to a known, local, target location. It was achieved by extracting a 1-D circular "location signature" (actually, the horizon line[6]) from each omnidirectional image and subsequently extracting features from these signatures. Matching was done using a normalised cross-correlation function. In all, 17 images of a corridor were acquired, 40cm apart. The robot then homed from one image to the next. Given the small distance traversed (6.8m), it is not clear how this approach would scale up to larger

[6]The horizon line has the property that landmarks projected onto it remain there as the robot moves.

environments. Additionally, relying upon a 1-D circular ring from the omnidirectional image, is not very robust.

A similar idea was developed by Hancock and Judd [54] in 1993. They named their mobile robot "Ratbot" and a silver Christmas tree bulb decoration was used as part of a very simple omnidirectional vision sensor. The image data acquired were 1-D panoramic image strips extracted at a given height above the ground. While they were not of high enough quality to uniquely determine position, they proved useful in correcting dead reckoning measurements. Localisation was achieved by matching features, simply vertical bars, in the run-time images to those in the database. No detailed experimental results were presented. In later work [86], a version of this system was placed on a car, again for dead reckoning correction.

Horswill [62] developed a low-cost robot, "Polly", capable of vision-based navigation. Efficiency was achieved by task specialisation and places in the world were ordered using qualitative co-ordinates. Position was estimated by landmark recognition. This was the system's weakest link, given the fact that the robot's standard camera was pointed towards the floor and so landmarks were restricted to corridor intersections. The system failed in the presence of occlusion but was a major step forward towards the goal of building low-cost robots.

Zheng [168] presented a system which moved along a given route under human guidance and autonomously memorised a side-view of that route. To obtain a wide field-of-view, two types of images were constructed: panoramic views and generalised panoramic views. The first was a projection of a scene onto a cylinder, taken by a stationary rotating camera through a vertical slit. The second was acquired along a path by a laterally facing camera, again through a vertical slit. These data were then used as a basis for route recognition so as the robot could autonomously locate and orient itself. Image matching was performed in a coarse-to-fine manner using a dynamic programming technique and successful results were achieved. This approach suffered

from the complexity of using large images, in addition to the necessity of acquiring two representations of the environment.

As a first attempt at efficient memory utilisation, Ishiguro [67] used the Fourier Transform of a set of omnidirectional images as an image-based memory of the environment. The images were acquired in a section of an unknown office environment. By using the similarity between images, they were organised so as to reflect the environmental geometry. An interesting result of this research was that, even though the arrangement of images was significantly distorted, the topology remained unchanged. Matching was achieved using the sum of absolute differences. Unfortunately, results of autonomous navigation were not presented.

A "View Sequence" of images for navigation was proposed by Matsumoto *et al.* [93] and images were captured using a standard narrow field-of-view camera. Localisation was achieved by template matching (in hardware), using the central rectangular section of each image as a template. Unfortunately, this led to a qualitative local control scheme, where the robot's displacement was determined by image shift: rotation and lateral translation could not be distinguished. Additionally, when large changes in appearance occurred, the robot was unable to localise itself. Clearly, this system would benefit from the use of an omnidirectional camera. Matsumoto *et al.* realised this and in [94], they presented a system with a hyperbolic mirror. The images acquired were transformed by cylindrical projection. Images were acquired every 0.5m - 1m apart or according to changes in the scene and thus did not describe important places within the environment. Localisation was achieved by using the entire cylindrical image as a template, a time consuming task. In addition, to keep the robot moving along a straight path, small front and rear template matching was used. If these regions were occluded the robot became laterally displaced.

In [92], Matsumoto *et al.* again extended this system by using a stereo head for free space detection and optical flow for junction detection. Unfortunately, this complicated

the system, given that two vision systems were required for effective navigation.

Continuing along this line, Maeda *et al.* [88] used the parametric eigenspace approach to image matching [102]. They noted that if one takes a single image from a *conventional* camera, multiple matches could be obtained in environments where similar images appeared a distance apart. Thus, robot position could not be reliably determined. The proposed solution to this problem was to take another image, close to the current one, by moving the robot (or the camera) and projecting this image into the eigenspace. Position estimation was considered successful if the difference between the estimated pose and the actual pose was less than 1m and 20°, a large error estimation band. Experiments were only undertaken in straight lines, no corner detection was evoked. The disadvantage of this approach was that, when building the eigenspace, multiple images had to be acquired at every location, thus increasing complexity. Indeed, this shows that building an appearance-based system with a standard camera leads to problems. An additional disadvantage was that the robot's ability to traverse its environment was reduced, since it was required to stop and start again in order to overcome ambiguities.

The approach taken to vision-based navigation by Aihara *et al.* [1] utilised *row* autocorrelated omnidirectional images and eigenspace matching. Autocorrelating the omnidirectional images makes them invariant to rotation. This is a nice property when undertaking such tasks as corridor following but fails if the rotational information is required, for example, when turning corners. Additionally, if their images contained any local deviations, matching failed. Since there was no local control strategy, they were forced to densely sample the environment and build multiple small eigenspaces (of 18 images each). Thus, a two-stage approach to localisation was used: first the nearest eigenspace had to be detected, followed by the closest image within that space. The mechanism for detecting the closest eigenspace was not detailed. Using this method, images, **from** the *a priori* set, were recognised in 100% of cases using more than 6

eigenvectors. More realistically, when localising using images **not from** the *a priori* set, the recognition rate dropped to approximately 65%, when using 6 eigenvectors and 85%, when using 18 eigenvectors. A possible reason for the low recognition result is that row autocorrelation is not a one-to-one mapping, and so different images can match to a single database image.

Pajdla and Hlaváč [112] attempted to overcome this one-to-many mapping problem by using what they termed a "Zero Phase Representation". Here, the phase of the first frequency of the Fourier Transform was set to zero, thus gaining a one-to-one mapping in the presence of no image deviation. Jogan and Leonardis [70] also tackled this problem by using a representation termed, spinning images. In their case, a single image was acquired at each location and subsequently shifted row-wise by 7.2° in order to simulate possible rotations. A disadvantage of this approach was that the number of images required to represent the environment increased 50-fold, although the dimension of the eigenspace did not exhibit such a profound increase.

Yagi *et al.* [163] presented a route recognition system for a mobile robot using a 2D Fourier power spectrum of polar panoramic images. Presuming that the density of environmental features changed, the frequency component of the power spectrum varied. Thus, it was used to differentiate between places in the environment. Matching was achieved using cross-correlation and the robot motion was assumed to be constant and linear.

Andersen *et al.* [2] presented an appearance based approach which defined visual processes for navigation. Here simple processes were used to transform images into commands for displacement and steering. Their system used odometry in association with a conventional camera (or a multiple camera configuration) on a distributed system. Image matching was done using zero mean energy normalised cross correlation. Such tasks as navigating along a rectangular 9m × 3m path were successfully achieved.

A method for robot navigation using image sequences was proposed by Rasmussen

and Hager [115]. Stable features were visually servoed upon and used to guide the robot. In order to be successfully applied, good features had to be in the field-of-view. This was not always the case, for example at turns and throughout long corridor stretches, thus impacting on the results obtained.

Kato *et al.* [72] detailed an environmental representation termed a T-Net. Here an omnidirectional camera was used to capture images of the environment and matching was achieved using templates. Sub-goals were defined by targets.

## 1.3 Attention Mechanisms: Handling Complexity

While navigating, the computational load on a robot can be reduced if it has the ability to identify, and focus its attention, upon *highly discriminating* regions within the environment. This is a real need when considering the complexity of the input imagery. In this way, the robot can concentrate its (limited) system resources on the most relevant sensory input. This input is then periodically memorized for future reference, thus mimicking the approach to navigation naturally adopted by humans and some animal species. One may classify such input as a "landmark". Our approach is to define discriminating regions as the pixels (within a set of images) which vary *significantly* from one image to the next, i.e. those which exhibit the most information change. Our review of related research shall only concentrate upon work which looks at the problem of selecting informative features (or landmarks) from image data.

A good starting point for selecting effective points is to use an interest operator, which defines an interest point as a location where the signal changes two-dimensionally, at corners, for example. In the literature, one can find many operators: Schmid *et al.* [128] carried out a comparison of different detectors and concluded that the OurHarris detector, an improvement over the original Harris [55] detector, provided the best results. Each of the points selected by the detector were characterised by their dif-

ferential structure. In [127], Schmid and Mohr applied this method to image retrieval from a large database.

Knapek *et al.* [75] address the problem of selecting, from a single image, landmarks which are both salient (i.e. "pop-out" from the background) and distinctive. Their work is based upon, and strongly influenced by, that of Schmid and Mohr [127, 128], although their application was mobile robot navigation. Here potential landmarks (points) were selected by first applying an interest operator. These potential landmarks were then characterised by a feature vector of partial $k^{th}$-order derivatives (known as a k-jet). Subsequently, the potential landmarks were ordered by distinctiveness, with the most distinctive being retained, thus forming landmarks for the robot. When navigating through the environment these landmarks were recognised by nearest neighbour classification using the Mahalanobis distance. Experiments were undertaken using three image sequences: a 2.5m linear trajectory, a 2.5m circular trajectory and rotation about the optical axis. It was shown that the most distinctive landmarks are more easily recognised from one image to the next. The advantage of this approach is that selected landmarks can be recognised under large changes of scale and orientation. Its major downfall is that, in order to achieve good results, highly textured environments are required. In addition, if viewpoint dependent features (T-junctions, for example) are selected by the interest operator, then as the robot moves, errors occur. This is due to a change in the interest point characterisation by the k-jet.

Schiele and Crowley [126] used Multidimensional Receptive Field Histograms [125] to build a network of salient points describing an object. In this case, the salient points were those which were most unique and maximised the distinctiveness *between* objects.

Yeh and Kriegman [166] considered the problem of automatically selecting, from a set of 3D features, the set (landmark) which was most likely to be recognised in a single image. The approach worked as follows: a subset of features were considered as a candidate landmark, after which a recognition function was used to find this landmark

and its associated Bayesian cost. The one with the lowest cost was selected as the most discriminating. For their experiments, the set of candidate landmarks were restricted to 12 vertical line features within an image. From these 12, 4 were selected by the recognition function as the optimal landmark. The goal was to recognise this landmark in 24 images acquired around a 90°circular arc at three depths. The selection of the optimal landmark assumed that all features were visible and that no other vertical lines were considered as additional features. This was achieved by manually selecting the 12 vertical lines from each of the 24 test images thus giving $12!(12-4)! = 11,880$ groups of possible landmarks. The recognition function was then applied to the 11,880 groups of features and those falling within the recognition interval were considered matches to the optimal landmark: it was found in 23 out of the 24 test images. Naturally, in order to achieve high landmark recognition rates, this approach relies upon highly textured environments.

Sim and Dudek [131] presented a method for vision-based robot localisation. As part of this work, they proposed an approach to selecting appropriate landmarks which were then used to encode images of the environment. Candidate landmarks were selected as subwindows of high edge density which exceeded a user-defined threshold. A low-dimensional eigenspace representation of these candidate landmarks was built and, in order to recognise the landmarks from different viewpoints, they were tracked over the configuration space. Position estimation was achieved by matching the landmarks, extracted from a test image, to those tracked landmarks in the database. In one experiment, database images of a simple structured scene were captured at 2cm intervals on a 30cm $\times$ 30cm grid. Then 100 test images were taken at random positions. The average deviation in position using a landmark set was 3.8mm. In a second experiment the grid size was increased to 1.2m $\times$ 3m and database images were captured at 20cm intervals. The average localisation error was found to be 6.8cm. It is not clear how the method would scale up to larger environments. Again, this method relied upon the

availability of good texture within the environment.

In [137], Thrun described a statistical technique which allowed a robot to automatically learn landmarks. Here the robot was presented with a set of sensor readings, labelled with the position at which they were acquired. This data was used to train a neural network to minimise the expected localisation error after taking a sensor reading. After training, the robot could recognise the landmarks which best estimated its position.

In the area of tracking, Shi and Tomasi [130] proposed a method for selecting easily trackable features which corresponded to real-world physical points. This selection was based on the monitoring of the quality of the image features. Each feature's RMS residue between the first frame and the current frame, assuming affine motion, was measured and when this value (termed dissimilarity) grew too large, the feature was abandoned. The main goal of their work was to discover *local* problems which may occur during tracking. In an experiment, features were selected from a 26 frame sequence of a structured scene acquired with a forward moving camera. Using affine motion dissimilarity, discriminating between good and bad features was possible.

A philosophically similar idea to Information Sampling (see Chapter 5) was implemented by Dellaert and Collins [33] in the area of real-time tracking. They proposed a method, termed Selective Pixel Integration, to select the most informative pixels from an image. In their case, they relied upon the fact that the change between any two images could be described by a 2D projective transform (or 2D homography [56]). This warping is governed by eight parameters. The main idea of Selective Pixel Integration was to find the pixels, in the original image, which provided the most information about the *change* in the parameters of the homography.

Our approach was influenced by that of Rendas and Perrone [117]. They addressed the problem of current mapping in coastal areas using *a priori* knowledge of the survey area.

In the area of object recognition, several authors have noted problems when using entire standard images to build a low-dimensional eigenspace, including sensitivity to occlusion, scale change and illumination. The problem of dealing with partial occlusion (in a bin-picking task) was investigated by Ohba and Ikeuchi [108]. Instead of projecting the entire image, as is usual, they proposed dividing each image into a number of smaller windows which they termed *eigenwindows*. Eigenspace analysis was then applied to each window. Their basic idea was that even if a number of the windows were occluded, the remaining ones would contain enough information to perform the bin-picking task. As they pointed out, a very large number of image windows need to be stored in order to obtain good results. For example, if one had an *a priori* set of 1000 images of size $256 \times 256$ pixels, and each window was $8 \times 8$ pixels in size, then one would require 1,024 non-overlapping windows to represent an image or 1,024,000 to represent the entire *a priori* set. Clearly the chances of one window, acquired at runtime being matched to a number of images from the *a priori* set is high. This could be due, for example, to having many ambiguous regions within an image. As noted by Colin de Verdière and Crowley [28, 29] this leads to the problem of deciding which eigenwindows contain discriminative information and therefore should be used in the recognition task. It is highly desirable that only the most *effective* windows are selected from each acquired image, and that only these chosen windows be matched to the *a priori* set.

As a solution to this problem, Ohba and Ikeuchi proposed using three criteria to eliminate the redundant windows, namely: detectability, uniqueness and reliability. Colin de Verdière and Crowley reformulated the problem as a question of whether to use the set of eigenwindows selected by a particular interest operator or to use those windows selected from a predefined grid. When using a predefined grid, the first task was to project *all* of the eigenwindows into the eigenspace. Since an image contained a number of windows, it was represented in the eigenspace as a surface and a set of images were represented by a set of surfaces. Naturally, on projection of an eigenwindow

many matches occurred. Thus, suppression of redundant windows was required. This was usually achieved by noting that a search for the closest point in the eigenspace produced too many matches. Alternatively, it was noted that suppression could occur at the training stage given that a redundant window will be projected many times. Nevertheless, this approach still required enough space and computational power to store and search for all of the eigenwindows.

## 1.4 Original Contributions

The original contributions of this dissertation are as follows:

1. We propose a holistic methodology for vision-based navigation, validating two different omnidirectional camera designs.

2. We show that by combining (*i*) a suitable camera geometry, (*ii*) an appropriate environmental representation, (*iii*) an adequate localisation scheme and (*iv*) a means of local pose control, successful navigation is possible.

3. We demonstrate how our methodology forms a core part of a larger navigation module, which allows a mobile robot to undertake both precise local tasks (docking, for example) and qualitative global navigation.

4. We detail a statistical method, termed Information Sampling, for finding the most discriminating pixels from an *a priori* set of images. A key point of this approach is that it is non-feature based and so can be applied to images exhibiting low texture.

5. Information Sampling is defined as an extension to a topological representation of the environment. In addition, it is shown to be beneficial for both navigation and object recognition.

## 1.5 Dissertation Structure

This dissertation is structured as follows:

**Chapter 2** introduces the reader to the field of *Omnidirectional Vision*. We detail the many camera designs available, in particular those based on a mirror-camera combination. The designs used in this work, namely a spherical mirror combined with a standard camera and a specialised mirror combined with a log-polar camera are presented in detail.

**Chapter 3** is concerned with defining an appropriate, easily implemented, environmental representation for a mobile robot. This is its internal model of the world. In common with what is known about how humans represent large-scale space, we describe how a topological representation can be used for navigation. For completeness, other environmental representations are discussed. The topological representation is encoded by a low dimensional eigenspace, obtained via Principal Component Analysis. Initial matching results proved to be successful.

**Chapter 4** details our approach to vision-based navigation using an omnidirectional camera. Building upon the information presented in Chapter 3, we show that, in order to achieve successful navigation, a synergistic combination of topological navigation and local pose control is required. Local control is based on the tracking of guidelines, extracted from *bird's-eye* view images. Navigation results obtained using the two omnidirectional camera designs detailed in Chapter 2 are presented. Additionally, we provide results from integrated experiments which rely on a path distance/accuracy trade-off in order to robustly solve the navigation task. During the many trials of our navigation methodology, the distance travelled by the mobile robot varied from 17m to 35m.

**Chapter 5** proposes an extension to our topological environmental representation. We define Information Sampling as a statistical method for selecting only the most discriminating data from the *a priori* image set. In this way the robot maximises use of its computational resources and can effectively handle the complexity of the perceptual process. Real-world results show that navigation is possible using only the discriminating information. Encouraging preliminary results from navigation experiments using very low resolution images, for example 16 pixels in size, are also detailed. In an extension of the method, successful object recognition results are presented.

**Chapter 6** presents our conclusions and directions for possible future research.

# Chapter 2

# Omnidirectional Vision: Systems, Principals & Camera Design

*This chapter presents the state-of-the-art in catadioptric sensor design. Our motivation for using omnidirectional vision is provided. The camera designs used in this work are presented in detail and the differences between them and other sensors are highlighted. The single centre of projection is discussed. The method used to remap omnidirectional images to scaled orthographic views of the ground plane is also described.*

## 2.1 Introduction

Visual systems are designed to collect, with the utmost effectiveness, real-world data. These data are subsequently used to derive important information for such everyday tasks as: navigation, person-following and recognition. Since many biological and artificial systems are endowed with limited computation means, they must utilise their resources to the best of their ability.

In nature, one can find a large variety of viewing geometries, or simply *eyes*, where each is "designed" to efficiently process visual information for particular tasks. For

example, flying insects such as bees have compound eyes with a very large field-of-view which is beneficial for 3D motion estimation. Primates have corneal eyes; they view the world through high resolution colour imagery. Evidently, these images facilitate such tasks as identification, tracking and surveillance.

Biologically speaking, successful eye designs are those which help solve particular tasks quickly and robustly, rather than producing highly accurate images of the environment [83]. Indeed, eye geometries have evolved tremendously over time [27].

Until the mid-1990's the use of standard narrow field-of-view cameras was ubiquitous. In recent years, computer vision researchers have gained broad exposure to wide field-of-view imaging systems. Within this community they are customarily referred to as **omnidirectional**, **panoramic** or **catadioptric** systems. These definitions exhibit nuances which are inconsistent with their current technical usage. **Omnidirectional**, in this case, refers, not to a view captured in every direction, but to one captured in all horizontal directions, although limited in vertical viewing angle. The degree of limitation depends on the particular system. Typically, vertical viewing angles range from $\sim 70^\circ$ to $\sim 120^\circ$. An omnidirectional image, captured by a system mounted on a mobile robot, is shown in Figure 2.1. Here the robot is located in the centre of the image. Omnidirectional images can be remapped to panoramic views, as shown in Figure 2.2.

**Panoramic** systems are defined as systems which are constrained to capturing panoramas with up to a $360^\circ$ horizontal field-of-view. They do not possess the capability of capturing omnidirectional images, such as those shown in Figure 2.1. Typically, high resolution panoramas can be imaged by mosaicking with a rotating camera.

The word **catadioptric**[1] refers to the set-up of the actual system. The word is derived from a combination of terminology from optics: catoptrics relates to the optics of mirrors (reflection) and dioptrics to the optics of lens (refraction). In the context

[1]The term *catadioptrics* was first coined by Hecht and Zajac [57] in 1974.

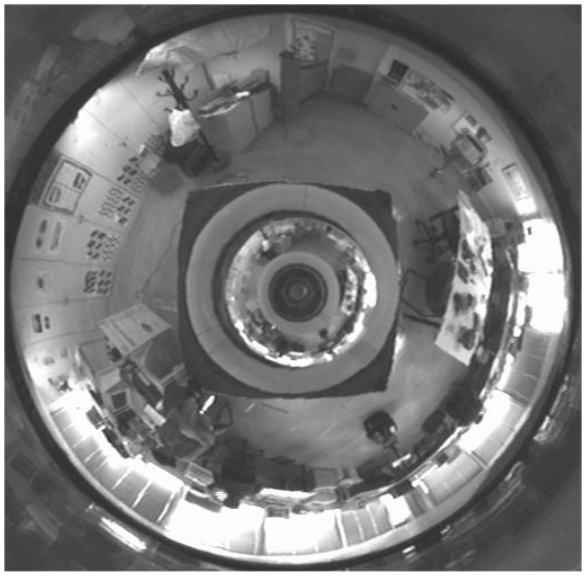


**Fig. 2.1**: An omnidirectional image.

**Fig. 2.2**: A panoramic image obtained by remapping Figure 2.1.

of computer vision, the term catadioptric was first used by Nayar [104] in 1997. As detailed in Section 1.1 (p. 6), there are other methods for capturing omnidirectional images.

Throughout this dissertation, we shall refer to the conventional phrase *omnidirectional vision* when signifying research related to wide field-of-view imaging systems. When detailing an actual system built to capture omnidirectional images[2], we shall use the term *catadioptric sensor*.

## 2.2 Omnidirectional Vision: Motivation

The main motivation behind the use of omnidirectional vision is that of obtaining a wide field-of-view. How this in itself is beneficial depends on the specific application. Certainly, in the area of motion estimation, for example, the goal of capturing the correspondences between images is significantly aided by using omnidirectional vision. This is simply due to the fact that, even though points may be occluded, they remain in the image rather than disappearing from view altogether, as is the case when conventional cameras are used.

The areas of internet streaming and remote reality benefit from the use of omnidirectional images primarily because they allow for the transmission of a large amount of visual information within a *single frame*. Tele-operation [154] is another area where this is beneficial: an operator can view the remote scene over a low-bandwidth link in a more natural manner than if he had to wait for a number of standard images of the environment to be transmitted.

It is highly significant that omnidirectional images, captured at a particular point, contain enough information to make them distinguishable from other images captured nearby. This is not the case with images obtained from a conventional camera, where

[2]The only exception to this definition are camera-only systems, as presented in Section 1.1.1 (p. 6), since by design they are not catadioptric.

multiple matches can easily occur. A closely related issue is that, implicitly, algorithms relying upon omnidirectional visual input can overcome occlusions with significantly more ease than those relying upon conventional imagery.

An advantage of catadioptric sensors over standard pan-and-tilt units lie in their simplicity: they have no moving parts and therefore exhibit increased reliability. When mounted on a mobile robot, and provided that the camera does not move, the orientation of the sensor is related to that of the robot by a rigid transformation. This imaging geometry has a number of properties that can be exploited in various navigation or recognition tasks. For example, vertical lines in the environment are viewed as radial image lines.

Overall then, as the visual competence of a mobile robot is substantially increased by using omnidirectional vision, this imaging modality clearly shows its superiority over conventional methods. Many works of art illustrate wide field-of-view imagery. Perhaps the most famous of these is *Hand with Reflecting Globe* by M.C. Escher, as shown in Figure 2.3.

**Fig. 2.3**: Hand with Reflecting Globe by M.C. Escher.

## 2.3 The Single Centre of Projection

The *single centre of projection*, otherwise known as a *single effective viewpoint*, is a defining theoretical characteristic of an omnidirectional vision system. The system is said to exhibit such a property if all light rays, captured by a particular design, meet at a single point. For example, if one were using a catadioptric sensor, then all rays would be reflected from a mirror (or set of mirrors), as if they emanated from a single point, located behind the mirror(s).

The single centre of projection is graphically illustrated in Figure 2.4. Here the sphere represents a truly omnidirectional view, i.e. $360° \times 360°$. The single centre of projection is denoted by **S**. If one imagines standing at this point and then looking in a particular direction, this is equivalent to looking at the world from a perspective point of view. If one looks further in the horizontal direction, the image viewed is a panorama.

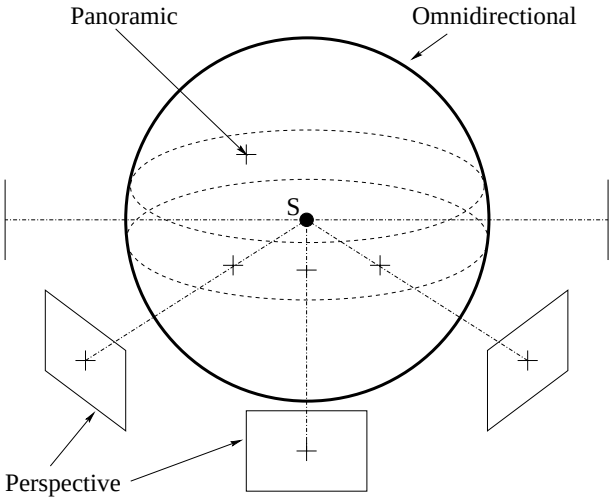



**Fig. 2.4**: Schematic of the Single Centre of Projection, **S**.

In more precise terms, a panoramic view is generated by a cylindrical projection and distortion-free perspective views by a planar projection. Thus, single centre of projection systems can be considered equivalent to purely rotating cameras.

| **Omnidirectional System** | **SCP?** | **Comments** |
|---|---|---|
| *Camera-Only Systems* | | |
| Rotating Camera | Yes | Focal point on axis of rotation |
| Combined Cameras | No | Different focal points |
| Fish-eye Lens | Yes | Exhibits radial distortion |
| Panoramic Annular Lens | Yes | Limited vertical viewing angle |
| *Multi-Camera – Multi-Mirror Systems* | Yes | Depends on mirrors used |
| *Single Camera – Multi-Mirror Systems* | Yes | Depends on mirrors used |
| *Single Camera – Single Mirror Systems* | | |
| Planar Mirror | Yes | Non-practical |
| Elliptical Mirror | Yes | Non-practical |
| Parabolic Mirror | Yes | Requires orthographic lens |
| Spherical Mirror | No | Loci of projection centres |
| Conical Mirror | No | Loci of projection centres |
| Hyperbolic Mirror | Yes | Difficult to calibrate |

**Table 2.1**: A summary of omnidirectional vision systems and whether or not they have a single centre of projection (SCP).

As cited by [49], a similar idea (related to concave mirrors) was in the mind of Zenodorous when he asked Diocles "to find a mirror surface such that when it is placed facing the sun the rays reflected from it meet at a point" [141].

We note here that most camera-only systems do obey the single centre of projection constraint. The exception is multiple camera-only systems as they view the scene from different, depth-dependent, directions. When using planar mirrors, multi-camera – multi-mirror systems do have a common single centre of projection. It is located on the axis of the mirror structure, where each of the virtual effective pinholes meet. In the case of both single camera – multi-mirror systems and single camera – single mirror systems, whether or not they obey the constraint depends upon the mirror and/or lens used. A summary is given in Table 2.1.

Recently, the need for systems with a single centre of projection, for real-world application, has come under scrutiny. This point is further analysed in Section 2.7 (p. 43).

## 2.4 Mirror Profiles for Single Camera – Single Mirror Systems

The catadioptric sensor designs used in this work were single camera – single mirror systems. We now detail both standard and specialised mirror profiles for use with this class of sensor.

### 2.4.1 Standard Profiles

Baker and Nayar [3, 4] define the complete class of mirrors satisfying the single centre of projection constraint. The only practical solutions are as follows: (*i*) a parabolic mirror with an orthographic lens and (*ii*) a hyperbolic mirror with a standard lens. Solutions using planar, conical, elliptical and spherical mirrors only obey the constraint in non-practical cases. For example, a planar mirror obeys the constraint but does not increase the field-of-view. The strict requirement of a single centre of projection for omnidirectional imaging is discussed in Section 2.7 (p. 43). We note here that the scope of application can be increased by concentrating on the mirror profile design rather than on the importance of the constraint.

We shall now discuss the merits and drawbacks of each mirror when designing a single camera – single mirror catadioptric sensor. We begin with the practical solutions.

Parabolic Mirror: When using a parabolic mirror, light rays are reflected from the surface of the mirror in parallel *and* perpendicularly to the image plane of the camera, i.e. an orthographic projection. This is graphically illustrated in Figure 2.5. In order to satisfy the single centre of projection constraint, a parabolic mirror must be used in conjunction with an *orthographic* lens.

Calibrating this system is relatively simple, as the optical axis of the mirror, and

that of the lens, do not have to be in exact vertical alignment; translations can be tolerated. Additionally, the focal point of the camera can be at any distance from the mirror. A catadioptric sensor with a parabolic mirror has been used for such applications as video-conferencing [114]. Multiple regions of interest were extracted simultaneously and broadcast as if each image was obtained with a conventional camera.

For vision-based robot navigation, this design is not the most suitable. Orthographic lenses are both large and heavy, making for a cumbersome catadioptric sensor. Additionally, if the camera to mirror distance shrinks, the lens induces a significant amount of self-occlusion.

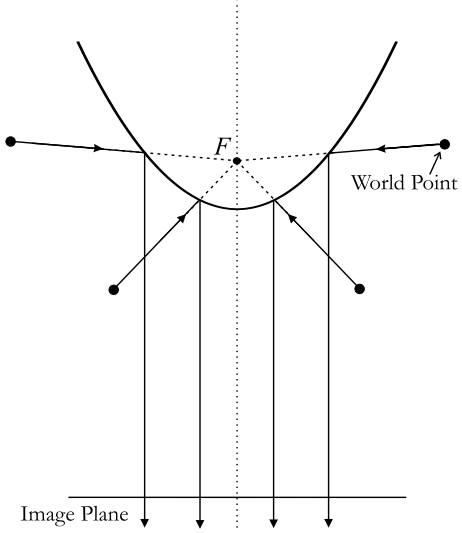



**Fig. 2.5**: Schematic of a Parabolic Mirror.

Hyperbolic Mirror: Catadioptric sensors built using hyperbolic mirrors have proven to be a practical solution to the generation of omnidirectional images [116, 136, 165]. A hyperbolic mirror satisfies the single centre of projection constraint *only* when the focal point of a conventional camera (with a standard lens) is precisely positioned at one of the foci of the hyperbolic mirror. This is shown in Figure 2.6. Given the high

cost of manufacturing hyperbolic mirrors, one wants to be confident that a single centre of projection is obtained. Unfortunately, this is extremely difficult; tolerance to the manual vertical movement/alignment of the camera, or mirror, is almost negligible. Therefore, invalidating the single centre of projection constraint is highly possible.

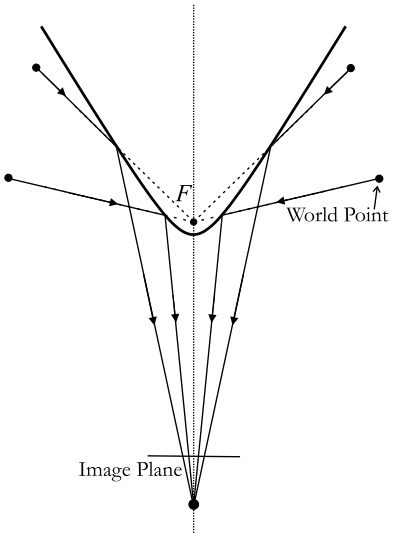



**Fig. 2.6**: Schematic of a Hyperbolic Mirror.

We now move on to discuss the non-practical solutions to the single centre of projection constraint. At this point, we reiterate the biological fact that, throughout time, successful eye designs did not necessarily produce optically perfect images of the environment.

Planar Mirror: A planar mirror satisfies the single centre of projection constraint but, as the camera view is not enhanced, it is a non-practical solution. We note here that if multiple mirrors are used, each has its own single centre of projection, and so to maintain the constraint, a camera must be associated with each mirror.

Ellipsoidal Mirror: An ellipsoidal mirror satisfies the single centre of projection

constraint if the pinhole of the camera is located at one of the foci of the ellipsoid. While this solution increases the field-of-view, the increase is minimal and so, in practice, this mirror is not used.

Conical Mirror: The conical mirror is a popular choice of mirror for catadioptric sensor design [18, 32, 164]. It only satisfies the single centre of projection constraint if the apex of the cone is placed at the pinhole of the camera. This is obviously impractical since objects within the environment cannot be viewed. By moving the mirror the environment becomes visible but the single centre of projection is lost.

Spherical Mirror: A spherical mirror satisfies the single viewpoint constraint if the pinhole of the camera lies at the centre of the sphere. This is a non-practical solution to the constraint. If the mirror is placed a distance from the camera we obtain a loci of projection centres. Images obtained from a spherical mirror have a higher resolution in the centre of the image but are distorted at the periphery. They do have the distinct advantage of yielding the widest field-of-view of all the sensors using convex mirrors. As with conical mirrors, real world vertical lines appear as radial lines originating from the image centre. Spherical mirrors have been used for robot navigation [48, 61, 153, 156, 157], (see Section 4.5 (p. 92)). A schematic is shown in Figure 2.7.

### 2.4.2 Specialised Profiles

As an alternative to using a standard mirror, one can design a profile suited to particular tasks, for example, imaging a scene at a constant vertical resolution. This idea of utilising specialised mirror profiles has become more common in the literature over the past few years. Hicks and Bajcsy [58] designed and constructed a mirror which directly imaged a bird's-eye view of the environment, without the need for remapping software. In [59] they went on to present two families of surfaces which provided a wide

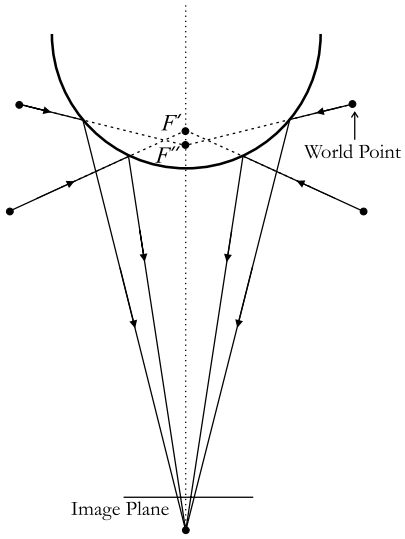



**Fig. 2.7**: Schematic of a Spherical Mirror.

field-of-view, while approximating a perspective projection. Chahl and Srinivasan [19] produced a mirror which ensured that a change in the elevation angle was proportionally mapped to a change in radial distance from the centre of the image. In [24], Conroy and Moore built on this work to produce a stereo resolution invariant mirror. Gachter *et al.* [45] had a similar design goal to Chahl and Srinivasan when they designed a mirror which, when used with a *log-polar sensor*, produced a uniform cylindrical projection. Thus, at a given distance from the camera, an object was the same size in the image, independent of its height in the real-world. This design was again improved upon by Deccó *et al.* [31]. A number of different mirror designs were produced, including a constant horizontal resolution mirror, for use with both standard and log-polar cameras, and a so-called mixed mirror design. Here, a single mirror profile was designed so that the outer part of the sensor imaged a scene with constant vertical resolution, while the inner part produced a constant horizontal resolution image.

More details on a constant vertical resolution mirror, combined with a log polar sensor, and its application to vision-based robot navigation, can be found in Sections 2.8.2 (p. 51) and 4.5 (p. 92), respectively.

## 2.5 A Unifying Theory for Single Centre of Projection Systems

Recently, Geyer and Daniilidis [50, 49] presented a unifying theory for all catadioptric systems *with* a single centre of projection. They showed that these systems (parabolic, hyperbolic, elliptical and perspective[3]) can be modelled by a two-step mapping, $\mathcal{M}_s$ via the sphere. This mapping of a point in space to the image plane is graphically illustrated in Figure 2.8. The two steps of the mapping are as follows:

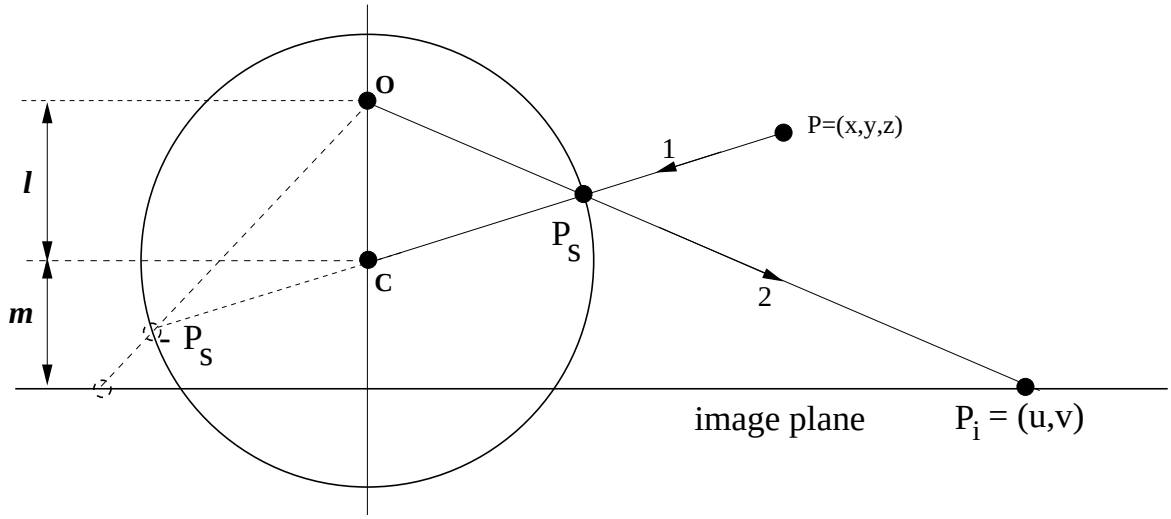



**Fig. 2.8**: A Unifying Theory for all catadioptric sensors *with* a single centre of projection.

1. Project a 3D world point, $\mathbf{P} = (x, y, z)$ to a point $\mathbf{P_s}$ on the sphere surface, such that the projection is normal to the sphere surface.

2. Subsequently, project *to* a point on the image plane, $\mathbf{P_i} = (u, v)$ from a point, $\mathbf{O}$ on the vertical axis of the sphere, through the point $\mathbf{P_s}$.

The mapping, $\mathcal{M}_s$ is mathematically defined by Equation 2.1:

[3]A parabolic mirror with an orthographic lens and all of the others with a standard lens. In the case of a perspective camera, the mirror is virtual and planar.

$$\left.\begin{bmatrix} u \\ v \end{bmatrix} = \frac{l+m}{l \cdot r - z} \begin{bmatrix} x \\ y \end{bmatrix}, \text{where } r = \sqrt{x^2+y^2+z^2} \right\} \mathcal{M}_s \tag{2.1}$$

As one can clearly see, this is a two-parameter, ($l$ and $m$) representation, where $l$ represents the distance from the sphere centre, **C** to the projection centre, **O** and $m$ the distance from **O** to the image plane. Modelling the various catadioptric sensors with a single centre of projection is then just a matter of varying the values of $l$ and $m$ in Equation 2.1. As an example, to model a parabolic mirror, we set $l = 1$ and $m = 0$. Then the image plane passes through the sphere centre, **C** and **O** is located at the north pole of the sphere. In this case, the second projection is the well known stereographic projection. We note here that the Unifying Theory can model standard perspective cameras (i.e. the pinhole model) when $l = 0$ and $m = 1$. In this case, **O** converges to **C** and the image plane is located at the south pole of the sphere.

In terms of camera self-calibration, it was shown that the image centre, the effective focal length (and the mirror eccentricity, when using a hyperbolic mirror) of *non-perspective* catadioptric sensors can be calculated from lines in a single image without the need for metric information[4]. It was assumed that that aspect ratio was one and the skew zero. In the parabolic case, three lines were required, while in the hyperbolic case, self-calibration was possible with just two.

## 2.6 Which Design to Use?

Omnidirectional images obtained with single camera catadioptric designs are, comparatively speaking, of medium resolution. This is because, when compared to a conventional camera, each pixel images a larger portion of the environment and so the ability to differentiate between environmental details is lessened. If one's application

[4]In the perspective case, when using lines and no metric information, the number of unknowns always exceeds the number of constraints and so calibration is not possible.

necessitates high resolution image capture (for example, surveillance or personal identification) then either multiple camera – multiple mirror or camera-only designs offer a better solution.

However, if one's objective is to use a catadioptric design that: (*i*) views a dynamic scene omnidirectionally, (*ii*) is the least likely to fail, (*iii*) has the lowest power consumption and, (*iv*) obtains adequate image resolution, then the best solution is given by the single camera – single mirror catadioptric design. Our motivation for using *particular* single camera – single mirror catadioptric designs for vision-based robot navigation is outlined in Section 2.8.1 (p. 46).

## 2.7 The Single Centre of Projection Revisited

As outlined in Section 2.3 (p. 34), the single centre of projection is an important theoretical property of a catadioptric sensor. When research into the design and use of catadioptric sensors first became widespread, the single centre of projection was viewed as the *primary* design parameter. Very recently, for many applications, this idea has come under scrutiny.

It is our belief that catadioptric sensors should be designed to facilitate the particular task at hand. The application of omnidirectional vision in this work is visual-based robot navigation and so we designed a catadioptric system with this objective in mind. Naturally, there are applications where the need for a single centre of projection is unquestionable: omnidirectional stereo or 3D reconstruction, for example. Nevertheless, in each case *application drives sensor design*, not the quest for optical perfection.

As related by Baker and Nayar in [3, 4], the only designs which adhere to the single centre of projection constraint are rotations of conic sections. If this constraint is relaxed, the number of possible designs, and thus applications, increases. For example, one can design a sensor with a constant vertical field-of-view or one which gives an

orthographic view of the ground plane.

A secondary question related to the use of non-single centre of projection sensors is the quality of the image obtainable from such systems. What is the degree of error induced by a loci of viewpoints? Gaspar and Santos-Victor [46] have studied this problem using a catadioptric sensor with a spherical mirror. As outlined in Section 2.5 (p. 41), the Unifying Theory covers all catadioptric sensors with a single centre of projection. A projection model governing a catadioptric sensor with a spherical mirror, termed the **Spherical Projection Model** is given in Section 2.8.1 (p. 47). If the Unifying Theory can approximate a non-single centre of projection camera, one would expect that - using both models - the error between projecting 3D points to the image plane would be small. It turns out that for real-world points further than 2m away from the catadioptric sensor the error in the image plane is less than *1 pixel*.

Derrien and Konolige [34] also approximated a single centre of projection but used a concept they termed *iso-angle mapping*. They constructed a virtual system by displacing all incoming rays, each having a unique Euler angle, so as they converged at a single point. Thus, their method produced a camera with a single centre of projection, imaging a distorted scene. Since they did not derive an analytical expression for the distortion, it was measured as a change in the height of a small object, given a change in its elevation angle and remained less than 2.5%.

In [59], Hicks and Bajcsy described a mirror design which approximated a perspective projection, i.e. a single centre of projection system. Their mirror design directly produced an orthographic view of the ground plane. For such a mirror to approximate a single centre of projection, an orthographic view of planes - other than the ground plane - should also be produced. Indeed, this was the case, as shown by Figure 2.9 (from [59]). Here a checkered pattern was placed on the floor and another on a table top. As can clearly be seen, both are nicely mapped, and although the exact error was not detailed, the result indicates a qualitative approximation to a single centre of

projection.

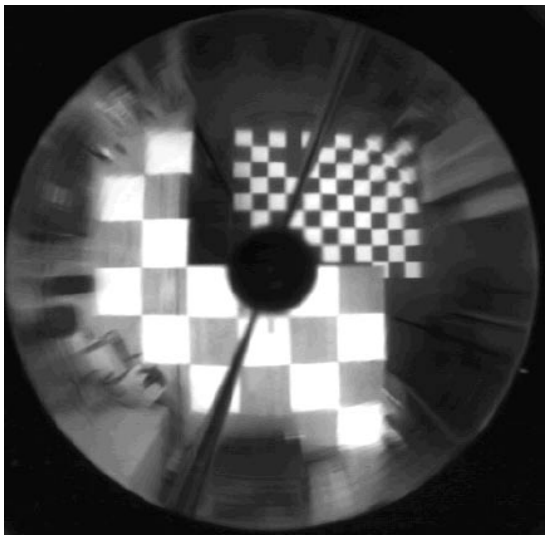


**Fig. 2.9**: Hicks and Bajcsy designed a mirror which approximates a perspective projection. In this case, two orthographic views of the ground plane are correctly mapped from the *same* mirror (from [59]).

## 2.8 Catadioptric Sensor Designs

For this work, two catadioptric sensor designs were used. Both were of the *single camera – single mirror* class. In the following sections, we detail the design of each sensor.

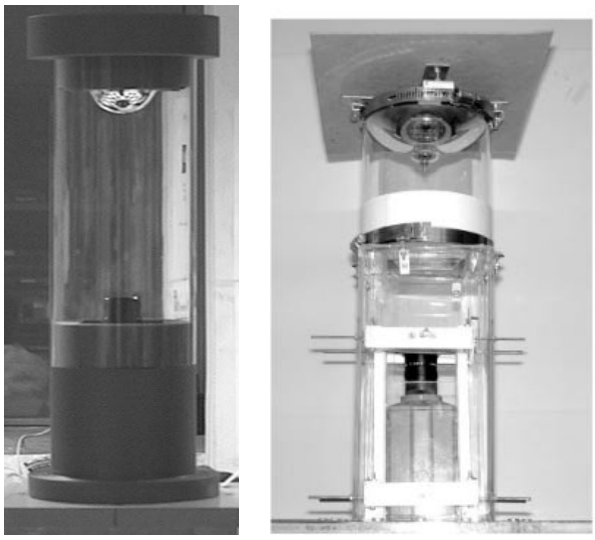
The first (older) design used a spherical mirror with a conventional camera [48, 153, 157]. Two systems were built and each was mounted on a Labmate mobile robot platform[5]. Details of the design are given in Section 2.8.1 (p. 46).

The second design utilised a specialised mirror with a *log-polar* camera [31, 135]. A number of systems were built, with one being mounted on a Scout mobile robot base. More details on this design can be found in Section 2.8.2 (p. 51).

[5]Identical set-ups were implemented at the Computer Vision Lab, Instituto de Sistemas e Robótica, and at the Computer Vision and Robotics Group, Department of Computer Science, University of Dublin, Trinity College.

### 2.8.1 Design of a *Single Camera – Single Mirror* Catadioptric Sensor with a Spherical Mirror

The first catadioptric sensor designed utilised a *spherical* mirror. Primarily, this was due to the fact that having a non-single centre of projection system is not a drawback for vision-based robot navigation. Additionally, this work did not require the generation of geometrically correct perspective views. The omnidirectional view given by the spherical mirror was ample for our needs. A secondary issue was the relative ease of calibration of a spherical system compared to that of a hyperbolic one. Finally, the inexpensive cost and easy availability of spherical mirrors [25] proved an effective lure. Spherical mirrors have previously found application in areas of autonomous navigation [48, 61], and tele-operation [5, 154].



(a) (b)

**Fig. 2.10**: Two of the omnidirectional cameras built: (a) The camera at TCD and (b) the camera at IST. Both use a spherical mirror.

We term our projection model the **Spherical Projection Model**. Essentially, we model the projection of a 3D point to the image plane, reflected via a point on a spherical mirror. The three key design parameters to be kept in mind are: (*i*) the

mirror radius, (*ii*) the camera-to-mirror distance and (*iii*) the vertical viewing angle. For example, we can specify a certain viewing angle and determine the mirror radius and camera-to-mirror distance required to achieve that viewing angle. Naturally, we must also keep in mind that the entire mirror should fill as much of the CCD as possible, thus giving good resolution. Additionally, the camera's optical axis and the axis of the mirror should be aligned.

The camera design presented below was developed at the Instituto Superior Técnico, Lisbon, Portugal. Preliminary work on catadioptric sensors with spherical mirrors was also done at the University of Dublin, Trinity College.

### The Spherical Projection Model

The geometry of image formation is obtained by relating the co-ordinates of a 3D point, $\mathbf{P}$, to the co-ordinates of its projection on the mirror surface, $\mathbf{P_m}$, and finally to its image projection $p$. Figure 2.11 shows the most relevant parameters of the geometric model for image formation.

A point $\mathbf{P_m} = (r_m, z_m)$ on the mirror surface has to fulfill the following equations:

$$\begin{aligned} r_m &= (z_m + L)\ \tan\beta \\ R^2 &= z_m^2 + r_m^2 \\ \gamma_r &= \gamma_i \ \Leftrightarrow -2\arctan\left(r_m/z_m\right) = \alpha - \beta \end{aligned} \tag{2.2}$$

where $\beta$ is the radial angle, $R$ is the mirror radius, $L$ is the distance to the camera projection centre and $\alpha$ is the elevation angle defining the size of the camera's vertical view. Naturally, the angle of incidence, $\gamma_i$ equals the angle of reflection, $\gamma_r$.

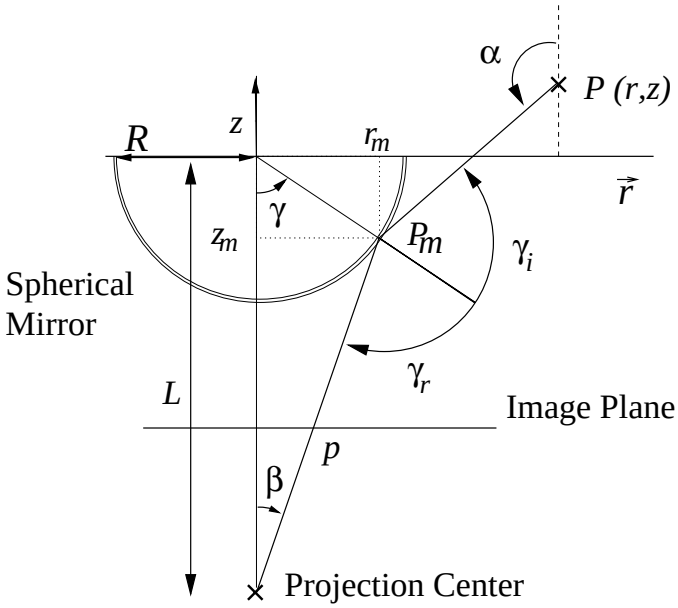



**Fig. 2.11**: Camera (spherical mirror) projection geometry. Symmetry about the z-axis simplifies the geometry.

These equations are reduced to a vertical plane containing the vertical $z$ axis, since the system is rotationally symmetric around that axis. The last equation can be expressed as a function of the vertical viewing angle, $\alpha$, or expressed using the co-ordinates of a 3D point, $\mathbf{P} = (r, z)$. Some of the parameters involved in Equations (2.2) are fixed by the physical setup $(R, L)$, whereas $(\alpha, \beta)$ depend on the co-ordinates of an observed 3D point.

### Projection of a 3D Point

The projection of a 3D point, $P = [x\ y\ z]^T$, is a two-step process: $(i)$ project the 3D point onto the mirror, $(ii)$ project this mirror point, $P_m$ onto the image plane. $P = [x\ y\ z]^T$ can be expressed in cylindrical co-ordinates as:

$$P = \begin{bmatrix} \varphi & r & z \end{bmatrix}^T = \begin{bmatrix} \arctan(y/x) & \sqrt{x^2+y^2} & z \end{bmatrix}^T$$

Noting that the vertical viewing angle, $\alpha$ for $P$ can be expressed as:

$$\alpha_P = \arctan\left(\frac{z - z_m}{r - r_m}\right) + \frac{\pi}{2}$$

we can replace $\alpha$ in Equations (2.2) and solve the resulting non-linear system of equations to determine $(r_m, z_m)$. This allows us to determine the value of $\beta$.

All that remains to be done is to project the mirror point, $P_m = [\varphi\ r_m\ z_m]^T$ onto the image plane, $p = (u, v)$. Using the perspective projection model and taking into account the camera's intrinsic parameters, we get:

$$\begin{bmatrix} u^* \\ v^* \end{bmatrix} = \tan\beta \begin{bmatrix} \cos\varphi \\ \\ \sin\varphi \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \end{bmatrix} \begin{bmatrix} u^* \\ v^* \\ 1 \end{bmatrix}$$

where $f_u, f_v$ denote the focal length expressed in (vertical and horizontal) pixels; and $u_0, v_0$ is the position of the principal point on the image co-ordinate system[6].

### Model Parameter Estimation

Points in 3D space, $P$, are projected as image points, $p$, by means of a projection operator, $\mathcal{P}$: $p = \mathcal{P}(P, \theta)$, where $\theta$ contains all the intrinsic and extrinsic parameters of the catadioptric panoramic camera: $\theta = [L\ f\ u_o\ v_o]^T$.

While the mirror radius can be measured easily, the camera-mirror distance, $L$, focal length, $f$ and principal point, $(u_0, v_0)$, can only be determined up to some error: $\delta\theta = [\delta L\ \delta f\ \delta u_0\ \delta v_0]^T$.

Thus, to estimate $\delta\theta$ we use a set of known 3D points, $P^i$, and the corresponding image projections $p^i$, then minimise the following cost function:

$$\delta\theta = \arg\min_{\delta\theta} \sum_i \parallel p^i - \mathcal{P}(P^i, \theta_0 + \delta\theta) \parallel^2 \tag{2.3}$$

At this point we have:

[6]We assume that the pixel aspect ratio is known and use $f = f_u = f_v$ as the focal length expressed in pixels.

1. defined the projection operator needed to obtain omnidirectional images with a conventional camera/spherical mirror set-up.

2. described a procedure to estimate the model parameters, starting from initial nominal settings.

For successful navigation, a mobile robot needs a module to estimate and control its pose (i.e. position and orientation) as it travels through an environment. It would be ideal if there was an easy method of obtaining this information visually. *Bird's-eye views* of the ground plane offer such a solution. Here, omnidirectional images are remapped to scaled orthographic views of the ground plane, thus greatly facilitating the measurement of distances and angles directly from the image.

### Obtaining a Bird's-Eye View of the Ground Plane

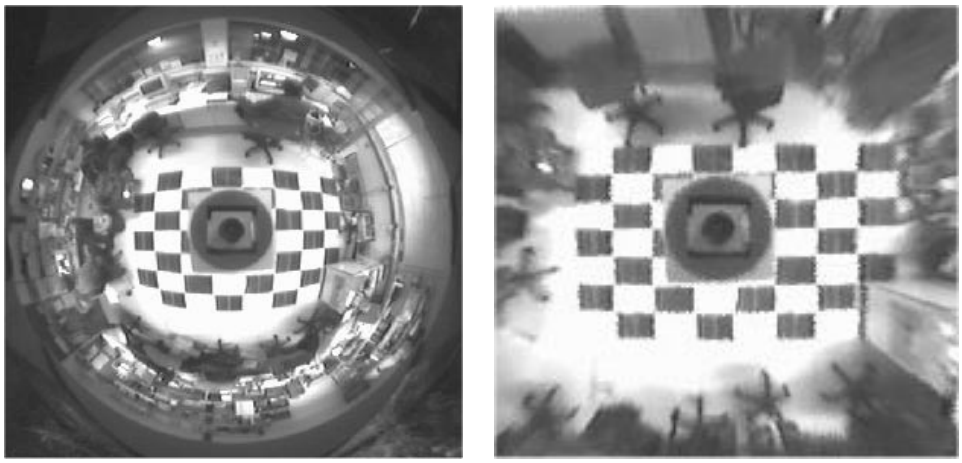
Due to the geometry of the mirror, the images acquired with our omnidirectional camera are naturally distorted. For instance, a corridor appears as an image band of variable width. In contrast, the bird's-eye view preserves all shapes on the ground plane (up to a scale factor).

To obtain a bird's-eye view [47], we rewrite the projection operator, $\mathcal{P}_\rho$ to relate radial distances, $\rho_{ground}$, measured on the ground plane, and radial distances, $\rho_{img}$, measured in the image:

$$\rho_{img} = \mathcal{P}_\rho(\rho_{ground}, \theta) \tag{2.4}$$

Then, using this information, we build a look up table which maps radial distances from the ground plane to their respective image co-ordinates. Since the inverse function cannot be expressed analytically, once we have an image point, we search the look up table to determine the corresponding radial distance on the ground plane. In this way, image remapping to a bird's-eye view is efficiently achieved.

Figure 2.12 shows an example of ground plane remapping to obtain the bird's-eye view. The ground pattern shown in the original image becomes a rectangular pattern in the bird's-eye view, as desired.



(a) (b)

**Fig. 2.12**: (a) The original omnidirectional image. (b) The ground plane remapped to a bird's-eye view image.

### 2.8.2 Design of a *Single Camera – Single Mirror* Catadioptric Sensor with a Specialised Mirror

The second camera design utilised a specialised mirror to image the scene around the robot [31, 151]. The difference between this mirror profile and the design utilising the spherical mirror is that the vertical dimension, at a given distance from the camera, is linearly mapped to the radial distance from the centre of the image plane. In addition, the sensor used was of a *log-polar* design. This system was developed as part of the EU IST project OMNIVIEWS involving three partners: DIST - University of Genova, CMP - Czech Technical University and ISR - Instituto Superior Técnico. It is shown in Figure 2.13.

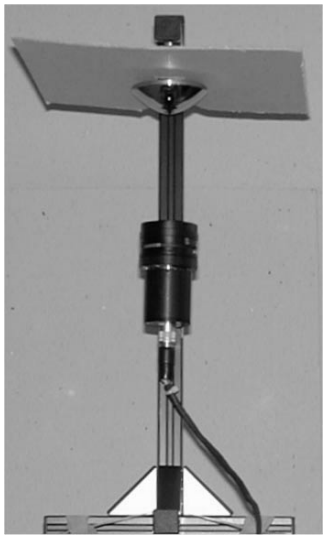


**Fig. 2.13**: The SVAVISCA omnidirectional camera with a specialised mirror

### Log-Polar Sensor

The rotational symmetry of the omnidirectional images immediately suggests the adequacy of using a polar pixel distribution. In this design the SVAVISCA log-polar sensor was used. For detailed information on the sensor, see [135]. By using this image sensor we gain the following:

- Panoramic images can be *directly* read out from the sensor without the need for any geometric transformations. Thus, we gain a speed increase over current omnidirectional camera designs.

- Panoramic images have constant azimuthal resolution due to the fact that the log-polar sensor is organised in concentric rings with a constant number of pixels.

The log polar sensor is shown in Figure 2.14.

Inspired by the resolution of the human retina, the log-polar sensor is divided into two parts: the fovea and the retina. The fovea is the inner part of the sensor and consists of 42 rings, with a uniform pixel density and a radius of $\rho_0 = 0.027273$cm. The retina is the outer part of the sensor, consisting of a set of 110 concentric circular

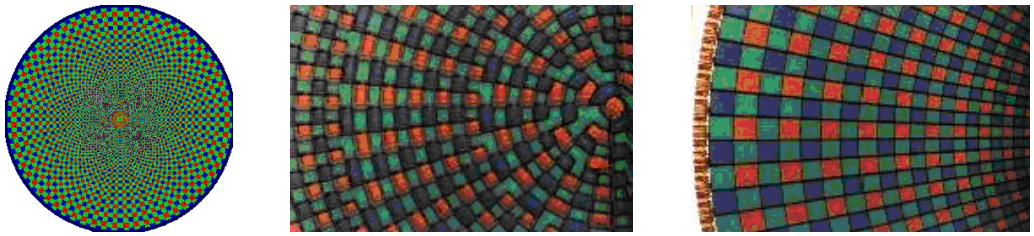


**Fig. 2.14**: General view of (a) the SVAVISCA Log Polar Sensor. Detailed views of the (b) foveal and (c) retinal regions.

rings, with 252 pixels each, whose resolution decays with a logarithmic law towards the image periphery.

In the retinal part of the sensor, the relationship between the linear distance, $\rho$, as measured on the sensor's surface, and the corresponding pixel coordinate, $p$ is specified by the following equation:

$$p = \log_k(\rho/\rho_0) \tag{2.5}$$

where $\rho_0$ and $k$ stand for the radius of the fovea and the rate of increase of pixel size towards the periphery, respectively.

### Mirror Profile Design

The image formation process is determined by the trajectory of rays that begin at a 3D point, are reflected by the mirror surface and finally intersect with the image plane. These reflections are governed by a projection function that specifies the SVAVISCA sensor. Overall, we want to define the mapping between some world distances, $y$, and corresponding distances measured in the image sensor, $\rho$. For the case of the log-polar sensor, the simplest mapping is to impose a linear relationship between 3D distances, $y$, and the logarithm of the distances measured on the SVAVISCA sensor, to account for the logarithmic pixel distribution:

$$y(\rho) = a \log_k(\rho/\rho_0) + b.$$

Here $a$ and $b$ mainly determine the visual field.

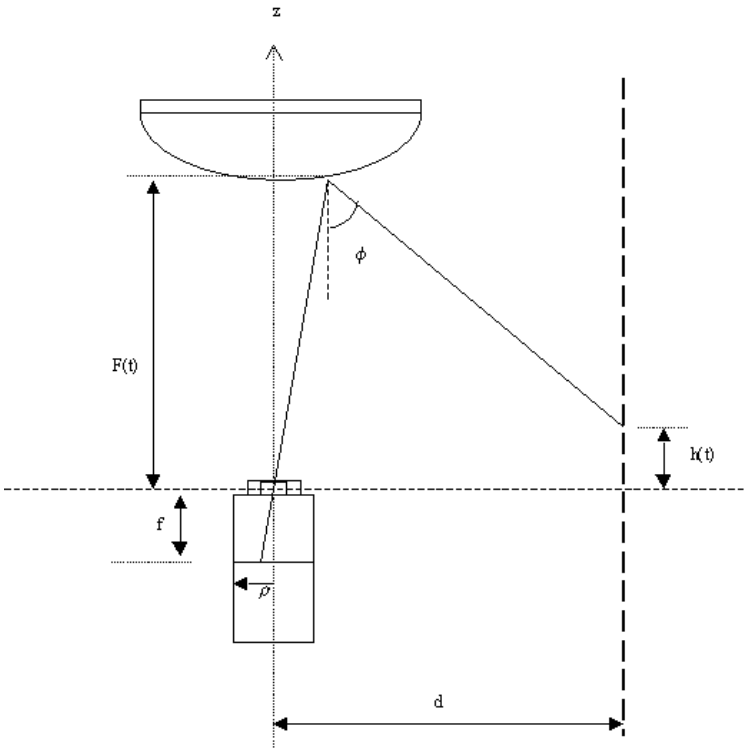



**Fig. 2.15**: Geometry of image formation using a catadioptric sensor with a constant vertical resolution mirror profile.

We now describe a mirror profile which images the world with constant vertical resolution.

### Constant Vertical Resolution

Due to the rotational symmetry of the system we need only consider the design of the mirror profile. The geometry of the image formation of our catadioptric sensor is shown in Figure 2.15.

We aim to preserve the relative vertical distances of points placed at a fixed distance, $d$, from the camera's optical axis. In other words, if we consider a cylinder of radius, $d$, aligned with the optical axis, we want to ensure that the ratios of distances, measured in the vertical direction, along the surface of the cylinder, would remain unchanged when

measured in the image. Such invariance should be obtained by adequately designing the mirror profile, yielding a constant vertical resolution mirror.

We start by deriving the relationship between the elevation angle, $\phi$, the mirror profile, $F(t)$ and the height, $h(t)$:

$$\tan(\phi) = \frac{d-t}{F(t)-h(t)} \tag{2.6}$$

which can be rewritten as:

$$h(t) = F(t) - \cot(\phi)(d-t) \tag{2.7}$$

Equating the angles of the incident and coincident rays on the mirror surface, we obtain:

$$h(t) = F(t) + \frac{2tF'(t) - F(t)(1-F'(t)^2)}{2F(t)F'(t) + t(1-F'(t)^2)}(d-t) \tag{2.8}$$

This equation relates the derivative of the mirror profile, $F'(t)$, to a given evolution of the height, $h(t)$. Solving this equation for $F'(t)$ we have:

$$F'(t) + \frac{t(d-t) + F(t)(F(t)-h(t))}{F(t)(d-t) - t(F(t)-h(t))} - \sqrt{\left[\frac{t(d-t) + F(t)(F(t)-h(t))}{F(t)(d-t) - t(F(t)-h(t))}\right]^2 + 1} = 0 \tag{2.9}$$

We can now relate the co-ordinate along the mirror profile, $t$, with the radial distance on the image plane, $\rho$, by introducing the perspective projection equation:

$$\rho = \frac{f\,t}{F(t)} \tag{2.10}$$

Finally, we can now introduce the constraint of invariance of vertical resolution by specifying that h(t) must be an affine map of the radial pixel co-ordinates. Now, considering the radial log-polar distribution of pixels, we have:

$$h(\rho) = a\log_k(\rho/\rho_0) + b \tag{2.11}$$

Hence, the procedure to determine the mirror profile is to integrate Equation (2.9), while $t$ varies from 0 to the mirror radius and replace $h(t)$ by Equations (2.10) and (2.11). The numerical integration was performed using MatLab's `ode45` function.

The initialisation of the integration process was done by computing the value of $F(0)$ that would allow the mirror rim to occupy the entire field-of-view of the sensor, while neglecting the thickness of the mirror shape (which is not available during initialisation). From Equation (2.10), we then have:

$$F(0) \approx F_0 = \frac{f\, t_{max}}{\rho_{max}} \tag{2.12}$$

where $t_{max}$ and $\rho_{max}$ represent the mirror and sensor radius, respectively.

## 2.9 Summary

This chapter presented the state-of-the-art in catadioptric sensor design. The designs used in this work were detailed: *a standard camera with a spherical mirror* and a *log-polar camera with a specialised mirror*. Our motivation for using omnidirectional vision was provided and the single centre of projection was discussed. The method used to remap omnidirectional images to scaled orthographic views of the ground plane was also presented.

# Chapter 3

# Environmental Representations

*This chapter defines an appropriate and efficient representation of the environment, suitable for use by a vision-based mobile robot. This representation should meet the criteria that it: (a) is easy to build, (b) requires a small amount of memory and (c) can be used for real-time localisation. Our motivation for choosing a* ***topological*** *approach is detailed, as are alternative environmental representations. A brief introduction to how humans model large-scale space is given. Our representation of the environment is omnidirectional image-based and is encoded by a low-dimensional eigenspace. The approach to building this subspace by using Principal Component Analysis is detailed and initial results are presented.*

## 3.1 Introduction

An essential component of a successful navigation system is an appropriate environmental representation, i.e. an internal model of the world stored by the mobile robot. Defining a successful system as one which can accomplish its given task, one can consider endowing a mobile robot with an environmental representation *tailored* to achieving the task. This is an important but subtle point: many previous research works chose to concentrate valuable computational resources on building an accurate representa-

tion of the environment, *whether it was required or not*. The building of full or partial 3D maps [167] is a case in point. In this work, we argue that shifting the emphasis from thinking about appropriate representations to the process of building these 3D maps, explains why most existing systems require large computational resources, but still lack the robustness required for many real-world applications.

Motivation for looking at the appropriate representation problem comes from studies of how humans store knowledge of large-scale spaces. From the available research, it seems that very parsimonious representations are memorised. This point is further addressed in Section 3.2 (p. 59).

A review of the current research shows two main environmental representations: *Geometric Maps* and *Topological Maps*. There are a number of benefits and shortcomings to each representation, which are detailed in Section 3.3 (p. 60). Naturally, geometric maps model the environment to a high degree of precision while topological approaches use a more functional representation. Considering that we wish to build a simple and efficient navigation system, the environmental representation we chose was topological in design. Our motivation for choosing a topological approach is detailed in Section 3.3 (p. 60) and the exact implementation details are presented in Section 3.4 (p. 66).

A secondary issue related to the use of any environmental representation is whether the robot should be required to build the representation or be provided with it. In the early days, robots were often loaded with detailed CAD models (i.e. maps) which specified both the structure and layout of the world. Needless to say, such maps did not deal well with dynamic environments. Additionally, they were often used in conjunction with sonar or laser data. Therefore, the robot's observation of the world, at a given instant, needed to easily fit within the map. Often, achieving this required overcoming sensor noise by using robust matching methods.

In order to provide any detailed *a priori* model to the robot, it first has to be

acquired. This is a difficult and time consuming task, primarily because effective localisation necessitates the extraction of **relevant** landmarks from the acquired data.

In our case, in order to build a map, the robot first traverses through the environment, simply acquiring images as it goes. These form the basis for the robot's representation of the environment.

## 3.2 Spatial Knowledge Representation

As previously stated, one of the key navigation components is knowledge of the environment. Given that humans and animals need to maintain an internal representation of the environment, inspiration can be taken from what is known about how they do so. Thus, before detailing how robots may represent their environment, we shall first give a brief overview of what is known about how humans and animals do so.

It is acknowledged that humans represent the world, internally, as a *Cognitive Map*. The development of this concept is widely credited to Tolman [140], who in 1948 suggested that stimuli from the world are "*worked over and elaborated in the central control room into a tentative, cognitive-like map of the environment*".

What exactly is the cognitive map? The answer is still unknown. One can say that this term is often associated with (what Kuipers describes as) the "Map in the Head" metaphor [81]. At a very basic level, this metaphor suggests that the way in which we store spatial information has a direct, isomorphic relationship to the everyday graphical map. Kuipers argued that while this metaphor is of some use, the implied, exact map-like qualities *do not exist*. In essence, if they did, then there would be a single central region in the brain where all information relating to large scale spaces is stored. This has been shown not to be the case. Therefore, it is highly probable that metric and topological information are stored separately. In 1960, Lynch [87] described results from experiments concerning how people navigate through urban environments. He clearly

showed that the correct topological relationships between locations were maintained, even when completely incorrect metric knowledge was remembered. This again suggests the separate storage of each. Additional support for the separation theory comes from the fact that we find bi-directional navigation in unknown environments highly difficult. For example, when a tourist in a city, finding one's way back to one's hotel is often confusing.

We shall now detail the available schemes a mobile robot can use to represent its environment. In deciding which to implement, we kept the above points in mind. To meet our goals successfully, a topological approach was chosen.

## 3.3 Environmental Representations

Clearly, the ability of a mobile robot to locate itself within its environment requires the availability of an appropriate environmental representation. Current representations can be placed into three distinct categories. These are as follows:

1. Geometric Representations
2. Topological Representations
3. Hybrid Representations

This categorisation is based on the level of environmental detail provided to the robot by each representation. Each class has its own particular merits and drawbacks, particularly when related to the task the robot is required to solve. Naturally, the greater the amount of information provided, the more precisely the robot can determine its position. How this aids in the navigation process shall be dealt with in the following sections.

### 3.3.1 Geometric Representations

In the literature, geometric representations are often referred to as Geometric Maps. Given that a scene is mapped in a detailed manner, expensive and accurate sensors (for example a laser scanner or a stereo head) are used to measure distances and angles from the robot to various objects within the environment. The key idea to keep in mind when talking about geometric mapping is that the world is represented in a *metric* manner. Thus, when navigating through an environment, accurate and reliable sensor measurements are required at each time instant.

Much of the earliest work in visual navigation concerned the geometric mapping of structured environments. Moravec's [96, 97] seminal work on the subject used a binocular set of cameras to recover the structure of an indoor scene. The position of cones on the floor was modelled by a 2D region on the ground plane. Thus, the environment was segmented into regions of free and occupied space allowing the robot to traverse a path from its current position to a goal point.

A particularly important case of geometric mapping is the *Grid-based* paradigm.

#### Grid-Based Mapping

Possibly, the most popular approach to geometrically mapping the environment is to build a grid-based map. One of the earliest methods was that proposed by Elfes and Moravec [38]. Here, the environment was divided into evenly spaced grids, termed Occupancy Grids. Each cell, within the grid, had an associated value which related information as to whether the cell was occupied or not. Initially, all values were set to 0.5, i.e. an equal probability that a cell was occupied or unoccupied.

It should be noted that the intrinsic geometric nature of grid-based maps directly corresponds to the structure of the robot's environmental surroundings. Thus, large amounts of memory and search time are required to store the information demanded to accurately capture this structure. In order to maintain this degree of accuracy,

real-world problems such as slippage and drift need to be overcome. The data received from the sensor(s) must also be integrated over time in order to keep the map up to date.

Grid-based mapping has been used in many research works, including [13, 139]. It obtains particularly successful results when navigating in cluttered environments. Although trap situations may occur, current research focuses upon overcoming such local minima [69].

A major problem with geometric approaches is that they tend to contain a large amount of irrelevant information. In particular, specific environmental cues which lead to effective means of navigating, are not explicitly represented; thus they are difficult to find. Clearly, this leads to an increase in the size of the search space; therefore, the computational cost of searching is high.

### 3.3.2 Topological Representations

In the last section, we noted that geometric representations are particularly suited to obtaining precise robot pose estimates. Given that this is not always a requirement for successful navigation [87], a second approach to map design was developed. This approach is topological in nature and was pioneered by Kuipers [80, 82] with his TOUR model. Essentially, it evolved from the available knowledge regarding how humans represented details of large-scale spaces (see Section 3.2 (p. 59)). It is the mapping scheme used in this work.

When using a topological map, the robot's environment is represented as a graph. *Nodes* in the graph correspond to recognisable scenes (distinctive landmarks) where specific actions may be elicited, such as entering a door, turning left, etc. *Links* connecting nodes in the topological map correspond to regions where some environmental structure can be used to control the robot.

A simple analogy can be made with human behaviour when walking around a city.

Figure 3.1 shows various landmarks in Lisbon, Portugal. In order to get from one particular locale to another, we do not have to think in *precise metric terms*. For example, to go from the city centre, **Rossio**, to **Saldanha** square, we may *go forward until we reach the statue of Marquês Pombal in* **Rotunda**, *turn right in the direction of* **Picoas** *and carry on until we finally reach* **Saldanha** *Square.*

In order to reach the final goal, the navigation problem is decomposed into a succession of subgoals that can be identified by recognisable landmarks. The required navigational skills are the ability to follow roads, make turns, etc., and recognise that we have reached a landmark.

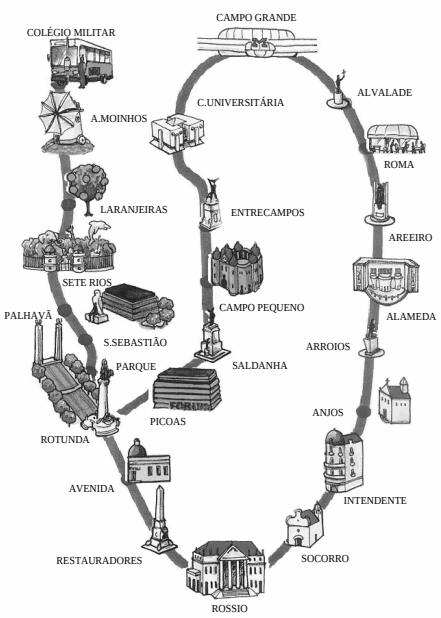



**Fig. 3.1**: A topological map of landmarks in Lisbon, Portugal.

In this work, a similar approach to robot navigation is adopted in order to accomplish missions such as *go to the third office on the left-hand side of the second corridor.*

### Motivation for the use of Topological Maps

Topological Maps exhibit a number of advantageous properties. As already detailed, cognitive scientists [129] have shown that a cognitive map is made up of successive layers and thus it has been suggested that a topological map is a natural description of the environment.

In [10], Brooks notes the advantages of topological maps over traditional geometric approaches. Given that uncertainties in measurement always occur, his idea was to utilise navigation algorithms which explicitly represented the uncertainties encountered in the real-world. He dropped the assumption that a map should be represented in a 2D co-ordinate system and stated that only relationships between parts of the map should be stored in a graph structure. Thus, while not using the word, topological, the aim was to produce a "*relational map, which is rubbery and stretchy*". In this way, when metric errors do occur, the location of certain places can still be correctly estimated. Thus, it is clear from the *qualitative* nature of topological maps that the inevitable problems of movement uncertainty are significantly reduced. One need only deal with proximity and order: since the robot navigates between nodes, global errors do not accumulate. Inherently, this means that the robot shall be able to overcome problems of drift and slippage more easily than if a geometric approach were used.

In addition, topological maps offer a parsimonious representation of the environment since their resolution is only dependent upon the complexity of their surroundings.

Lastly, the computational cost of path planning is significantly reduced by using topological maps. This was most strikingly demonstrated by Thrun and Bücken [138]. They generated 23,881,062 paths and found that when generating shortest paths, topological maps produced a performance loss of just 1.82% but that planning using a metric map was $4.9 \times 10^3$ times more expensive.

Conversely, topological maps can suffer from the sometimes ambiguous nature of scenes within the environment, especially when using sonar sensing. However, the

correct estimate of position can usually be determined by taking into account the robot's travel history. Traditionally, topological approaches have been sensitive to the camera's point of view, although this particular disadvantage is significantly reduced by the use of an omnidirectional camera. We believe that this camera design is particularly suited to capturing topology.

### 3.3.3 Hybrid Mapping

Hybrid mapping is the name given to the extraction of topological information from a geometric map. Thrun [137] presented a system which used sonar to sense the environment and learnt grid-based maps using neural networks and Bayesian integration. The key idea of the approach to extracting topological information is as follows: free space on the geometric map is partitioned into a smaller number of regions; each region is then separated by a link, or critical line. This partitioned map is then projected onto an isomorphic graph.

Fabrizi and Saffiotti [41] extracted what they termed "topology maps" from grid-based maps. Using the discipline of digital topology, they defined topological properties in a discrete space, as opposed to mathematical topology which relies upon continuous space. Primarily they defined nodes as large open spaces and used fuzzy mathematical morphology to extract them. They did not deal with localisation and were only concerned with map-building.

### 3.3.4 Our Approach

Our approach to building a topological representation of the environment uses an omnidirectional camera: no other sensing modality is required. The robot directly builds a topological representation of the environment, with no intermediate geometric mapping by simply traversing through the environment, acquiring images as it goes. This *a priori* information forms the basis of its internal representation. On subsequent

runs through the same environment, all that is required for **qualitative localisation** is that the current image be matched to one acquired *a priori*.

As input, we use gray level images, although naturally other types of images: edge detected or gradient intensity, for example, can be used. A typical sequence of omnidirectional images, acquired in a corridor environment, is shown in Figure 3.2. Here the robot is in the centre of each image. The corridor lines appear curved in the images due to reflection by the spherical mirror.

For the successful application of appearance-based techniques, matching must be reliable. It is clear that matching may fail if *similar* regions in the environment cannot easily be distinguished. As noted in Section 1.2.2 (p. 16), this problem was encountered in some previous research works. By using an omnidirectional camera the probability of an incorrect classification is significantly reduced, given the fact that the robot acquires visual information in 360° about the vertical.

## 3.4 Image Eigenspaces as Topological Maps

In general, sizeable learning sets are required to map the environment and so matching using traditional techniques, such as correlation, would incur a very high computational cost. If one considers the images as points in space, it follows that they shall be scattered throughout this space, *only* if they differ significantly from one other. However, many real-world environments (offices, highways etc.) exhibit homogeneity of structure, leading to a large amount of redundant information within the image set. For example, in Figure 3.2 the robot is common to all images and many show white corridor walls. Consequently, the images are not scattered throughout a high dimensional space but - due to their similarity - lie in a lower dimensional subspace. This subspace is often termed an **eigenspace**.

To present this idea in an informative manner, let us take an elementary example.

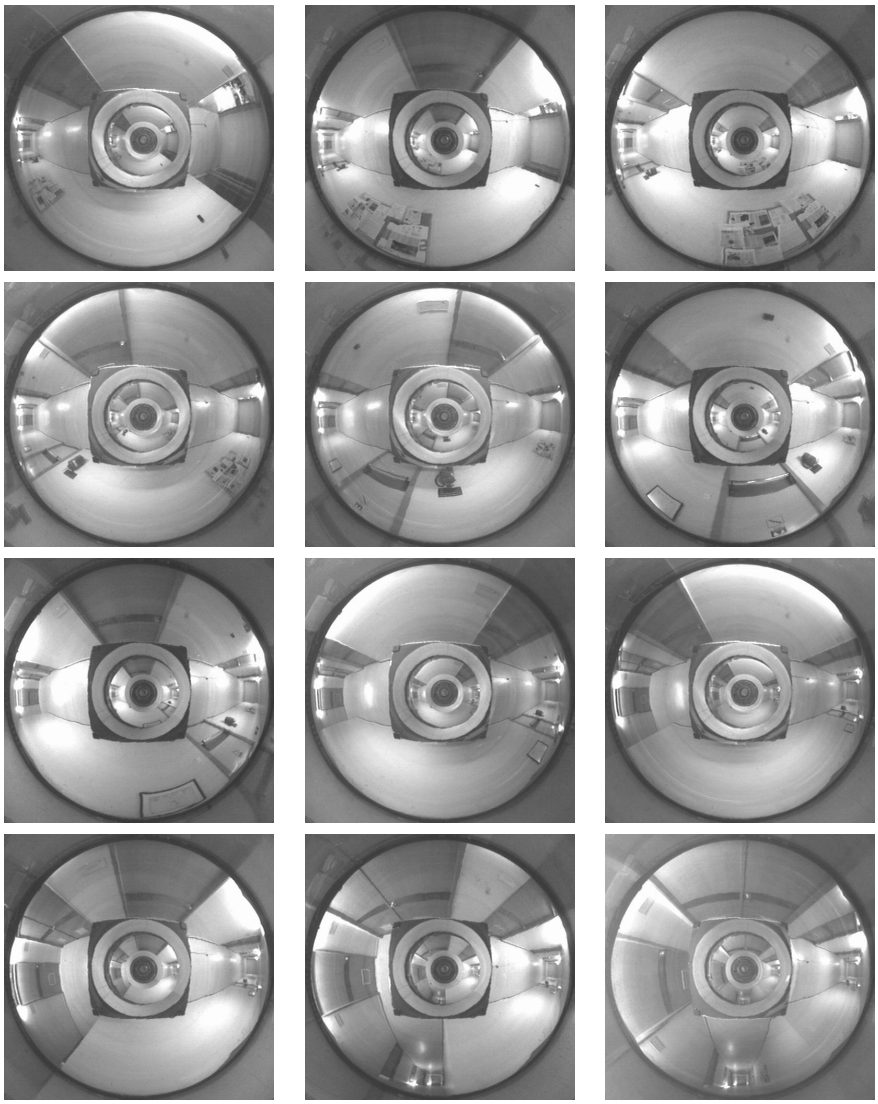


**Fig. 3.2**: A sequence, from left-to-right and top-to-bottom, of omnidirectional images acquired along a corridor at a full resolution of $516 \times 508$ pixels. Before applying Principal Component Analysis, these were reduced to a resolution of $128 \times 128$ pixels.

Figure 3.3(a) shows points in the 2D $xy$ co-ordinate system. Notice that the data

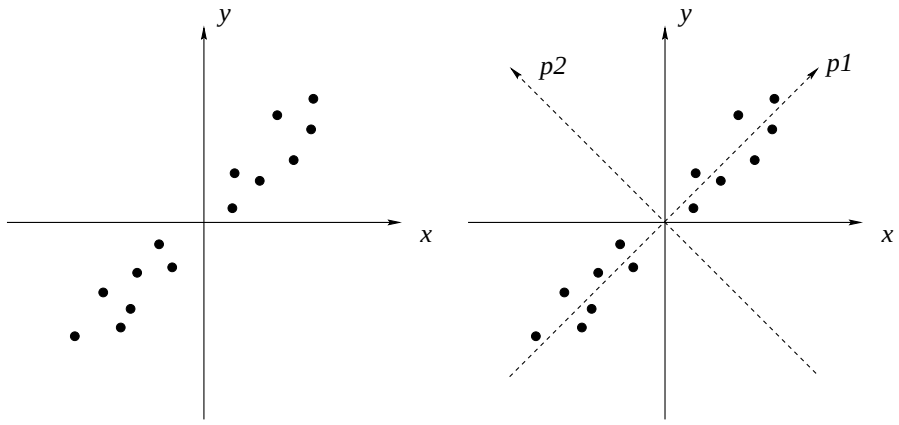



**Fig. 3.3**: A simple example showing how (a) 2D points can be represented by (b) a 1D line, i.e. dimensionality reduction.

is highly correlated. Given this fact, Figure 3.3(b) shows there is a more compact way to represent them: the projection of the data points along the line, $p1$ is a good approximation of the original 2D data.

The only remaining question to be answered is how to implement the actual process of dimensionality reduction, given a set of images. The curse of dimensionality has long been addressed by the statistics, communication and signal processing communities. A classical procedure to solve it is *Principal Component Analysis* (PCA)[1] [71], or as it is sometimes known, the application of the *Karhunen-Loève transform* [109, 147]. Simply put, Principal Component Analysis **reduces the dimensionality** of a set of linearly independent input variables, while still accurately representing most of the original data. The reconstruction of this original data is optimal in the sense that the mean square error between it and the original data is minimised.

Perhaps the first use of Principal Component Analysis in the field of computer vision was by Murase *et al.* in 1981 when they used it for hand-written character recognition

[1]The idea of principal components was first put forth in 1901 by Pearson [113] but was developed by Hotelling [63] in 1933.

[100]. In 1987, Sirovich and Kirby brought the technique to a wider audience by using it for the characterisation of human faces [132]. This work was followed, in 1991, by Turk and Pentland, who are widely cited as popularising the technique in the area of face recognition [142]. Other applications include visual inspection [105] and object recognition [102]. Recent work has focused upon tracking [9], position estimation using laser [145, 146], motion and gesture recognition [9, 22, 90].

Subspaces have been used in other information domains to represent structured information. For Kohonen [76] to explain the properties of the optimality of associative mappings, subspaces proved essential. Subspaces and logic have been linked by Watanabe [148], where a subspace was assigned to each proposition of a modular lattice. It was this connection that prompted Watanabe to put forward the idea of a subspace method of pattern recognition.

### 3.4.1 Building the Eigenspace

Let us now formally address the mathematical details of constructing the low-dimensional subspace. Imagine that we represent images as $L$-dimensional vectors in $\mathbb{R}^L$. Due to the similarity between images (data redundancy) these vectors will not span the entire space of $\mathbb{R}^L$ but rather, they will be confined (or close, to be more precise) to a lower-dimensional subspace, $\mathbb{R}^M$ where $M << L$. Hence, to save on computation, we can represent our images by their co-ordinates in such a lower-dimensional subspace, rather than using all of the pixel information. Each time a new image is acquired, its capture position can easily be determined by projecting it into the lower-dimensional subspace and finding its closest match from the *a priori* set of points (images).

The only question which remains to be answered is that of determining the lower-dimensional subspace from the original input data. A basis for such a linear subspace can be found through PCA, where the basis vectors are denominated **principal components**. They can be computed as the **eigenvectors** of the **covariance matrix** of

the normalised set of images acquired by the robot. The number of eigenvectors that can be computed in such a way is the same as the number of images in the input data, and the eigenvectors are the same size as the images.

### Preliminaries

The input data are composed of $N$ images, where each image, $\mathbf{I}_k$, is arranged as a single column vector of size $L$. We start by determining the average of all images, $\bar{\mathbf{I}}$ and subtract it from each image:

$$\tilde{\mathbf{I}}_k = \mathbf{I}_k - \bar{\mathbf{I}} \qquad \text{where} \qquad \bar{\mathbf{I}} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{I}_k \tag{3.1}$$

Subtracting the average image, $\bar{\mathbf{I}}$ ensures that the first eigenvector captures most of the variance of the set of images. An estimate, $\mathbf{R}$ of the covariance matrix of the set of images $\tilde{\mathbf{I}}_k$ can now be determined as follows:

$$\mathbf{R} = \mathbf{B}\mathbf{B}^{\mathbf{T}} \qquad \text{where} \qquad \mathbf{B} = \left[\tilde{\mathbf{I}}_\mathbf{1}\ \tilde{\mathbf{I}}_\mathbf{2}\ \cdots \tilde{\mathbf{I}}_\mathbf{N}\right] \tag{3.2}$$

where the columns of $\mathbf{B}$ are the input images $\tilde{\mathbf{I}}_k$. As we usually have a small number of images, i.e. $N << L$, from Equation 3.2, it follows that $\mathbf{R}$ will, at most, be of rank $N$. Hence, we only need to compute the first $N$ eigenvectors which are those representing the used data.

The input images can only be exactly represented by the entire set of eigenvectors. In general, however, the first few eigenvectors account for most of the information available within the input image set. Hence, we can use a small number of these eigenvectors as an orthonormal basis for a *lower* dimensional eigenspace, that is an efficient approximation to the larger input image space.

### How to Compute the Principal Components

There are a number of methods by which we can obtain the eigen-structure of $\mathbf{R}$. Conjugate Gradient [109] takes an iterative approach and finds the eigenvector that

maximised a scalar function thus giving the corresponding eigenvalue. The covariance matrix is then updated and the process begins again until the eigen-structure has been determined. Singular Value Decomposition (SVD) [74] is a well-known and powerful method for determining the eigen-structure. It is the method we chose due to its easy implementation, efficiency and numerical stability. The Spatial Temporal Adaptive algorithm [101] is faster than SVD, as it works on blocks of image data.

By using the SVD we can determine the eigenvectors, $\mathbf{e}_j$, and eigenvalues, $\lambda_j$, of $\mathbf{R}$.[2] These eigenvectors (actually only the first N are meaningful), form an orthonormal basis that can represent the entire input image set. Each eigenvector contains components from all of the images and thus an eigenvector is sometimes referred to as an *eigenimage*.

Unfortunately, computing the SVD of $\mathbf{R}$ is highly computationally intensive. One solution [99] to this problem lies in computing the SVD of $\tilde{\mathbf{R}} = \mathbf{B^T B}$. Thus, we obtain the eigenvectors, $\tilde{\mathbf{e}}$ and eigenvalues, $\tilde{\lambda}$ which are then converted to the required eigenvectors, $\mathbf{e}$ and eigenvalues, $\lambda$ by the equations:

$$\begin{aligned} \lambda_j &= \tilde{\lambda}_j \\ \\ \mathbf{e}_j &= \tilde{\lambda}_j^{-1/2} \mathbf{B} \tilde{\mathbf{e}}_j \qquad j \in 1..N. \end{aligned} \tag{3.3}$$

The next stage is to build a lower dimensional subspace using only the first $M \ll N$ eigenvectors. This lower dimensional eigenspace has the advantage of accounting for most of the variance in the images. The coefficient vector, $\mathcal{C}_k = [c_1^k \; c_2^k \cdots c_M^k]^T$, that represents the projection of an image, $\mathbf{I}_k$, into the eigenspace is obtained as follows:

$$c_j^k = \mathbf{e}_j^T . (\mathbf{I}_k - \bar{\mathbf{I}}) \tag{3.4}$$

For our navigation experiments (see Section 4.5) we keep the 10 eigenvectors with the highest eigenvalues, which are denominated, the *Principal Components*. Images are

[2] As $\mathbf{R}$ is symmetric and positive definite its eigenvalues are the same as its singular values. A proof is given in Appendix A.

coded by a vector of 10 coefficients that represent the projection of each input image along the principal components of the reduced-order manifold (eigenspace). These coefficients can be used to approximately reconstruct the input image using only the reduced-order manifold:

$$\mathbf{I}_k \approx \sum_{j=1}^{M} c_j^k \, \mathbf{e}_j \; + \bar{\mathbf{I}} \tag{3.5}$$

Each reference image, $\mathbf{I}_k$ is associated with a *qualitative* robot position (e.g. half way along the corridor). To find the robot position in the topological map, we have to determine the reference image that best matches the current view, $\mathcal{I}$.

The distance, $d_k^2$, between the current view and the reference images can be computed directly using their projections, $\mathcal{C}$ and $\mathbf{C}^k$, in the lower dimensional eigenspace:

$$d_k^2 = (\mathcal{C} - \mathbf{C}^k)^T \Lambda (\mathcal{C} - \mathbf{C}^k) \tag{3.6}$$

where $\Lambda$ is a diagonal matrix containing the (ordered) eigenvalues which express the relative importance of the various directions in the eigenspace. Notice that $d_k^2$ is computed between M-dimensional coefficient vectors (10 in our case), as opposed to image size vectors ($128 \times 128$). The position of the robot is that associated with the reference image, $\mathbf{I}_k$ having the lowest distance, $d_k^2$.

### 3.4.2 Properties

The first question we wish to address is: how many eigenvectors to use as a basis for the low-dimensional eigenspace? This information can be garnered from the eigenvalues of $\mathbf{R}$, associated with each of the eigenvectors. The largest eigenvalues correspond to the eigenvectors which capture most of the variance within the image set. The Frobenius norm (which quantifies the signal energy) of $\mathbf{R}$ can be determined by the sum of its eigenvalues. Hence, by taking the first 10 eigenvalues, we see that their associated

eigenvectors capture 84.9% of that norm, while the first 20 eigenvectors capture 93.3%. Thus, the first 10 capture a great portion of the variance within the omnidirectional image set and so high matching rates can be achieved using a 10D eigenspace. Figure 3.4 shows a graph of the eigenvalue drop-off.

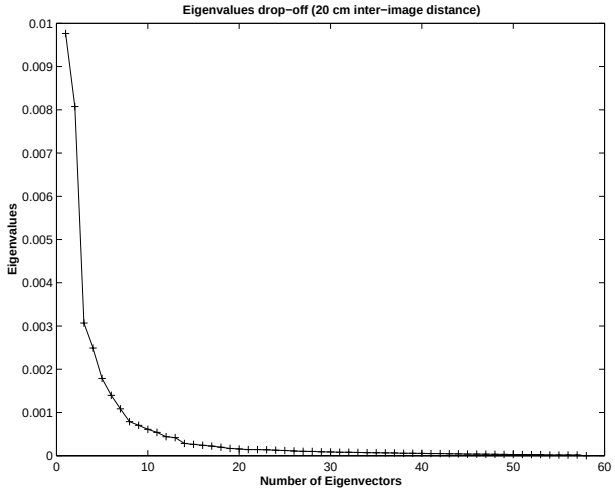



**Fig. 3.4**: The eigenvalue drop-off. Good matching results were obtained using the first 10 eigenvectors.

Figure 3.5 shows the first 9 eigenimages (eigenvectors) computed from 50 omnidirectional images, representing one corridor. They are shown in descending order, from left-to-right and top-to-bottom, in accordance with their eigenvalues. Here, we can easily see that the most general information (i.e. that common to all images) is stored by the first eigenvector, while more specific details are stored by those of a lower rank. Thus, for example, information only seen in one image of the *a priori* data set will not be captured by the low-dimensional eigenspace.

An important property of the eigenspace is that it is optimal in the correlation sense: points in the eigenspace which are closely related correspond to similar images in terms of the $l^2 norm$.

There is an additional benefit to building a low-dimensional eigenspace representation of topological structure using omnidirectional images: the *same* eigenspace can be

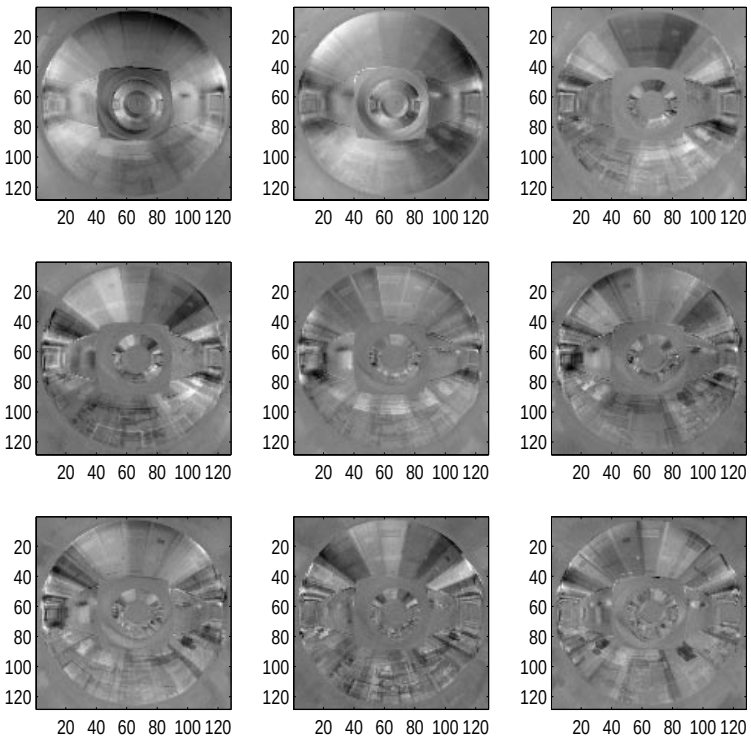



**Fig. 3.5**: The first 9 (omnidirectional) eigenimages obtained via Principal Component Analysis.

| **Distance** | **20cm** | **30cm** | **40cm** | **50cm** |
|---|---|---|---|---|
| *Dimension* | | | | |
| *2D* | 73.7% (42/57) | 84.2% (32/38) | 82.1% (23/28) | 47.8% (11/23) |
| *5D* | 100% (57/57) | 100% (38/38) | 100% (28/28) | 91.3% (21/23) |
| *10D* | 100% (57/57) | 100% (38/38) | 100% (28/28) | 100% (23/23) |
| *20D* | 100% (57/57) | 100% (38/38) | 100% (28/28) | 100% (23/23) |

**Table 3.1**: Matching Results using eigenspaces of differing dimensions.

used along both the forward and return trajectories, simply by rotating, in real-time, the acquired omnidirectional images by 180°.

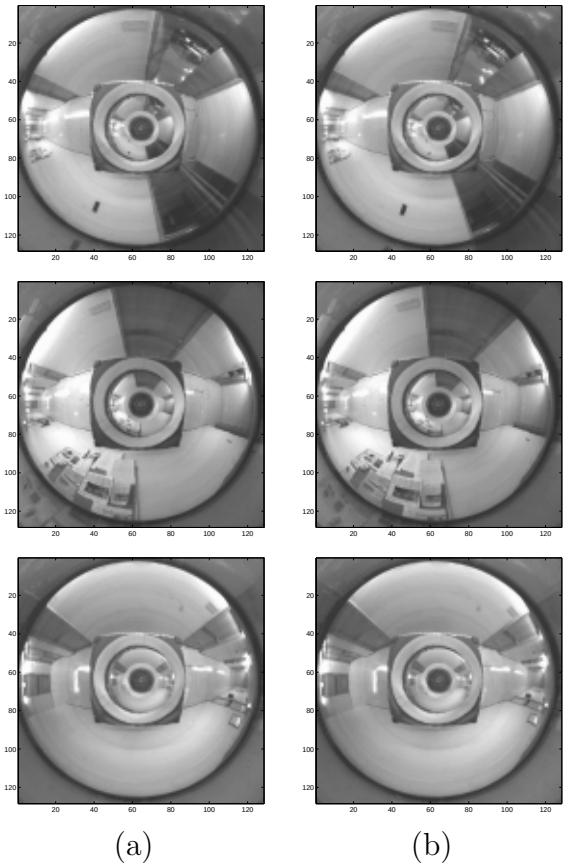
### 3.4.3 Initial Matching Results

We now go on to detail some initial matching results. For these experiments, no robot navigation took place: the goal was to assess the applicability of our approach. Real-world navigation results are presented in Chapter 4.

A sequence of 115 omnidirectional images were acquired at 10cm intervals along a corridor environment. A selection of these images are shown in Figure 3.2. Each was acquired at full resolution, then filtered and subsampled to a resolution of $128 \times 128$ pixels in size. The odd numbered images, 58 in all, were used to construct a number of topological representations of the environment, i.e. a number of low-dimensional eigenspaces. The other 57 even numbered images were used as a test set for matching. Significantly, two parameters required consideration: (*i*) the image sampling density and (*ii*) the number of eigenvectors.

Given these parameters, a number of solutions were tested, with differing inter-image distances: 20cm (57 images), 30cm (38 images), 40cm (28 images) and 50cm (23 images), respectively. For each image density, four associated eigenspaces - 2D, 5D, 10D and 20D - were constructed. The matching results achieved are summarised in Table 3.1.

Clearly, these results show that in our indoor office environment, a 10D eigenspace



(a) (b)

**Fig. 3.6**: IST Set: A selection of (a)omnidirectional test images and (b) their closest matches obtained by projection into a 10D eigenspace. The *a priori* inter-image distance was 20cm and each image was $128 \times 128$ pixels in size.

was required to correctly match images in all cases. This result is also borne out by the eigenvalue drop-off graph shown in Figure 3.4. A higher dimensional eigenspace yields the same results. Naturally, as the inter-image distance increases, the similarity between images decreases. Thus, if an eigenspace of a very low-dimension is used to encode topological information, incorrect matches occur. This is particularly evident when using a 2D or 5D eigenspace with an inter-image distance of 50cm. A selection of image matches are shown in Figure 3.6. For the real-world navigation experiments outlined in Chapter 4, a 10-dimensional eigenspace was used.

In order to test the method in a different environment, images were acquired at

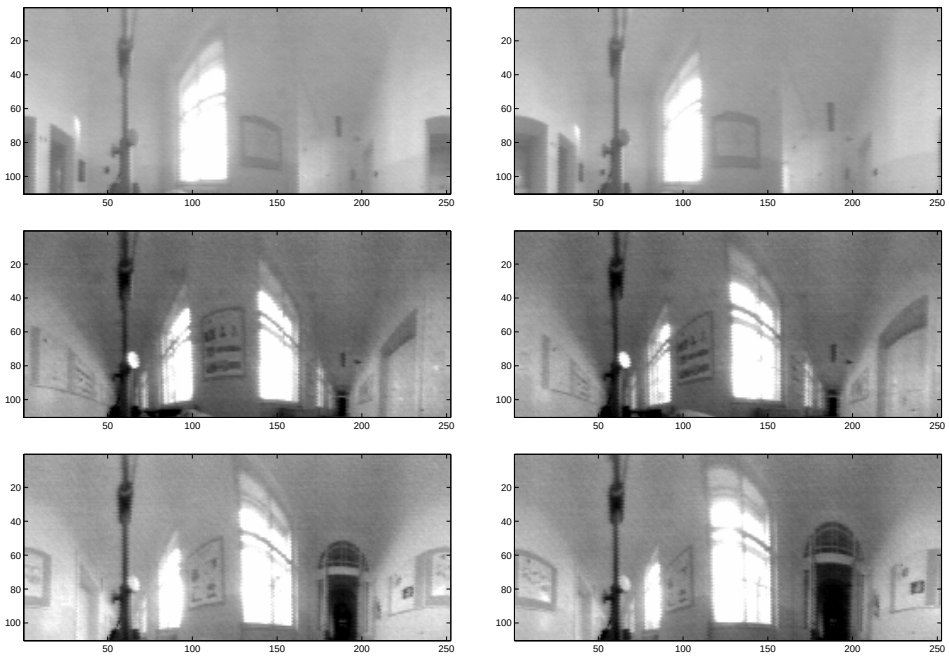



**Fig. 3.7**: CMP Set: A selection of (a) panoramic test images and (b) their closest matches obtained by projection into a 10D eigenspace. The *a priori* inter-image distance was 50cm and each image was $252 \times 110$ pixels in size.

the Center for Machine Perception in Prague[3]. Here the SVAVISCA camera design, outlined in Section 2.8.2, was used to capture a set of 64 images of $252 \times 110$ pixels in size. The inter-image distance was 50cm. The 32 odd-numbered images were used to build a low-dimensional eigenspace representation of the environment, while the 32 even-numbered images were used for testing. Using a 10D eigenspace, correct matching was achieved in 100% of cases. Figure 3.7 shows a selection of image matches.

## 3.5 Summary

This chapter presented a solution to the problem of building an appropriate environmental representation for an indoor mobile robot. The key idea presented was that any

[3]My thanks to T. Pajdla & B. Mičušik for acquiring the image set.

representation should be specifically *tailored* to the task at hand. Since our robot was designed to navigate globally, we used a topological approach. When constructed with input from an omnidirectional camera, this environmental representation produced initial matching results which were very encouraging. We presented a low-dimensional eigenspace, obtained from Principal Component Analysis, as an effective method of encoding the topological map.

We now go on to address the vision-based navigation problem in detail. Our approach to global navigation relies upon the environmental representation detailed in this chapter. For local pose control, we visual servo upon bird's-eye views of the ground plane.

# Chapter 4

# Vision-based Navigation

*This chapter concerns vision-based robot navigation using an omnidirectional camera. Our topological environmental representation is built upon with the addition of local pose control. This control mechanism visually servos on environmental features, extracted from bird's-eye view images. Real-world navigation results are presented using the camera designs detailed in Chapter 2. We identify the path distance/accuracy trade-off as being crucial to the success of any overall navigation scheme and results from integrated experiments are presented. Finally, we detail navigation results obtained in areas exhibiting strong non-uniform illumination change.*

## 4.1 Introduction

In Chapter 3, we detailed how to build an appropriate environmental representation for a vision-based mobile robot. Endowing a robot with this representation is critical to its ability to navigate, but note that sole reliance upon the representation does not solve the navigation problem. This is because successful navigation requires three components [84]: (*i*) the availability of an environmental representation, (*ii*) the ability to localise and (*iii*) the ability to path plan. These issues are further addressed in Section 4.1.1 (p. 82).

One of the strengths of our technique lies in the fact that we closely relate navigation components (*i*) and (*ii*). This allows our robot to make *effective* use of the environmental representation for **qualitative localisation**. This effectiveness is guaranteed by the use of a local control strategy which is visually based and allows the robot to maintain its position and orientation as it travels through the environment. Pose is controlled by visually servoing upon corridor guidelines, extracted from *bird's-eye views* of the ground plane, which simplifies the task. These views are obtained by the remapping of omnidirectional images. Details of the remapping method were presented in Section 2.8.1 (p. 50). Additionally, the close coupling of components (*i*) and (*ii*) allows for *actions* initiated by the robot to be directly linked to its *perception* of the world. For example, in order to effectively use the image eigenspace, images captured by the camera should be similar to those in the eigenspace. This similarity is ensured by implementing an action to keep the robot centred in each corridor, where perception closely matches the environmental representation.

The vision-based navigation method presented in this work is undertaken by *combining* appearance-based methods and visual servoing.

**Appearance-based methods** are defined as those which rely upon direct image matching techniques. In many research works this involves feature extraction or template matching, rather than matching the actual images themselves. In cases where actual images are used, often they are from narrow field-of-view cameras and so are less appropriate for topological mapping. More details about previous research in this area and how it relates to this work were presented in Section 1.2.2 (p. 16).

**Visual servoing** [40, 64, 150] is defined as the ability to control the position of a robot, relative to some feature in its environment, and within its field of view. Suitable examples include corridor guidelines, door frames or table edges.

Navigation represents a core research area of interest to the robotics community worldwide. Indeed, one may say that *independent* and *autonomous* robot navigation

represents their "Holy Grail". Ongoing work seeks to design a robot that can function autonomously in real-world environments. Given that these include both indoor and outdoor scenes, the prevailing research is delineated along these same lines: methods that work well outdoors may not do so inside. Moreover, it is usual that outdoor navigation requires large and expensive robots, equipped with a diverse range of sensing capabilities: sonar, laser, GPS, vision, gyroscopes etc. Such commercially available robots can cost in the region of $40,000 - $60,000, while designs built in-house cost significantly more.

Given the difficult nature of navigating in outdoor environments, complex solutions were often adopted. The general approach used was to generate the maximum amount of information about the structure of the environment and then to furnish the robot with significant computation power in order to integrate, and then, quickly process information from the sensors. By using this approach, successful results were achieved, including rough terrain traversal [53] and highway navigation [7, 35].

Research into indoor mobile robot navigation has been undertaken for approximately the last 25 years, often using complex navigation algorithms which relied upon multifaceted sensory information. In this work, we view the problem from the opposite end of the spectrum: we wish to obtain information from the environment in the simplest manner possible while reliably achieving the navigation task. Both robustness and an efficient usage of computational and sensory resources can be achieved by using visual information in closed loop to accomplish specific navigation tasks or behaviours [120, 121]. However, this approach cannot deal with global tasks or co-ordinate systems (e.g. going to a distant goal), because it lacks adequate representations of the environment. Hence, a challenging problem is that of extending these local behaviours, without having to build complex 3D representations of the environment. We address this problem by combining an omnidirectional camera, presented in Chapter 2, an appropriate environmental representation, presented in Chapter 3 and a local control

strategy, presented in this chapter.

In this work, we do not aim to integrate information from a number of sensors: our goal will be successful if we show that it is possible to navigate through structured environments using the visual input from each of the camera designs, respectively, outlined in Chapter 2.

### 4.1.1 Navigation Components

The ability to navigate is vital to the successful application of mobile robot technologies. Broadly speaking, mobile robot navigation requires three components:

- An Environmental Representation
- Position Estimation (also known as Localisation)
- Path Planning

Simply put, the robot is required to: (*i*) possess knowledge of the environment, encapsulated in an internal representation (or map) (*ii*) know its current position on the map and (*iii*) be able to plan a route from one point on the map to another. There are a number of sub-issues related to each of these requirements. Chapter 3 dealt with environmental representations; we shall now detail sub-issues related to position estimation and path planning.

**Position Estimation:** The main issue to be considered here is the degree of localisation required for the task at hand. We note that when travelling long distances, one does not need an extremely accurate estimate of one's position in the environment: it is sufficient to know one's qualitative position. Alternatively, for tasks undertaken over a short distance (passing through a door, for example) a far more accurate estimate of position is usually required. While the approach outlined in this work is suitable for

qualitative localisation during global navigation, tasks requiring very precise measurements are not dealt with. On the other hand, in a large-scale experiment, we show how our methodology forms part of an overall navigation scheme encompassing both qualitative and precise navigation tasks [48]. This observation of a *path distance/accuracy* trade-off is further addressed in Section 4.6 (p. 94), and associated experimental results are presented in Section 4.6.1 (p. 95).

**Path Planning:** Path Planning involves determining how to get from one point to another, usually in the shortest manner possible. In this work, we do not deal with this problem explicitly, although the topological map produced can be used as input to any standard graph planning (search) algorithm, $A^*$ for example. In our case, the robot's destination is simply specified by selecting an image of the goal location and if required, corners along the route.

## 4.2 Experimental Set-up

For the real-world experiments outlined in this dissertation, two robots and omnidirectional cameras were used. The first was a TRC Labmate from HelpMate Robotics, Inc. It was equipped with an omnidirectional vision system as shown in Figure 4.1.

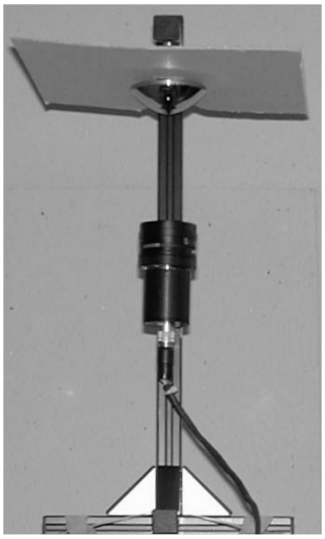
The design of this camera was detailed in Section 2.8.1 (p. 46). The system contained a Cohu CCD camera pointed upwards, viewing a spherical mirror. Grayscale images were captured with a full resolution of $768 \times 576$ pixels and subsequently subsampled to images of $128 \times 128$ pixels in size. PCA was then applied to build image eigenspaces, representing topological information. In order to visually servo upon corridor guidelines, bird's-eye view images of $600 \times 600$ pixels in size were used. All the processing was carried out on-board the mobile platform by a Pentium II 350MHz PC with 160MB of RAM. This system ran Matlab 5.3 under Windows 95 and achieved a

(a) (b)

**Fig. 4.1**: (a) The omnidirectional camera with a spherical mirror and (b) the camera mounted on a Labmate mobile platform.

sampling frequency of 2Hz.

The second set-up was a SCOUT mobile platform from Nomadic Technologies, as shown in Figure 4.2. It too was equipped with an omnidirectional camera but this design was that outlined in Section 2.8.2 (p. 51). It has a specially designed mirror profile combined with a log-polar camera. Panoramic images were obtained directly from this camera. Images were captured at a resolution of $252 \times 110$ pixels and PCA was directly applied to them. Visual servoing was undertaken using bird's-eye view images of only $200 \times 200$ pixels in size. The on-board processing power available was limited to a Pentium 166MHz with 64MB of RAM. This system ran Matlab 5.1 under Red Hat Linux 6.0 and achieved a sampling frequency of 1Hz. A challenging problem was to determine if such limited processing power, along with low resolution images, allowed for successful navigation.



(a)

(b)

**Fig. 4.2**: (a) The SVAVISCA omnidirectional camera with a specialised mirror and (b) the camera mounted on a SCOUT mobile platform.

## 4.3 Qualitative Localisation

As previously stated, we use a topological representation of the environment for global navigation. Each reference image is associated with a qualitative robot position. As detailed in Section 3.4.1 (p. 69), localisation is achieved by directly computing the distance, $d_k^2$, between the current view and the reference images by using their projections, $\mathcal{C}$ and $\mathbf{C}^k$, into the low-dimensional eigenspace:

$$d_k^2 = (\mathcal{C} - \mathbf{C}^k)^T \Lambda (\mathcal{C} - \mathbf{C}^k) \tag{4.1}$$

where $\Lambda$ is a diagonal matrix containing the (ordered) eigenvalues which express the relative importance of the various directions in the eigenspace. The position of the robot is that associated with the closest reference image.

The first real-world tests show that appearance-based methods can reliably provide qualitative estimates of the robot's position in the world, using a single corridor as an example. For this purpose we acquired an *a priori* set of images, $P$, and subsequently

ran the robot in the corridor to acquire a different set of run-time images, $R$. Although the *a priori* images and test sets used for the initial matching results, detailed in Section 3.4.3 (p. 75) were different, they were captured on a *single* run through the environment. The advantage of this approach was that similarity between images was easy to guarantee. As can be seen from previous research, other appearance-based techniques relied upon this approach. This is not so in our case: for navigation we wish to use the *a priori* set of images as our topological representation over multiple runs.

Figure 4.3 shows the distance, $d_k^2$ between the *a priori* and run-time images, $P$ and $R$.

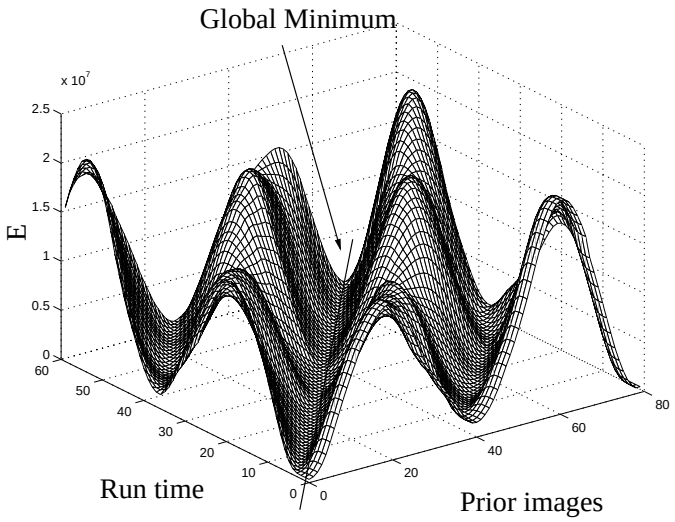



**Fig. 4.3**: A 3D plot of images acquired at run time, R versus those acquired *a priori*, P. This plot represents the traversal of a single corridor. The global minimum is the estimate of the robot's topological position.

It can be seen that the *global* minimum, at each time instant, is the correct estimate of the robot's current topological position. The error surface degrades gracefully in the neighbourhood of the correct estimates: the valley is smooth and relatively wide. Local minima appear, as side lobes of this surface, because some distant areas of the corridor

look very similar. These local minima are easily avoided by restricting the search space to images close to the previously estimated position, since images are captured sequentially according to the direction of motion.

In the vast majority of cases we have always obtained the correct estimate for the robot position, even in the presence of occlusions. If an incorrect classification occurs, the results are always in the vicinity of the correct answer, due to the smooth nature of the error function.

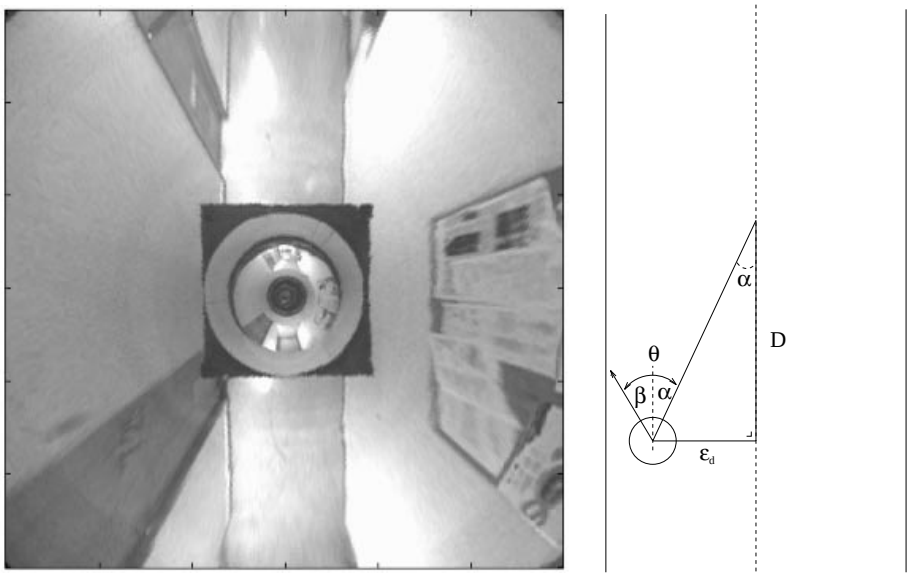
## 4.4 Adding Local Control

In this section, we detail our local control strategy to aid in the process of navigation. We wish to emphasise that it should easily combine with our topological environmental representation, thus forming a core part of our holistic approach to navigation.

In order to effectively navigate using the topological map, we must define a suitable vision-based behaviour for corridor following (*links* in the topological map). The goal is to control the robot in an efficient manner and thus inspiration is again taken from humans. As an example, when driving, humans make effective use of demarcations along the road for guidance. Similarly, since most corridors have parallel guidelines, we can exploit this information to drive the robot along the corridor. As we use an omnidirectional camera, these guidelines are always identifiable and so our corridor following behaviour can be easily implemented. Significantly, in different environments one can always use simple knowledge about the scene geometry to define other similar behaviours.

To keep the robot centred in the corridor, we control the heading direction by using visual feedback from the omnidirectional camera. The use of bird's-eye views of the ground plane significantly simplifies the servoing task, because the images become a scaled orthographic projection of the ground plane co-ordinates (i.e. no perspective

effects). Figure 4.4 (a) shows a bird's-eye view of the corridor, while Figure 4.4 (b) shows a top view of the corridor guidelines, the robot and the trajectory to follow in the centre of the corridor.




(a) (b)

**Fig. 4.4**: (a) A bird's-eye view of the corridor and (b) the measurements used in the control law: the robot heading, $\beta$, the distance, $\epsilon_d$ to the corridor centre, and the angle, $\alpha$ towards a point ahead in the corridor central path. The error used for controlling the robot orientation is $\theta$.

From the images we can measure the robot heading, $\beta$, with respect to the corridor guidelines, and the distance from the robot to the central reference trajectory, $\epsilon_d$. Driving $\beta$ to zero ensures that the robot moves parallel to the corridor, but is not necessarily centred.

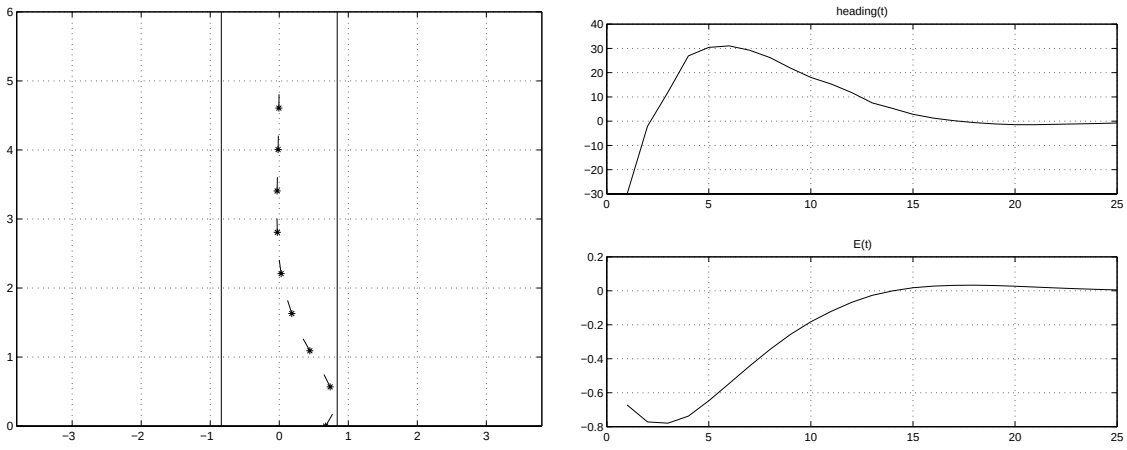
We use a simple kinematic planner to control the robot position and orientation in the corridor, using the angular velocity as the single degree of freedom. We consider a *look-ahead* distance, $D$, that defines, at each instant, the goal point that the robot should aim for. This goal point can be translated into an angular error, $\alpha$ as follows:

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

The value of D can be increased or decreased to control the influence of the displacement on the overall control scheme and may be a function of velocity. For our purposes, it was set to 7.5% of the image width, which corresponds to about 1 metre. Finally, the robot angular velocity, $\omega$, is driven by an angular error that combines both the robot heading error, $\beta$ and the angle, $\alpha$ which represents the position error, $\epsilon_d$:

$$\omega = -\mathcal{K}(\beta_i + \alpha_i)$$

where the proportional gain $\mathcal{K}$ was set to 0.45, thus ensuring a smooth trajectory back to the centre of the corridor (in general, it depends on the sampling period). Figure 4.5 shows a simulation of applying this control scheme to our navigational problem. The left plot shows the trajectory of the robot (starting from the bottom of the plot), while the right plot shows the change in the robot's heading and displacement over the same period.




(a) (b)

**Fig. 4.5**: Simulated results of the proposed control scheme: (a) Robot trajectory and (b) heading direction and translation. Distances are expressed in metres and the heading in degrees.

Notice that by using the bird's-eye view of the ground plane to extract the corridor guidelines, it becomes much simpler to translate that information into the robot

position and orientation errors with respect to the central path in the corridor. In the following subsection, we discuss the method to track the corridor guidelines over time.

### 4.4.1 Line Tracking using Prediction

Tracking the corridor guidelines is simplified because we only deal with scaled orthographic views of the ground plane. Hence, it is easy to determine projective-planar transformations (homographies) [56] that predict points and lines from one image to the next.

The corridor lines are extracted by, first finding edges within pre-definable bounding boxes, and then using a robust line fitting procedure based on RANSAC [42]. Predicting the position of these bounding boxes is a determinant for the robustness of the tracker, which could otherwise start tracking door frames and other image lines. In the following paragraphs we will describe the prediction process.

When a camera observes a plane under perspective projection, it induces a one-to-one projective-planar transformation that can map image points to the ground plane directly, and vice-versa. Similarly, one can map corresponding image points at two time instants by using inter-image homographies.

The homography, $\mathcal{H}_{iw}$, that determines the image projection, ${}^{i}p$, of a ground plane point, ${}^{w}P$, encapsulates the projection model, the camera's intrinsic parameters and the world-robot-camera co-ordinate changes:

$$ {}^{i}p = \mathcal{H}_{iw}\,{}^{w}P, \quad \text{with} \quad {}^{i}p = \gamma[u \;\; v \;\; 1]^T, \qquad \text{and} \quad {}^{w}P = [x \;\; y \;\; 1]^T $$

where $\gamma$ is a projective scale factor. Similarly, if we know the homography that relates two consecutive images, $\mathcal{H}_{im}$, we can predict where an image point ${}^{i}p_t$ will appear in the following frame, ${}^{i}p_{t+1}$:

$$ {}^{i}p_{t+1} = \mathcal{H}_{im}\,{}^{i}p_t $$

Since the bird's-eye view images are equivalent to a scaled orthographic projection of the ground plane, $\mathcal{H}_{iw}$ basically contains a scale factor and the inter-image homographies, $\mathcal{H}_{im}$ can be computed reliably using differential odometric information. Future improvements to our vision processing algorithm will eliminate the need for this odometric information altogether.

We can go even further by applying this prediction process directly to the corridor guidelines. If $\mathcal{H}_{im}$ maps points in the 2-dimensional projective space, the *inverse transpose* of this transformation can be used to map lines directly:

$$^{i}l_{t+1} = \mathcal{H}_l \, ^{i}l_t \quad \text{with} \quad \mathcal{H}_l = \mathcal{H}_{im}^{-T}$$

where $^{i}l$ are the projective co-ordinates of a line on a plane.

Figure 4.6 shows a sequence of bird's-eye view images acquired during tracking. The dashes represent the predicted positions of the central point of the bounding box extremities. The prediction is very accurate and vastly improves the probability of extracting the corridor guidelines rather than erroneous data such as door frames.

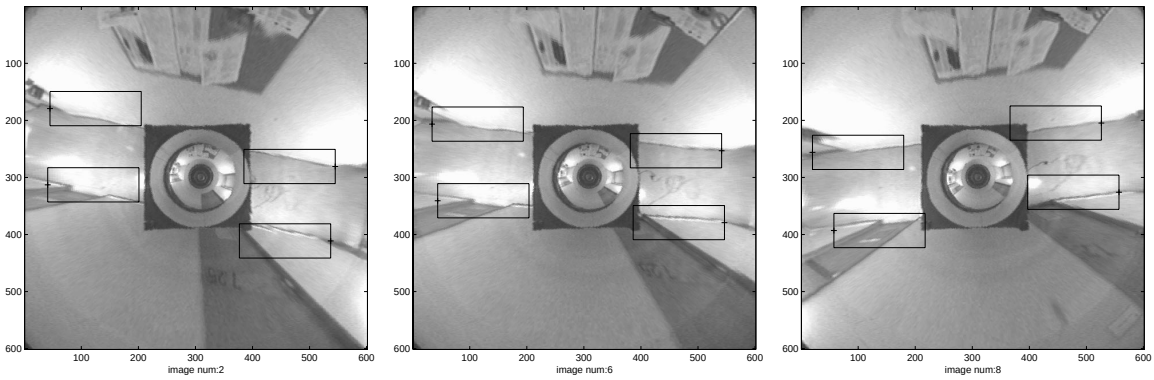



**Fig. 4.6**: Ground plane views of the robot's orientation and translation over time. The dashes represent the predicted position of each of the bounding box extremities.

## 4.5 Navigation Results

The experiments described in this work were undertaken in a typical indoor environment with corridors, offices and laboratories. A series of experiments were conducted to test the navigation behaviours required to perform global missions using the topological representation of the environment. As previously stated, the topological representation was built with omnidirectional images. For navigation experiments these were captured every 50cm along the corridors. Reference positions were ordered according to the direction of motion, thus maintaining a causality constraint. As detailed in Section 4.3 (p. 85), localisation was achieved by comparing the current view to those reference images acquired *a priori*.

To drive the robot along a central trajectory in the corridor, we used the behaviours described in Section 4.4 (p. 87). This was accomplished by using bird's-eye views of the ground plane to track the corridor guidelines and determine the robot's heading and position errors, relative to the desired path. This information was then 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.

Figure 4.7 shows results obtained with the Labmate mobile platform when navigation along a section of the $7^{th}$ floor at the Institute for Systems and Robotics. The distance travelled was approximately 21 metres. Odometry was used to display the path graphically.

Figure 4.8 shows a sequence of images of the SCOUT mobile robot navigating in a corridor environment. It began navigating outside an office, travelled down the corridor and turned 180°, before returning to its start position. The total distance travelled was approximately 17 metres.

These results show that the proposed control scheme can successfully drive the robot along the designated route. Additionally, they illustrate that when the robot arrives at the end of a corridor, it can switch to a different behaviour. In these examples, the

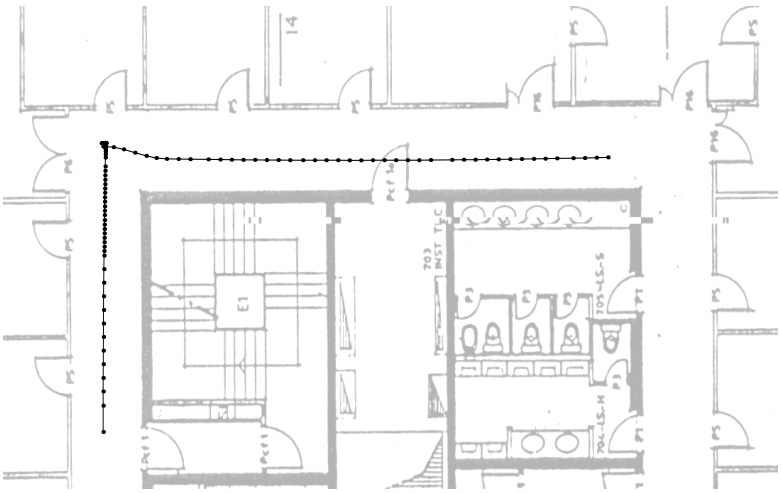


**Fig. 4.7**: One of the paths travelled by the robot at IST. The total distance travelled was approximately 21 metres.

**Fig. 4.8**: A sequence of images of the SCOUT mobile robot navigating in a typical indoor environment.

behaviour launched at the end of each corridor is simply to perform either a 90°, in order to proceed to the next corridor or a 180° turn, so as to travel back down the same corridor.

Overall, the holistic combination of a suitable camera geometry, an appropriate environmental representation, a means of qualitative localisation and a local control scheme provided an autonomous system with the ability to successfully achieve the goal of navigation.

## 4.6 Path Distance Versus Accuracy

We have just detailed successful *global* navigation results for a visually guided mobile robot. Here a topological environmental representation was used for navigating between distant environmental sites, using only qualitative and topological properties of the environment. Thus, we can characterise this approach by its ability to cover large distances to a given accuracy. This accuracy is specified by the distance between images in the *a priori* acquired set. Significantly, notice that the goal of reaching the final destination specifies the success of this task: whether very high accuracy is achieved or not is of secondary concern.

However, a different approach and success criterion are necessary when precise guidance or localisation are required for tasks such as docking or navigating in cluttered environments. For these precise navigation problems, Gaspar and Santos-Victor [47] proposed an approach they termed *Visual Path Following*. Visual Path Following allows a robot to follow a pre-specified path to a given location, relying upon the visual tracking of features (landmarks) [30, 47]. Here the approach is highly accurate but only suitable for local navigation; therefore it should be seen as complementing our topological approach to navigation.

This observation of a *path distance/accuracy* trade off between long-distance/low-

precision and short-distance/high-accuracy mission segments plays an important role in finding efficient solutions to the overall robot navigation problem. Indeed, integrating these two approaches is a powerful approach that leads to an *overall system* which exhibits improved robustness, scalability and simplicity, with respect to traditional approaches. In the next Section, the results obtained using this integrated approach are presented.

### 4.6.1 Integrated Experiments

In this experiment global and local navigation tasks are integrated by combining the topological approach to navigation with Visual Path Following. The robot used was the TRC Labmate.

The mission started in the Computer Vision Lab. Visual Path Following was used to navigate inside the Lab, traverse the Lab's door and drive the robot out into the corridor. Once in the corridor, control was transferred to the topological navigation module, which drove the robot all the way to the end of the corridor.

At this position, a new behaviour was launched, consisting of the robot executing a 180° turn, after which the topological navigation mode drove the robot back to the Lab entry point. During this backward trajectory, we used the *same image eigenspaces* as during the forward motion by simply rotating, in real-time, the acquired omnidirectional images by 180°. Once again, the use of an omnidirectional camera proved highly advantageous.

Finally, and once the robot was approximately located at the lab entrance, control was passed to the Visual Path Following module. Immediately it located appropriate visual landmarks and drove the robot through the door. It followed a pre-specified path until the final goal position, well inside the lab, was reached.

Figure 4.9 shows an image sequence of the robot during this experiment. The total path traversed was approximately 34m. Figure 4.10 shows the robot trajectory during

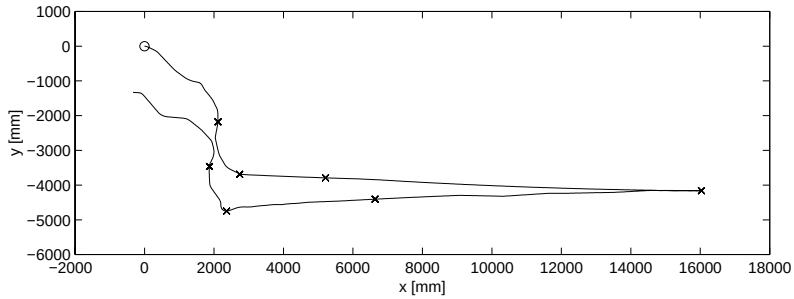
**Fig. 4.9**: A sequence of images of an experiment combining Visual Path Following for door traversal and topological navigation for corridor following.

the experiment, and its estimate using odometry. Interestingly, when returning to the laboratory, the uncertainty in odometry was approximately 0.5m. Thus, door traversal would not be possible without the use of visual control.

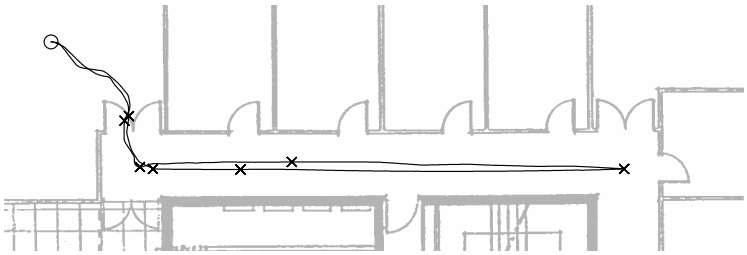
This integrated experiment shows the successful application of our methodology for navigating between distant environmental points and Visual Path Following for accurate path traversal. The resulting system can robustly solve various navigation problems and makes parsimonious use of the available computational resources.

## 4.7 Dealing with Large Illumination Changes

Our approach to building a topological representation of the environment works well in indoor environments, where the illumination can be relatively easily controlled. Unfortunately, if *large non-uniform* deviations in illumination occur (see Figure 4.11) as happens, for example, when a scene contains direct sunlight, the robot is prone to




(a)



(b)

**Fig. 4.10**: The experiment combining Visual Path Following for door traversal and topological navigation for travelling long distances. Trajectory estimate from (a) odometry and (b) the true trajectory.

miscalculating its location.

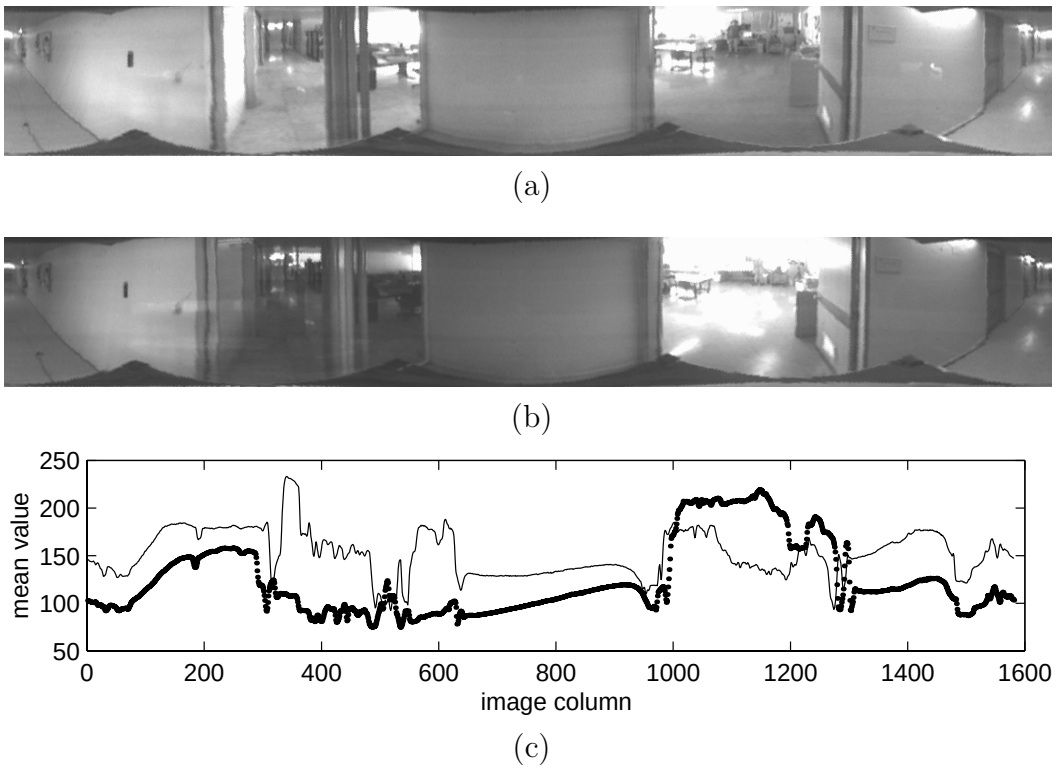



**Fig. 4.11**: Images acquired at (a) 5pm and (b) 11am. (c) Image intensity shows large non-uniform deviation in brightness. The thin line represents image (a).

This problem can be overcome by using edge images to represent the environment. Matching is achieved by using an eigenspace approximation to the *Hausdorff fraction* [66]. Previous research using this method concerned object recognition when faced with minor occlusions [65]. As an alternative implementation, one could build a larger learning set to represent the illumination variability. The disadvantage of this approach is the degradation in performance due to the extra cost of memory and computation.

### 4.7.1 The Hausdorff Distance

In this Section, we present a brief overview of the Hausdorff distance. The Hausdorff distance [118] (of which the Hausdorff fraction is a subset) is a technique whereby one can measure the distance between two sets of points, in our case edge images. A number of Hausdorff distance measures are defined by the following equations:

$$H(A, B) = \max(h(A, B), h(B, A)) \tag{4.2}$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} \parallel a - b \parallel \tag{4.3}$$

Here A and B represent sets of points. $h(A, B)$ measures the distance from each point in $A$ to the nearest point in $B$. These distances are ranked and the largest distance is termed the *directed* distance from $A$ to $B$. In the same manner, $h(B, A)$ measures the *undirected* distance from $B$ to $A$. $H(A, B)$ is the maximum of both. Unfortunately, the general Hausdorff distance $H(A, B)$ is highly sensitive to outlining points and in practice, the max in Equation 4.3 is replaced by a quantile:

$$h_k(A, B) = k^{th}_{a \in A} \min_{b \in B} \parallel a - b \parallel \tag{4.4}$$

where $0 < k \leq 1$. For example, the maximum is defined by the $1^{st}$ quantile and the median by the $\frac{1}{2}^{th}$ quantile.

Now, let us suppose that we wish to determine the fraction of points in $E$, where $0 < E \leq 1$ that are within a distance, $d$ of points in $F$, where $0 < F \leq 1$. This is termed the *Hausdorff fraction*, $h_f$:

$$h_f(E, F) = \frac{E_n \wedge F^d_m}{E_n} \tag{4.5}$$

where $E_n$ and $F_m$ are the number of points in $E$ and $F$, respectively. Here the points in $F$ are dilated by $d$. If they were not so, Equation 4.5 would simply represent normalised

correlation.

We now proceed to discuss the method used to build an eigenspace approximation to the Hausdorff fraction. This forms the environmental representation for our mobile robot in areas of large non-uniform illumination change.

**Eigenspace Approximation to the Hausdorff Fraction**

The eigenspace approximation [66] is built as follows: Let $I_m$ be an observed edge image and $I_n^d$ be an edge image from the topological map, arranged as a column vector. The *Hausdorff fraction*, $\hbar(I_m, I_n^d)$, which measures the similarity between these images, can be written as follows:

$$\hbar(I_m, I_n^d) = \frac{I_m^T I_n^d}{\|I_m\|^2}$$

An image, $I_k$ can be represented in a low-dimensional eigenspace by a coefficient vector, $\mathcal{C}_k = [c_1^k, \cdots, c_M^k]^T$, as follows:

$$c_j^k = e_j^T.(I_k - \bar{I}).$$

Here, $\bar{I}$ represents the average of all the intensity images and can be also used with edge images. Thus, the eigenspace approximation to the Hausdorff fraction can be efficiently computed as:

$$\widehat{\hbar}(I_m, I_n^d) = \frac{C_m^T C_n^d + I_m^T \bar{I} + I_n^{dT} \bar{I} - \|\bar{I}\|^2}{\|I_m\|^2}$$

### 4.7.2 Illumination Results

To test this eigenspace approximation we collected a sequence of images, acquired at different times, 11am and 5pm, near a large window. Importantly, this method was not required for the small changes in illumination which usually occur indoors. The image set acquired exhibited large non-uniform illumination change. Figure 4.12 shows the significant change in illumination, especially near the large window at the bottom left hand side of each omnidirectional image.

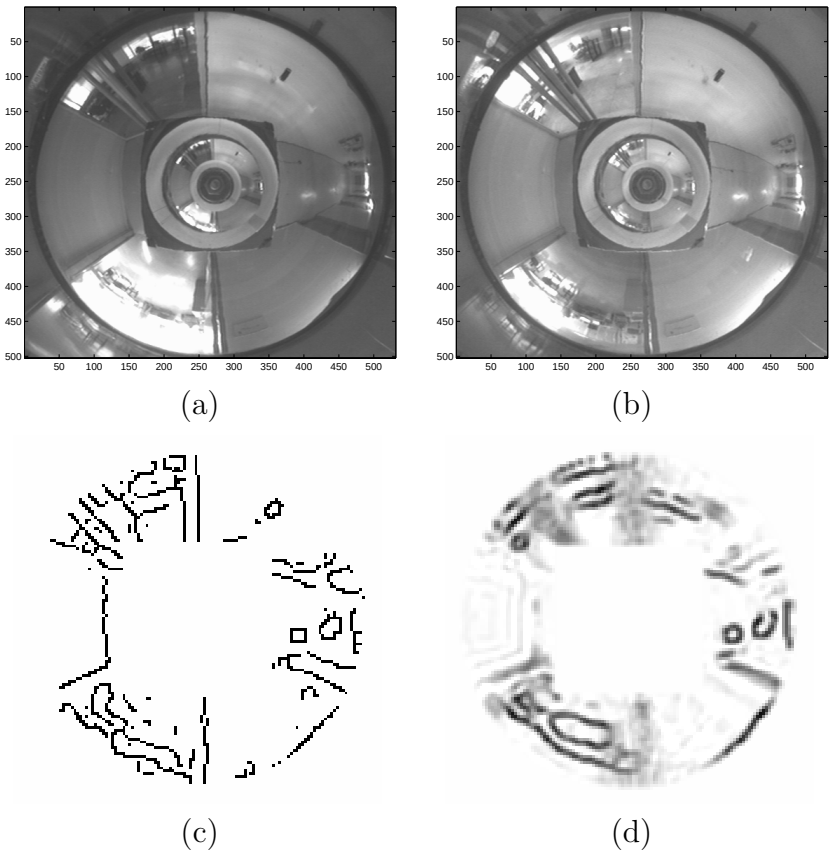



**Fig. 4.12**: (a) An omnidirectional image obtained at 11 am, (b) one obtained at 5 pm (c) An edge-detected image and (d) its retrieved image.

Even so, the eigenspace approximation correctly determined that the unknown image shown in Figure 4.12(a) was closest to the reference image shown in Figure 4.12(b), while PCA based on brightness distributions failed. For completeness, Figures 4.12 (c) and (d) show a run-time edge image and its corresponding retrieved image using the eigenspace approximation to the Hausdorff fraction.

Figure 4.13 shows the robot's qualitative position estimated over time, using both (a) gray-level distributions and (b) the Hausdorff fraction. Images for topological localisation were acquired at 11am and experiments were conducted at 12pm and 5pm. The $x$-axis represents traversal from a region of small illumination change to a region of large non-uniform illumination change. As shown in Figure 4.13(a), in the area of large illumination change, the robot miscalculated its position when relying upon intensity images. Figure 4.13(b) shows that only when using the Hausdorff fraction was qualitative localisation possible, independent of the changing illumination conditions.

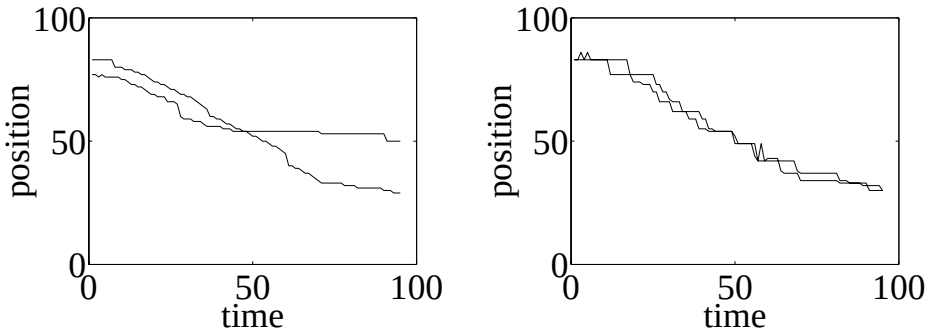



**Fig. 4.13**: Position estimation with large non-uniform illumination changes (a) using brightness distributions and (b) the Hausdorff fraction.

## 4.8 Summary

This chapter was concerned with the problem of vision-based mobile robot navigation. It built upon the topological environmental representation described in Chapter

3. From the outset of this work, the goal was to build a system which could solve the navigation problem by applying a holistic combination of omnidirectional vision, a topological environmental representation, appearance-based methods and visual servoing. This approach was shown to be successful.

We showed that by applying the eigenspace approximation to the Hausdorff fraction, when building the environmental representation, large non-uniform changes in illumination could be overcome.

Finally, results from an integrated experiment involving both global and very precise navigation were detailed. This approach relied upon the observation of a path distance/accuracy trade-off in order to robustly solve various navigation problems.

# Chapter 5

# Information Sampling

*This chapter presents an extension to the environmental representation component of our holistic approach to navigation. In order to handle the complexity of the perception process while increasing computational efficiency, our robot focuses its attention on the image data points, from the a priori set, which contain the most discriminating information. The discriminating pixels are selected by a statistical, non-featured based method termed Information Sampling. In particular, we show how to use the selected data for robot navigation. As a further extension of the Information Sampling method, object recognition results are detailed.*

## 5.1 Introduction

The last three chapters explained in detail our approach to vision-based robot navigation. This methodology required the synergistic combination of omnidirectional vision, a topological representation of the environment, appearance-based methods and visual servoing for local pose control.

As previously detailed in Chapter 3, an appropriate environmental representation is a key component of a successful navigation system. For the navigation experiments outlined in Section 4.5 (p. 92), **entire** omnidirectional or panoramic images were used

as a basis for the environmental representation and successful results were achieved.

Images usually convey a large amount of information and, as a consequence, the the perception and navigation systems must find ways of handling such a complex data flow. One way to overcome the complexity problem consists of focusing the system's attention (resources) upon the most relevant parts of the visual data. This chapter presents a method of achieving this by determining the pixels, from the *a priori* image set, which yield the most information, in terms of position estimation, about the environment. We term this approach **Information Sampling** [155, 159, 160]. Since we no longer require entire images, there is an obvious computational saving. Information Sampling was also applied to another problem domain within computer vision: object recognition, thus broadening the applicability of our work.

One can view Information Sampling as a "landmark" selection process, although *not* in the traditional sense. We shall back up this claim by using a simple example. A landmark is a significant, easily identifiable structure which contrasts greatly with the background. Therefore, one's attention is naturally drawn to it: a stop sign on a road, for example. Indoors, researchers have tended to use artificial or easily identifiable natural landmarks for localisation. Information Sampling approaches the problem from a different perspective. It relies upon the intensity changes between images from the *a priori* dataset. Therefore, a "landmark" is defined as a set of pixels which vary *significantly* from one image to the next, i.e. those which exhibit the most information change. For example, if a highly textured region is in all of the images, it is not a discriminating landmark and so shall not be chosen by the Information Sampling method. A major benefit of this approach is that it is non-feature based and so can be applied to images exhibiting low texture.

When navigating from one area to another, the mobile robot now has the ability to identify and focus upon *highly discriminating* regions within the environment. These are then periodically memorised for future reference, thus mimicking the approach to

navigation naturally adopted by humans and some animal species.

As an aside, we shall see how the same methodology can be applied to object recognition, by selecting the areas which can most easily distinguish between a set of objects.

## 5.2 The Information Sampling Method

As previously noted, our approach requires the use of *a priori* image data. Theoretically, it can be applied on a *pixel-by-pixel basis,* ***independent*** *of image type.* For the navigation experiments outlined in this chapter, images were acquired from an omnidirectional camera with a spherical mirror. Once the images were captured, we determined which regions contained the most relevant information, i.e. which were the most discriminating for position estimation, by applying Information Sampling. Significantly, our approach is *non-feature based* and was applied to images of low texture: the environment consisted of plain white walls and brown doors. As a first step in explaining the Information Sampling method, Section 5.2.1 outlines the procedure for reconstructing an image, given only a small amount of data.

### 5.2.1 Image Reconstruction

We assume that the images captured by the robot's camera can be modelled as a random vector $I$, characterised by a Gaussian distribution with mean $\bar{I}$ and covariance $\Sigma_I$:

$$I \sim \mathcal{N}(\bar{I}, \Sigma_I) = p(I)$$

Usually, one can take an ensemble of images of the environment $[I_1 \dots I_m]$, which can be utilised for computing $\bar{I}$ and $\Sigma_I$, so that $p(I)$ can be computed *a priori.* When the robot is navigating, we assume that the observations, $d$, consist of a selection of (noisy) image pixels (or sub-regions), rather than the entire image. Accordingly, the

observation model can be expressed as:

$$d = SI + \eta \tag{5.1}$$

where $d$ stands for the observed data and the measurement noise, $\eta$ is assumed to follow a Gaussian distribution with zero mean and covariance, $\Sigma_n$. We further assume that $I$ and $\eta$ are independent. The selection matrix, $S$ is composed of a series of ones and zeros, the ones correspond to the data points extracted from an image. We select a number of pixels to test by moving the set of ones in the selection matrix.

Having prior knowledge of $I$, in the form of a statistical distribution, $p(I)$, the problem now consists of estimating the (entire) image based on partial (noisy) observations of a few pixels, $d \in \mathbb{R}^n$. This problem can be formulated as a *Maximum a Posteriori* estimation of $I$. The posterior probability can be determined from Bayes rule as follows:

$$p(I|d) = \frac{p(d|I)p(I)}{p(d)} \tag{5.2}$$

where $p(d|I)$ is the likelihood of the observation of a pixel value (or set of pixels) given a known image, $I$. From Equation (5.1) we can determine $p(d|I)$ as follows:

$$p(d|I) = \frac{1}{2\pi^{\frac{n}{2}}\sqrt{\det(\Sigma_n)}} \exp[-\frac{1}{2}(d - SI)^T \Sigma_n^{-1}(d - SI)] \tag{5.3}$$

The prior distribution, $p(I)$ is assumed to have been learnt *a priori* and is given by the following Equation:

$$p(I) = \frac{1}{2\pi^{\frac{n}{2}}\sqrt{\det(\Sigma_I)}} \exp[-\frac{1}{2}(I - \bar{I})^T \Sigma_I^{-1}(I - \bar{I})] \tag{5.4}$$

Finally, $p(d)$ is given by:

$$p(d) = \frac{1}{2\pi^{\frac{n}{2}}\sqrt{\det(S\Sigma_I S^T + \Sigma_n)}} \exp[-\frac{1}{2}(d - S\bar{I})^T (S\Sigma_I S^T + \Sigma_n)^{-1}(d - S\bar{I})] \tag{5.5}$$

Now, taking $\mathcal{C}(I) = -\ln(p(I|d))$ and removing all terms that do not depend on $I$, yields the following equation:

$$\mathcal{C}(I) \propto [(I - \bar{I})^T \Sigma_I^{-1}(I - \bar{I}) + (d - SI)^T \Sigma_n^{-1}(d - SI)] \tag{5.6}$$

Hence, maximising the posterior probability is equivalent to minimising the criterion $\mathcal{C}(I)$. Therefore, from Equation (5.6), we can compute the maximum a posteriori estimate of $I$ [119] as follows:

$$\hat{I}_{MAP} = \arg\max_{I} p(I|d) = (\Sigma_I^{-1} + S^T\Sigma_n^{-1}S)^{-1}(\Sigma_I^{-1}\bar{I} + S^T\Sigma_n^{-1}d) \tag{5.7}$$

Thus, $\hat{I}_{MAP}$ is the reconstructed image obtained using the pixel (or set of pixels), $d$. Notice that by combining the prior image distribution with the statistical observation model, we can estimate the entire image based on the observation of a limited number of pixels.

### 5.2.2 Choosing the Best Data: Information Windows

Once we have reconstructed an image using the selected data, we can compute the error associated with this reconstruction. From Equation (5.7), the error covariance matrix, $\Sigma_{error}$ is given by:

$$\Sigma_{error} = \mathrm{Cov}(I - \hat{I}_{MAP}) = (\Sigma_I^{-1} + S^T\Sigma_n^{-1}S)^{-1} \tag{5.8}$$

Of course, the quality of the estimate and the "size" of $\Sigma_{error}$ depend not only on the observation noise, $\eta$ but also on the observed image pixels, as described by the selection matrix, $S$. Equation (5.8) quantifies the quality of an estimate obtained using a particular set of image pixels. In theory, we can evaluate the *information content* of any individual image pixel or combination of pixels, simply by selecting an appropriate selection matrix, $S$, and determining the associated $\Sigma_{error}$.

This problem can be formulated as an experiment design process [119], in which we look for the optimal selection matrix, $S^*$ that minimises (in some sense) the error covariance matrix. Taking the determinant of $\Sigma_{error}$ as an indication of the "size" of the error, the optimal selection of image pixels is given by:

$$S^* = \arg\min_{S}\{\ det((\Sigma_I^{-1} + S^T\Sigma_n^{-1}S)^{-1})\ \} \tag{5.9}$$

In practice, to avoid computing the inverse we define the following equivalent optimisation problem in terms of a modified uncertainty metric, $U$:

$$U = -\log\{\, det(\Sigma_I^{-1} + S^T \Sigma_n^{-1} S)\,\}\,; \quad S^* = \arg\min_S U \tag{5.10}$$

So far, we have described Information Sampling as a process for $(i)$ reconstructing an entire image from the observation of a few (noisy) pixels and $(ii)$ determining the *most relevant* image pixels, $S^*$.

Unfortunately, determining $S^*$ is computationally impractical since we would have to compute $\Sigma_{error}$ for all possible combinations of pixels scattered throughout the image[1]. Instead, we partition the image into non-overlapping square windows of $(l \times l)$ pixels. We term these regions *Information Windows*, denoted by $\mathbf{w} = [w_1 \ldots w_n]$.

By using Equation (5.10), we can rank Information Windows or combinations of such windows, in terms of their information content. Again, as searching for all possible combinations of windows within the image, in order to minimise Equation (5.10), would be computationally intensive, we instead use two sub-optimal (greedy) algorithms. These algorithms are described in Section 5.3.

Again, we reiterate that the information criterion is based on the entire set of images and not, as with other methods, on an image-by-image basis. For instance, a highly textured image region is defined as a good "landmark" only if it varies significantly from one image to the next. In other words, data common to all images have a large reconstruction error and so are unreliable for position estimation. On the other hand, data which varies throughout the image set are associated, at any instant, with a single image and so allow for reliable position estimation.

[1]If we wished to compute all possible subsets of $m$ pixels from $n$, the number of pixels in an image, then the number of subsets = binomial(m n) with $m \sim 10^6$ and $n \sim 10^2$.

## 5.3 Ranking the Information Windows

Information Sampling was applied to two sets of omnidirectional images acquired by the Labmate mobile robot in an indoor office environment. We show that by using Information Sampling to focus upon attentive regions within the environment, effective navigation is possible. Real-world experiments verify this thesis.

An image set, consisting of 89 omnidirectional images, was acquired every 10cm in an indoor environment. Each image was acquired at a resolution of $768 \times 576$ pixels, low pass filtered[2] and subsampled to an image resolution of $128 \times 128$ pixels. These formed the database set, $\mathbf{T_{128}}$. In order to perform Information Sampling these images were further low pass filtered and subsampled to a set of images, $\mathbf{T_{32}}$, $32 \times 32$ pixels in size. The reason for such a small image size relates to the complexity of determining the error covariance matrix, $\Sigma_{error}$ in Equation (5.8). To find the most discriminating Information Windows, over $\mathbf{T_{32}}$, we found and ranked the 16 *non-overlapping* windows of size $8 \times 8$ pixels. We then calculated the equivalent $32 \times 32$ Information Windows (extracted from the $128 \times 128$ training images, $\mathbf{T_{128}}$) to the $8 \times 8$ windows (extracted from the $32 \times 32$ images, $\mathbf{T_{32}}$). For the initial set of experiments, the $32 \times 32$ size windows were used.

Additionally, we improved the process by finding and ranking the 225 *overlapping* windows of size $4 \times 4$ pixels in $\mathbf{T_{32}}$. The overlapping windows were generated by shifting each window in the horizontal and vertical direction by 2 pixels, thus generating an overlap of 50%. We then calculated the equivalent $16 \times 16$ Information Windows (extracted from the $128 \times 128$ training images, $\mathbf{T_{128}}$) to the $4 \times 4$ windows (extracted from the $32 \times 32$ images, $\mathbf{T_{32}}$).

[2] An average filter was used.

### 5.3.1 Searching for the Best Information

As previously noted, finding the set of pixels to select (from the *a priori* set of images) as the best information is a highly computationally intensive problem. In order to overcome this problem, we implemented two greedy search algorithms: *Combinatorial Search* and *Simple Search*. These were used to find and rank the best Information Windows.

#### Combinatorial Search

We first search for the best Information Window. Then, the search for the next best window is made keeping the first window *fixed*, thus locating the best *pair* of windows. As the method continues it determines the best triplet of windows, etc. If we denote $n$ as the number of windows within an image, this method requires the evaluation of Equation (5.10), $n!$ times. The method automatically groups the Information Window(s) into a single window, a pair of windows, a triplet of windows etc. Notice that this method is not a true combinatorial search, which would require the evaluation of *all* possible combinations of windows.

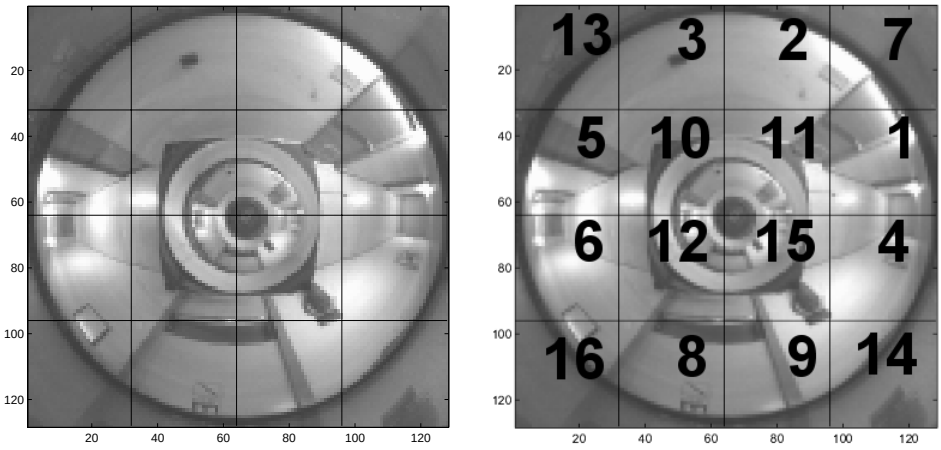
#### Simple Search

This is a faster search algorithm. We rank each of the information windows *independently*. In this case, Equation (5.10) has only to be evaluated $n$ times. As distinct from Combinatorial Search, if we wish to group the best (single, pair, triplet etc. of) window(s) we must do it manually based on the initial ranking.

### 5.3.2 Ranking Results

Figure 5.1(a) shows the Information Windows available for selection and Figure 5.1(b) these Information Windows, *individually ranked* from the most (number 1) to the least

discriminating (number 16) using Simple Search. Figure 5.2 shows the ranking when using panoramic images.




(a) (b)

**Fig. 5.1**: Ranking Results: (a) The 16 non-overlapping Information Windows. (b) Those windows ranked, according to the amount of information they contain, using Simple Search.

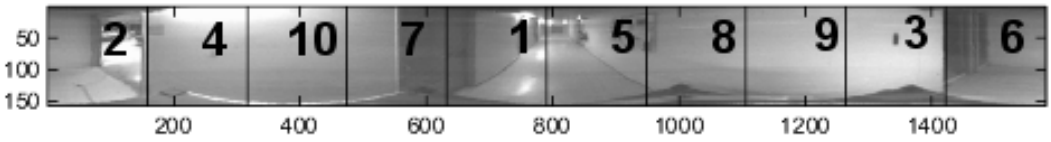



**Fig. 5.2**: The information windows obtained using panoramic images, ranked, according to the amount of information they contain, using Simple Search.

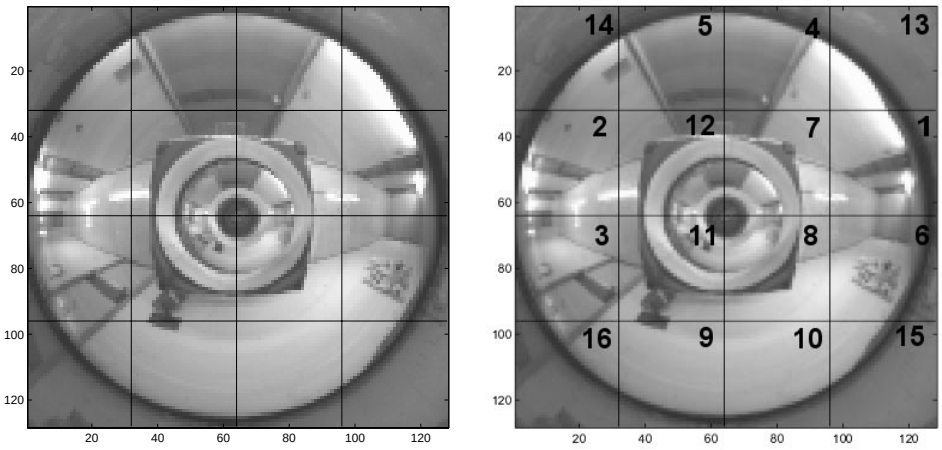
The following example provides an intuitive idea of the Information Sampling method. All of the omnidirectional images in this dissertation show the robot in the centre of each image[3]. Any Information Window which contains the robot is not a discriminating one and so it follows that such a window should have a relatively low ranking. As shown in Figure 5.1(b), this proves to be the case: the four Information Windows which contain the robot are ranked from numbers ten to fifteen. Additionally,

[3]All omnidirectional images acquired by the catadioptric sensor with a spherical mirror show the robot. The mirror of the SVAVISCA sensor was designed so as this was not the case.

the four windows at the periphery of the image also have a low ranking, since they only contain a portion of the omnidirectional image. It should be noted that the corridor in which the *a priori* set of images were acquired had a number of offices on one side (the top half of the omnidirectional images) and only a single door and notice-board on the other (the bottom half of the omnidirectional images). Thus, as the robot travels down the corridor more information change occurs in the top half of the omnidirectional images. Again, this is borne out by the window ranking, where the three highest ranking Information Windows are all in the top half of the omnidirectional image.

A second set of 53 (as opposed to 89) omnidirectional images, obtained every 20cm (as opposed to 10cm) were acquired in the same corridor but from a different starting position. Figure 5.3(a) shows the Information Windows available for selection and Figure 5.3(b) the window ranking when using *non-overlapping* windows. The four Information Windows which contain the robot are ranked from numbers seven to twelve. Additionally, the four windows at the periphery of the image have the lowest ranking. Significantly, the **same** highest ranking Information Window (window 8) is selected from both sets. The other Information Windows are ranked in a different order, but note that, Information Sampling chooses the same six top ranking windows from both sets.

To further test the technique, we found and ranked the 225 *overlapping* windows of $16 \times 16$ pixels in size. The advantage of this approach is that we can focus upon smaller areas of the omnidirectional image. Naturally, relevant information contained within the image received a high ranking and, advantageously, portions of the image which were close to the background, and relevant, were highly ranked. As expected the dark background received the lowest ranking. An image of the 10 best overlapping Information Windows is shown in Figure 5.4.




(a) (b)

**Fig. 5.3**: Ranking Results: (a) The 16 non-overlapping Information Windows. (b) These windows ranked, according to the amount of information they contain, using Simple Search.

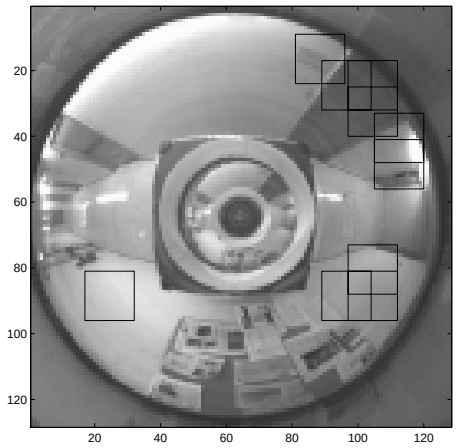



**Fig. 5.4**: The 10 best overlapping Information Windows.

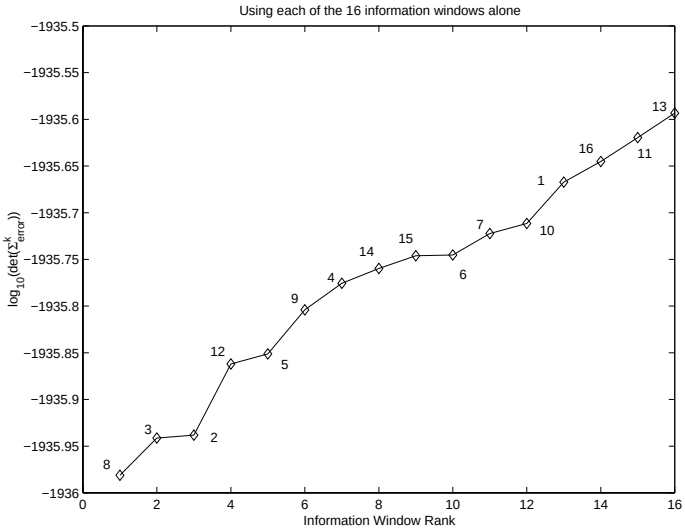
### Graphing the Information Content

Figure 5.5 shows the graphs of the Information Windows, obtained from omnidirectional images, ranked using (a) Simple Search and (b) Combinatorial Search. In both cases, the $x$-axis corresponds to the window ranking, from first to sixteenth and the $y$-axis corresponds to the uncertainty metric, $U$, calculated using Equation (5.10) (p. 109). The numbers along the graph line correspond to the 16 non-overlapping Information Windows per omnidirectional image. For example, using Simple Search, Figure 5.5(a) tells us that the eighth Information Window exhibits the lowest uncertainty value and so is individually ranked in first position, while the third window, having a higher uncertainty value, is individually ranked in second position etc.

Using Combinatorial Search, Figure 5.5(b) tells us that the eighth window is ranked in first position. This window is then fixed and the best *pair* of windows, in this case the eighth plus the third, are found. Thus, the third window contains the next best amount of information and is ranked in second position. Using Combinatorial Search the next best window added at each stage matches the window rank chosen by Simple Search.

Combinatorial Search continues until all windows have been combined. As can be seen from Figure 5.5 (b) each *combination of Information Windows* exhibits a lower uncertainty measure than the previous one. Intuitively, this makes sense as the more information available, the better the image reconstruction (see Equation 5.7) should be. However, the payoff for using many Information Windows is not significant, as can be seen from the small drop in uncertainty. This result is also borne out by Figure 5.7, as detailed in Section 5.3.3 (p. 117). Clearly, the fact that the highest ranking Information Window is not only the most relevant, but is the most relevant by a significant factor, is the reason why we need use only it for reconstruction.

In terms of computation time, Simple Search took an average of 6.2 seconds to rank the Information Windows while Combinatorial Search took an average of 63.9 seconds




(a)

(b)

**Fig. 5.5**: Graphs of the data contained in each Information Window versus the window rank when using (a) Simple Search and (b) Combinatorial Search. The numbers along the graph line are the window numbers.

to determine the same information. The trade-off is accuracy versus computational power.

Figure 5.6 shows the ranking results when using (a) non-overlapping and (c) overlapping windows from the second set of images. Again, when using (a) non-overlapping Information Windows, the eighth exhibits the lowest uncertainty value and so is ranked in first position. When using (c) overlapping windows, window number 74 is ranked in first position. Figure 5.6 shows (b) the best non-overlapping and (d) overlapping Information Windows in an image.

### 5.3.3 Reconstruction Results

The reconstruction results obtained using Information Windows of size $8 \times 8$ pixels and omnidirectional images of $32 \times 32$ pixels in size are shown in Figure 5.7. Figure 5.7(a) shows an omnidirectional image from the *a priori* set, Figure 5.7(b) its reconstruction using only the *most discriminating* $8 \times 8$ Information Window and 5.7(c) its reconstruction using all of the Information Windows. Reconstruction was achieved using Equation (5.7). As can be seen from the images, a good reconstruction is obtained using only the best Information Window. This is an indication of the power of Information Sampling.

## 5.4 Information Sampling for Robot Navigation

In Chapter 4, we presented our vision-based mobile robot navigation results obtained when using entire omnidirectional images. As detailed in Section 1.2.2, previous appearance-based approaches also used entire images for position estimation. In contrast to other areas of computer vision, (tracking, for example) no attempt was made to select informative regions from the images. In Section 5.2, we outlined our approach to this problem.

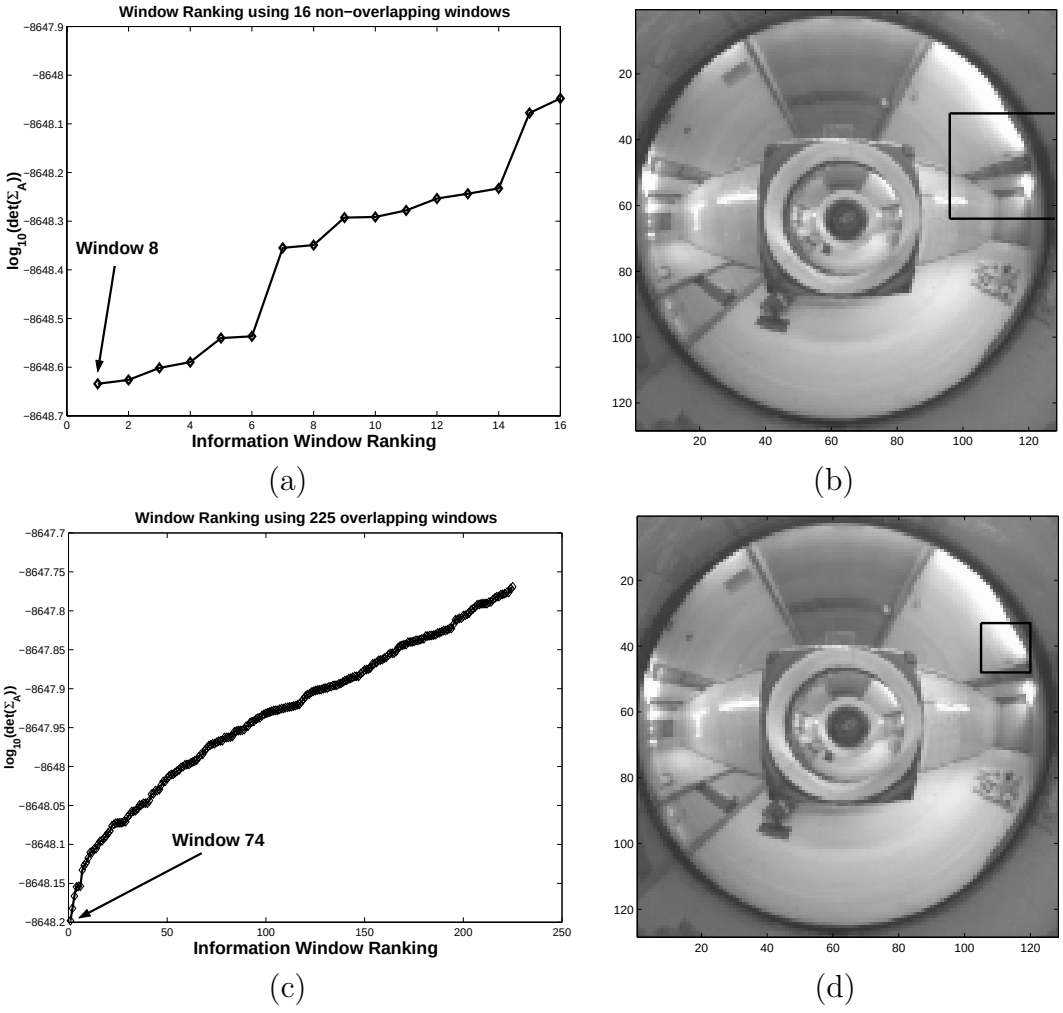
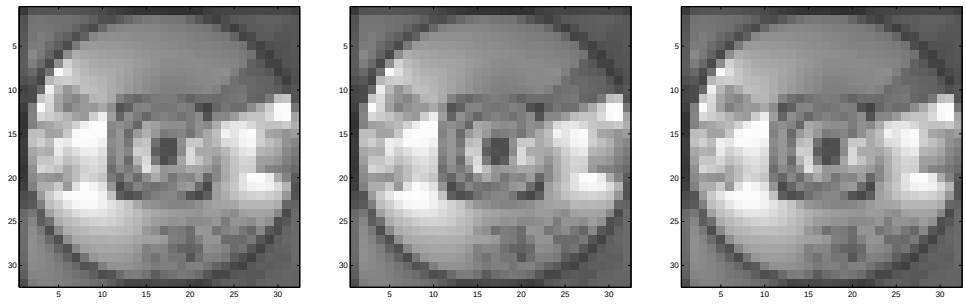



**Fig. 5.6**: Graphs of the information contained in each Information Window versus the window rank using (a) non-overlapping and (c) overlapping windows. The best (b) non-overlapping and (d) overlapping Information Window in an image.



(a) (b) (c)

**Fig. 5.7**: (a) A $32 \times 32$ omnidirectional image acquired at run-time. (b) Its reconstruction using the *most discriminating* Information Window. (c) Its reconstruction using all of the Information Windows. Each Information Window is $8 \times 8$ pixels in size.

**Building a Local Appearance Space Using Information Sampling** Information Sampling selects the most informative data from an *a priori* image set. This was defined as the most discriminating $32 \times 32$ non-overlapping Information Window, i.e. just 6.25% of the original amount of information. When applied to the second image set, we used the 10 most discriminating $16 \times 16$ overlapping Information Windows, i.e. just 10.93% of the original amount of information.

Conceivably, navigation could be achieved by matching the reconstructed image, $\hat{I}_{MAP}$ to the set of omnidirectional images, although this would be computationally expensive. Thus, as before, this matching is achieved in real-time by projection into a low-dimensional eigenspace built using Principal Component Analysis (PCA). In this case, the key difference (i.e. the extension to the environmental representation component of our holistic approach to navigation) is that we use *only* the most discriminating information obtained by *Information Sampling* to build a **local appearance space** (low-dimensional eigenspace). In this way, we directly project the best information, significantly reducing the number of projected windows and therefore, the level of possible ambiguity. The local appearance space has an orthonormal basis of eigenvectors of size $(l^2 \times 1)$, where $l$ is the length of the side of a square Information Window. It is our topological environmental representation.

### 5.4.1 Navigation Results

So far we have outlined our procedure to:

1. Select the best Information Window using Information Sampling, thus focusing the robot's attention on the most discriminating information.

2. Build a local appearance space using only the best Information Windows from each image in the *a priori* set.

As a first test of the method, we ran our Labmate mobile robot in a corridor environment. *Only* the best Information Window, from each image, was projected into the local appearance space. The images in Figure 5.8(a), (b) and (c) show the results obtained using windows of $32 \times 32$ pixels in size. The top row shows (a) the most relevant Information Window from an unknown image, (b) its closest match from the *a priori* set of best Information Windows and (c) its reconstruction using PCA. Figure 5.8 shows (d) the best Information Window in the unknown $128 \times 128$ image and (e) its closest match from the *a priori* set, obtained by projecting only the most relevant Information Window. We note here that we could in principal, given enough computing power, use Equation (5.7) to reconstruct a $128 \times 128$ *image* using only the most relevant *window*.

To further test the applicability of the Information Sampling technique, we ran three more position estimation experiments along a corridor. The first experiment used entire $128 \times 128$ images for matching, the second the most informative *non-overlapping* $32 \times 32$ Information Window and the third the 10 most informative *overlapping* $16 \times 16$ Information Windows.

Figure 5.9 shows (a) an unknown image, (b) its closest match from the *a priori* set of entire images and (c) the distance travelled ($\sim$10.5m) by the robot under closed loop control. Figures 5.10 and 5.11, respectively, show navigation experiments using *non-overlapping* and *overlapping* Information Windows. Significantly, in these latter

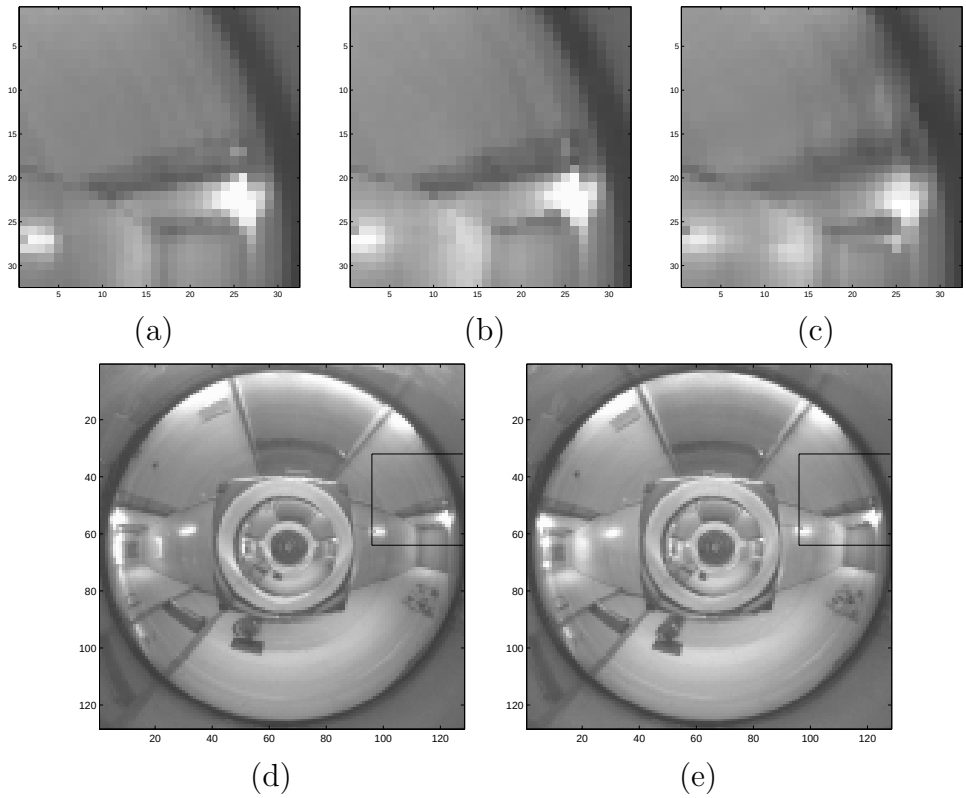



**Fig. 5.8**: Close-up of the $32 \times 32$ Information Windows from Set A: (a) unknown (b) closest and (c) reconstructed. The position of (d) the unknown and (e) the closest images in their respective omnidirectional images.

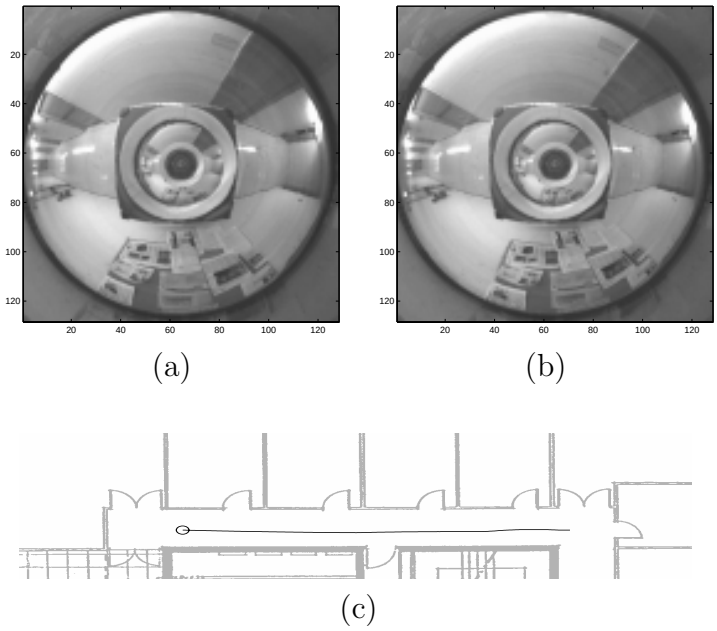



**Fig. 5.9**: (a) An unknown image, (b) its closest match and (c) the path travelled by the robot when using entire $128 \times 128$ images.

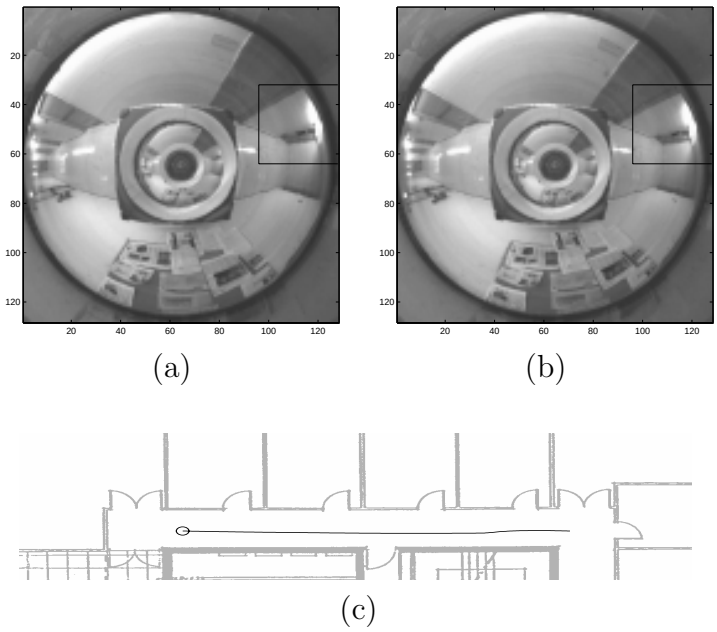



**Fig. 5.10**: a) An unknown image, (b) its closest match and (c) the path travelled by the robot when using the best $32 \times 32$ non-overlapping Information Window.

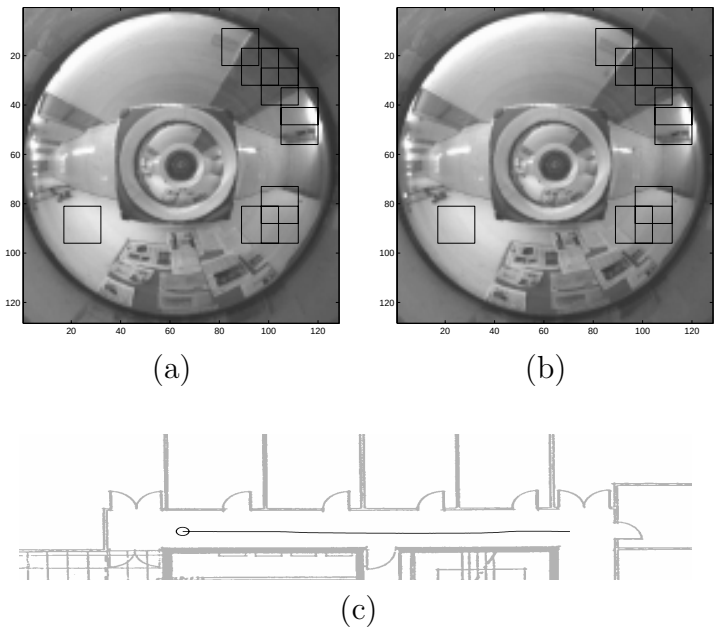


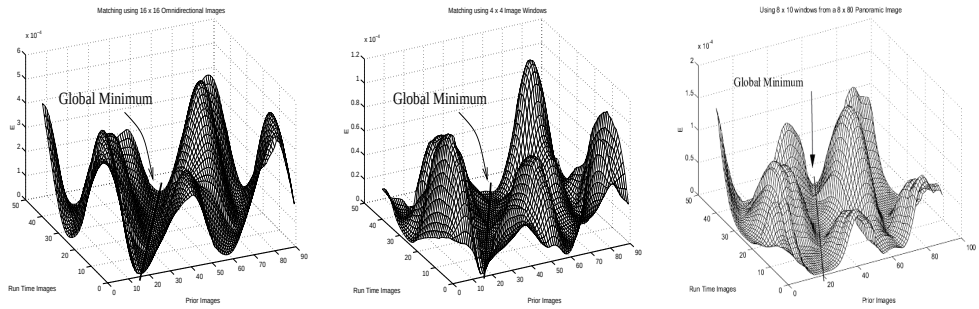

**Fig. 5.11**: a) An unknown image, (b) its closest match and (c) the path travelled by the robot when using the 10 best $16 \times 16$ overlapping Information Windows.

cases the number of pixels used for position estimation was 6.25% of the total, when using non-overlapping windows, and 10.93% in the case of overlapping windows, thus allowing the robot to maximise the use of its limited computational resources. Clearly, vision-based navigation in a corridor environment was successfully completed in each case.

### 5.4.2 Navigation Results using Low Resolution Images

As a final test of our holistic approach to navigation, we undertook preliminary experiments using low resolution images. Use of these may prove beneficial for large scale experiments, although our experiments were performed over a short distance ($\sim$7m) in a single corridor environment. The input data for three experiments consisted of (*i*) low resolution $16 \times 16$ omnidirectional images, (*ii*) the best $4 \times 4$ information window, extracted from the $16 \times 16$ omnidirectional images and (*iii*) the best $8 \times 10$ infor-

mation window, extracted from $8 \times 80$ panoramic images, respectively. Figure 5.12 shows the results obtained. Each graph shows the distance, $d_k^2$, between the prior and run-time images. While the local minima are different, significantly the global minimum is maintained in each case: the robot can determine a qualitative estimate of its position. Although requiring further testing and validation, our results show that navigation can be undertaken using extremely limited amounts of data, only 16 pixels representing 6.25% of the original 256 pixels. This allows the robot to concentrate its resources on other tasks.




(a) (b) (c)

**Fig. 5.12**: Graphs showing images acquired at run-time versus those acquired *a priori* when using (a) $16 \times 16$ Omnidirectional Images, (b) $4 \times 4$ Information Windows and (c) $8 \times 10$ Information Windows. Experiments were undertaken along a $\sim$7m path.

## 5.5 Object Recognition

In order to test further applications of the Information Sampling method, we applied it to the object recognition problem [158]. The object set used was the well known COIL-20 database [107] from Columbia University. A selection of images from the database are shown in Figure 5.13.

Each image is $128 \times 128$ pixels in size. Experiments were undertaken using 36 evenly spaced views of 20 objects as the database set and a different 36 evenly spaced views of the same 20 objects as the test set.


**Fig. 5.13**: A selection of images from the COIL-20 database.

We ran our Information Sampling method on the COIL-20 database, to determine the most discriminating Information Windows. Due to computational constraints, each $128 \times 128$ image was first subsampled to $32 \times 32$ pixels in size. Each Information Window was chosen to be $8 \times 8$ pixels in size, thus giving 16 non-overlapping Information Windows per image, ordered from left-to-right and top-to-bottom. Once the Information Windows were ranked, corresponding $32 \times 32$ Information Windows in the $128 \times 128$ sized images were found. For our navigation experiments only data corresponding to the best Information Window were used to build a local appearance space. In the case of object recognition, six local appearance spaces were built, each termed an *Informative Local Appearance Space (ILAS)*. They are numbered from one (the most informative) to six (the least informative), i.e. they correspond to the six highest ranking Information Windows. Recognition was achieved by using the first Information Window (only 6.25% of the pixels in an image) projected into the associated 10D local appearance space, ILAS 1. If the best window was occluded, or had a significant amount of non-uniform background change, recognition was achieved by jumping to the *next best* local appearance space and so on. Results are presented in Section 5.5.2 (p. 128).

Figure 5.14 shows six objects from the database in a number of differing poses along with their associated most (shown mid-image) and least (shown at the bottom-right of the images) discriminating Information Windows, as yielded by the Information Sampling method. Notice that each Information Window discriminates over the entire set of images, not on an image-by-image basis.

### 5.5.1 Matching Results

Object recognition and pose estimation experiments were first undertaken on unperturbed images using only ILAS 1, i.e. the highest ranking appearance space. The images in Figure 5.15 show the results obtained. Here, Figures 5.15 (a) and (d) show

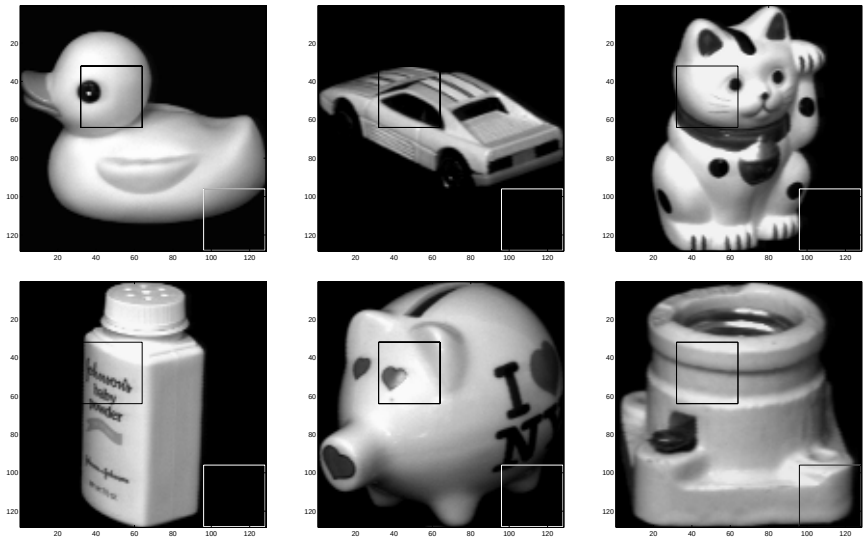


**Fig. 5.14**: A selection of images showing the highest (mid-image) and lowest ranking (bottom-right) Information Windows, respectively in a selection of images.

the best Information Window, extracted from two *unknown* objects (Figures 5.15 (b) and (e)) which we wish to recognise. Figures 5.15 (c) and (f) show the closest match, from the database set, in the correct pose. Results obtained using a large set of 720 unknown images reveal that the correct object was determined in 95.3% of cases and the correct pose in 73.8% of cases. Thus, the recognition results compare very favourably to those using entire images but utilise significantly less image data. In order to test the discriminating power of each Information Window we compared matching results using the $1^{st}$ and $3^{rd}$ most discriminating Information Windows. In this case, ILAS 3 yielded a correct object recognition rate of 82.5% and a pose estimation rate of 65.3%. Thus, as expected the discriminating power of ILAS 1 is superior. Naturally, the lower the ILAS level, the more our approach degrades.

Importantly, if regions, other than the Information Window used for recognition, were occluded the recognition results did not deteriorate. On the other hand, if an Information Window was occluded then the method outlined in the next Section was used to overcome the problem.

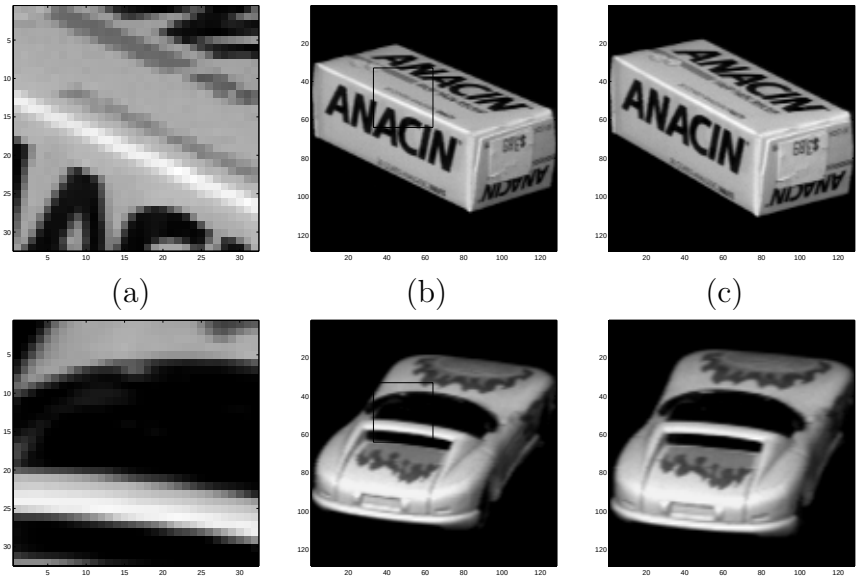



**Fig. 5.15**: Object recognition and pose estimation without background variation. When using the *most* discriminating Information Window, the object recognition rate was 95.3% and the pose estimation rate 73.8%.

### 5.5.2 Results: Non-Uniform Background Change

As a further test of our method we decided to run it on images with *non-uniform* background variation. This is a particularly difficult problem, as PCA is well known to be susceptible to such changes. Since an Information Window may contain some background data or may be partially occluded, we wish to minimise the effect of such aberrations. Thus, we added an additional step to our method. Once we determined each Information Window, *we subdivided it into* 16 *sub-regions*. These sub-regions were then used to build each Informative Local Appearance Space. Background variation was dealt with by associating a confidence level to each Information Window. If a high percentage of the sub-regions identify the same object we trust the result. If this is not the case, i.e. if most of the sub-regions fall on the background region and not on the object itself, then object recognition can be achieved by jumping to the next ILAS, and repeating the process. This is shown in Figure 5.16, where the most

| Image Change | Correct | False Positive | No ID(6) |
|---|---|---|---|
| Unperturbed (ILAS 1) | 95.3% | 4.7% | - |
| Unperturbed (ILAS 3) | 82.5% | 17.5% | - |
| Non-Uniform Background Change | 87.6% | 5.5% | 6.9% |

**Table 5.1**: Object Recognition Results Summary.

| Image Change | Correct | False Positive |
|---|---|---|
| Unperturbed (ILAS 1) | 73.8% | 16.2% |
| Unperturbed (ILAS 3) | 65.3% | 34.7% |
| Non-Uniform Background Change | 50% | - |

**Table 5.2**: Pose Estimation Results Summary.

discriminating Information Window is identified as containing a large amount of non-uniform background variation. In this case, object recognition and pose estimation were successfully achieved using ILAS 3. For Information Windows with less background variation, jumping to the next ILAS was not necessary. Using 612 test cases and the first six Information Windows, the correct object was identified in 87.6% of cases, with a false positive rate of 5.5%. An object was unidentifiable in 6.9% of cases. Tables 5.1 and 5.2 summarise the results obtained.

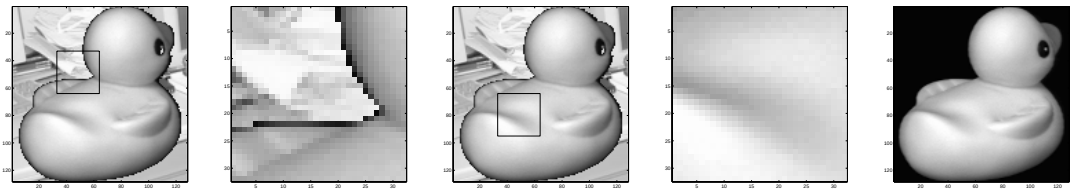


**Fig. 5.16**: Object recognition and pose estimation with non-uniform background variation.

## 5.6 Summary

This chapter presented improvements to the topological navigation aspect of our holistic approach to mobile robot navigation. Information Sampling was detailed as a method to select the most discriminating information from an *a priori* image set. This allowed the mobile robot to effectively use its (limited) computational resources. We showed how to use the data selected by Information Sampling for successful navigation and associated experimental results were presented. Furthermore, the applicability of Information Sampling was broadened by applying it to the domain of object recognition.

# Chapter 6

# Conclusion

*This chapter concludes the dissertation. The main contributions and results obtained are summarised. Finally, possible future avenues of research are suggested.*

## 6.1 Dissertation Summary

This dissertation presented a novel holistic methodology for vision-based robot navigation using an omnidirectional camera. One of the key observations is that successful navigation systems should result from the synergistic combination of a set of fundamental principals:

- A suitable camera geometry.
- An appropriate environmental representation which relates closely to a method for global (qualitative) localisation and a means of local pose control.
- An attention mechanism for handling the complexity of the perception process, thus allowing for an efficient use of (limited) computational resources.

The way in which these fundamental aspects have been solved in nature explains both, the diversity of (specialised) solutions encountered in biological navigation sys-

tems, as well as their performance, even when using relatively modest resources. Our approach focused on these various aspects, proposed a number of promising solutions and demonstrated how their integration may lead to simple, but yet, flexible and robust navigation systems.

We used an omnidirectional camera to sense the environment. Two designs were presented: (*i*) a conventional camera, pointed upwards, viewing a spherical mirror and (*ii*) a log-polar camera viewing a constant vertical resolution mirror. Neither of these systems exhibited a single centre of projection but this did not limit the applicability of our approach. Indeed, the opposite was true: by dropping the constraint we were able to utilise systems which broaden the usefulness of an omnidirectional camera. In addition, calibrating such systems was easier than if we had used a hyperbolic mirror, for example. The simplicity and increased reliability of these designs proved highly advantageous.

An important contribution of this work was to define an *appropriate* environmental representation as a key element of a robot's ability to navigate. In many previous works this aspect was often overlooked. We argued that the emphasis should be placed on building appropriate representations rather than always relying upon highly accurate information about the environment. Therefore, given the fact that our robot was designed to travel long distances within the environment, we chose a *topological* representation. The decision to use this representation was partly inspired by the way in which humans and animals model spatial knowledge.

This topological representation was required to meet the criteria that it: (a) was easy to build, (b) utilised a small amount of memory and (c) could be used for real-time localisation. The robot acquired omnidirectional images of the environment which were then used to build a low-dimensional eigenspace representation, obtained via Principal Component Analysis. For the navigation results provided in this dissertation, eigenspace matching proved effective for qualitative localisation.

We noted that sole reliance upon the representation does not solve the navigation problem. Thus, another of the strengths of our holistic approach lies in the fact that we presented a method whereby the environmental representation could be easily combined with a local control strategy. This allowed for our robot to make effective use of the environmental representation to significantly aid its ability to navigate. Control was visually based and the robot maintained its position and orientation as it travelled through the environment. Pose was controlled by visually servoing upon corridor guidelines, extracted from *bird's-eye views* of the ground plane.

We showed that the above holistic approach to navigation, which combined omnidirectional vision, a topological environmental representation, appearance-based methods and visual servoing achieved successful navigation in structured environments. In order to undertake both global (qualitative) and local (precise) navigation we presented results from a large-scale experiment where the robot travelled from the Computer Vision Lab, traversed the door, travelled down the corridor, turned around and travelled back to its starting position. This showed that our approach to navigation could be easily integrated into an overall navigation methodology. Additionally, navigation results obtained in areas of strong non-uniform illumination change, using an eigenspace approximation to the Hausdorff fraction, were presented and shown to be successful.

Finally, an attention mechanism, termed Information Sampling, was developed to allow a mobile robot to focus upon discriminating information, contained within an image set, acquired *a priori*. This reduced the computational load upon the robot, allowing it to maximise use of its limited resources. Information Sampling was a statistical, non-featured based method and theoretically, could be applied on a pixel-by-pixel basis to any type of image. Navigation was shown to be possible using only the data contained within Information Windows. In an effort to broaden the scope of application, object recognition results were detailed.

## 6.2 Future Research Directions

A number of topics can be viewed as viable directions for further research:

1. **Camera Design**: One could certainly investigate other camera geometries which could have a positive impact on the navigation system design. For the case of catadioptric sensors, what other mirror shapes and sensor layout could be utilised? Also, a number of optical aberrations affect the quality of the omnidirectional images obtained from our systems. These include astigmatism and field curvature. Modelling these would allow for corrective measures to be taken. A second goal is to minimise the design in order to widen the potential fields of application for omnidirectional cameras.

2. **Automatic Knowledge Extraction**: In the current implementation, nodes within the environment must be specified by a human. Thus, while our robot is autonomous, it is not independent. An extension of the approach would allow the robot to use various criteria to evaluate a particular scene's importance.

3. **Construction of the Environmental Representation**: In this work, the environmental representation was built with images acquired *a priori*. An interesting area of future research would be to endow the robot with the ability to extend its representation *online*, as it traverses through the environment. Information Sampling is a useful first step in this process as it allows one to build a highly compact environmental representation. Thus, given the size of the eigenvectors, real-time rebuilding of the low-dimensional eigenspace is possible.

4. **Teleoperation**: As detailed in [46, 52] it is relatively simple to build a 3D representation of the environment traversed by the robot. Each image in this model could be associated to an image in the topological map. Thus, a user could easily select a target for the robot to reach autonomously.

5. **Information Window Selection**: As it stands, Information Sampling is computationally constrained to finding the best subwindows within a set of images. It would be beneficial to enhance this approach by randomly sampling an image to find the uncertainty associated with the selected data.

6. **Object Recognition**: A short term research goal would entail finding objects in large cluttered scenes. All this would require is searching for the Information Window(s) which define a particular object. Additionally, Information Windows could be found *per object*, rather than over the entire object database.

# Bibliography


[1] N. Aihara, H. Iwasa, N. Yokoya, and Haruo Takemura. Memory-based self-localization using omnidirectional images. In *International Conference on Pattern Recognition (ICPR'98)*, pages 1799–1803, 1998.

[2] C. Andersen, S. Jones, and J. L. Crowley. Appearance based processes for visual navigation. In *Proceedings of the 5th International Symposium on Intelligent Robotic Systems (SIRS'97)*, pages 551–557, July 1997.

[3] S. Baker and S. Nayar. A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, 35(2):175–196, 1999.

[4] S. Baker and S. K. Nayar. A theory of catadioptric image formation. In *Proceedings of the International Conference on Computer Vision (ICCV'97)*, pages 35–42, Bombay, India, January 1998.

[5] D. Bapna, E. Rollins, J. Murphy, M. Maimone, W.L. Whittaker, and D. Wettergreen. The atacama desert trek: Outcomes. In *Proceedings of International Conference on Robotics and Automation (ICRA'98)*, pages 597–604, 1998.

[6] M. Barth and C. Burrows. A fast panoramic imaging system and intelligent imaging technique for mobile robots. In *Proceedings of the International Conference on Intelligent Robotics and Systems (IROS'96)*, pages 35–42, 1996.

[7] P. Batavia, D. Pomerleau, and C. Thorpe. Predicting lane position for roadway departure prevention. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, October 1998.

[8] R. Benosman and S. B. Kang. *Panoramic Vision: Sensors, Theory and Application*. Springer-Verlag, 2001.

[9] M. J. Black and A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *Interational Journal of Computer Vision*, 26(1):63–84, 1998.

[10] R. Brooks. Visual map making for a mobile robot. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'85)*, pages 824–829, 1985.

[11] R. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2(1):14–23, 1986.

[12] A. Bruckstein and T. Richardson. Omniview cameras with curved surface mirrors. In *Proceedings of the IEEE Workshop on Omnidirectional Vision at CVPR 2000*, pages 79–86, Hilton Head, SC, USA, June 2000. First published in 1996 as a Bell Labs Technical Memo.

[13] J. Buhmann, W. Bugard, B. Cremers, D. Fox, T. Hofmann, F. Schneider, J. Strikos, and S. Thrun. The mobile robot rhino. *AI Magazine*, 16(1), 1995.

[14] P.J. Burt and E.H. Adelson. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics*, 2(4):217–236, October 1983.

[15] Z. Cao and E. L. Hall. Beacon recognition in omni-vision guidance. In *Proceedings of the International Conference on Optoelectronic Science and Engineering*, pages 778–790, 1990.

[16] Z. L. Cao, S. J. Oh, and E.L. Hall. Dynamic omni-directional vision for mobile robots. *Journal of Robotic Systems*, 3(1):5–17, 1986.

[17] B. A. Cartwright and T. S. Collett. Landmark maps for honeybees. *Biological Cybernetics*, 57:85–93, 1987.

[18] R. Cassinis, D. Grana, and A. Rizzi. Self-localization using an omnidirectional image sensor. In *Proceedings of the 4th International Symposium on Intelligent Robotic Systems (SIRS'96)*, pages 215–222, 1996.

[19] J. S. Chahl and M. V. Srinivasan. Reflective surfaces for panoramic imaging. *Applied Optics (Optical Society of America)*, 36(31):8275–8285, November 1997.

[20] P. Chang and M. Herbert. Omni-directional structure from motion. In *Proceedings of the 1st International IEEE Workshop on Omnidirectional Vision (OMNIVIS'00) at CVPR 2000*, Hilton Head, SC, USA, June 2000.

[21] R. Chatila and J. Laumond. Position referencing and consistent world modeling for mobile robots. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '85)*, pages 138–145, 1985.

[22] O. Chomat and J. L. Crowley. Recognizing motion using local appearance. In *Proceedings of the 6th International Symposium on Intelligent Robotic Systems (SIRS'98)*, pages 271–279, Edinburgh, United Kingdom, July 1998.

[23] T. S. Collett, E. Dillmann, A. Giger, and R. Wehner. Visual landmarks and route following in the desert ant. *J. Comp. Physiology A*, 170:435–442, 1992.

[24] T. Conroy and J. Moore. Resolution invariant surfaces for panoramic vision systems. In *Proceedings of the International Conference on Computer Vision (ICCV'99)*, pages 392–397, Greece, 1999.

[25] Edmund Scientific Corporation. Edmund scientific catalogue, 1997.

[26] J. L. Crowley. Navigation for an intelligent mobile robot. *IEEE Journal of Robotics and Automation*, RA-1(1):31–41, March 1985.

[27] R Dawkins. *Climbing Mount Improbable*. Norton, New York, 1996.

[28] V. Colin de Verdière and J. L. Crowley. Local appearance space for recognition of navigation landmarks. In *Proceedings of the 6th International Symposium on Intelligent Robotic Systems (SIRS'98)*, pages 261–269, Edinburgh, United Kingdom, July 1998.

[29] V. Colin de Verdière and J. L. Crowley. Visual recognition using local appearance. In *5th European Conference on Computer Vision, (ECCV 1998)*, pages 640–654, Freiburg, Germany, June 1998.

[30] C. Canudas de Wit, H. Khennouf, C. Samson, and O. J. Sordalen. Chap.5: Nonlinear control design for mobile robots. In Yuan F. Zheng, editor, *Nonlinear control for mobile robots*. World Scientific series in Robotics and Intelligent Systems, 1993.

[31] C. Deccó, J. Gaspar, N. Winters, and J. Santos-Victor. Omniviews mirror design and software tools. In *EU IST Project: Omniviews - Delieverable DI-3*, September 2001.

[32] L. Delahoche, C. Pégard, B. Marhic, and P. Vasseur. A navigation system based on an omnidirectional vision sensor. In *Proceedings of the International Conference on Intelligent Robotics and Systems 1997 (IROS'97)*, pages 718–724, 1997.

[33] F. Dellaert and R. Collins. Fast image-based tracking by selective pixel integration. In *Proceedings of the FRAME-RATE Workshop at the IEEE International Conference on Computer Vision (ICCV'99)*, 1999.

[34] S. Derrien and K. Konolige. Approximating a single viewpoint in panoramic imaging devices. In *Proceedings of the 1st International IEEE Workshop on Omnidirectional Vision at CVPR 2000*, pages 85–90, Hilton Head, SC, USA, June 2000.

[35] E.D. Dickmanns, R. Behringer, D. Dickmanns, T. Hildebrandt, M. Maurer, J. Schiehlen, and F. Thomanek. The seeing passenger car vamors-p. In *Proceedings of the International Symposium on Intelligent Vehicles*, Paris, France, 1994.

[36] K. Daniilidis (Ed.). *1st International IEEE Workshop on Omnidirectional Vision at CVPR 2000*, June 2000.

[37] P. Greguss (Ed.). *IEEE ICAR 2001 Workshop on Omnidirectional Vision Applied to Robotic Orientation and Non-destructive Testing*, August 2001.

[38] A. Elfes. Sonar-based real world mapping and navigation. *IEEE Journal of Robotics anf Automation*, RA-3(3):249–265, June 1987.

[39] R. Elkins and E. Hall. 3d line following using omnidirectional vision. *Proceedings of the SPIE International Robots and Computer Vision XIII: 3D Vision, Product Inspection and active Vision*, 2354:130–144, 1994.

[40] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3):313–326, June 1992.

[41] E. Fabrizi and A. Saffiotti. Extracting topology-based maps from gridmaps. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'00)*, 2000.

[42] 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 ACM*, 24(6):381–395, June 1981.

[43] J. Foote and D. Kimber. Flycam: Practical panoramic video and automatic camera control. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, volume III, pages 1419–1422, August 2000.

[44] M. Franz, B. Schölkopf, P. Georg, H. Mallot, and H. Bülthoff. Learning view graphs for robot navigation. In *Proceedings of the First International Conference on Autonomous Agents (Agents'97)*, 1997.

[45] S. Gachter, T. Pajdla, and B. Mičušik. Mirror design for an omnidirectional camera with a space variant imager. In *Proceedings of the IEEE ICAR 2001 Workshop: Omnidirectional Vision Applied to Robotic Orientation and Nondestructive Testing*, pages 99–106, Budapest, Hungary, August 2001.

[46] J. Gaspar, E. Grossmann, and J. Santos-Victor. Interactive reconstruction from an omnidirectional image. In *Proceedings of the 9th International Symposium on Intelligent Robotic Systems (SIRS'01)*, Toulouse, France, July 2001.

[47] J. Gaspar and J. Santos-Victor. Visual path following with a catadioptric panoramic camera. In *Proceedings of the 7th International Symposium on Intelligent Robotic Systems (SIRS'99)*, pages 139–147, Coimbra, Portugal, July 1999.

[48] J. Gaspar, N. Winters, and J. Santos-Victor. Vision-based navigation and environmental representations with an omni-directional camera. *IEEE Transactions on Robotics and Automation*, 16(6):890–898, December 2000.

[49] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*. to appear.

[50] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical implications. In *Proceedings of the 7th European Conference on Computer Vision (ECCV 2000)*, pages 445–461, Dublin, Ireland, June 2000.

[51] P. Greguss. Panoramic imaging block for 3d space. US patent 4,566,763, January 1986. Hungarian Patent granted in 1983.

[52] E. Grossmann, D. Ortin, and J. Santos-Victor. Algebraic aspects of reconstruction of structured scenes from one or more views. In *Proceedings of the British Machine Vision Conference (BMVC'01)*, pages 633–642, 2001.

[53] A. Hait, T. Simeon, and M. Taix. Robust motion planning for rough terrain navigation. In *Procedings of the IEEE International Conference on Intelligent Robots and Systems*, 1999.

[54] T. Hancock and S. Judd. Ratbot: Robot navigation using simple visual algorithms. In *Proceedings of the IEEE Regional Conference on Control Systems*, pages 181–184, August 1993.

[55] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.

[56] Richard Hartley and Andrew Zissermann. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[57] E. Hecht and A. Zajac. *Optics*. Addison Wesley, 1974.

[58] R. Hicks and R. Bajcsy. Reflective surfaces as computational sensors. In *Workshop on Perception for Mobile Agents at CVPR'99*, Colorado, USA, June 1999.

[59] R. Hicks and R. Bajcsy. Catadioptric sensors that approximate wide-angle perspective projections. In *Proceedings of the IEEE Computer Vision and Pattern*

*Recognition Conference (CVPR'00)*, pages 545–551, Hilton Head, SC, USA, June 2000.

[60] J. Hong, X. Tan, B. Pinette, R. Weiss, and E. M. Riseman. Image-based navigation using 360° views. In *Proceedings of the Image Understanding Workshop*, pages 782–791, 1990.

[61] J. Hong, X. Tan, B. Pinette, R. Weiss, and E. M. Riseman. Image-based homing. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'91)*, pages 620–625, 1991.

[62] I. Horswill. Polly: A vision-based artifical agent. In *Proceedings of the AAAI-93*, Washington, DC, USA, 1993.

[63] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:498–520, 1933.

[64] S. Hutchinson, G. Hager, and P. Corke. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12(5):651–670, October 1996.

[65] D. Huttenlocher, R. Lilien, and C. Olsen. Object recognition using subspace methods. In *Proceedings of the Fourth European Conference on Computer Vision*, volume 1, pages 536–545, 1996.

[66] D.P. Huttenlocher, R. Lilien, and C. Olsen. View-based recognition using an eigenspace approximation to the hausdorff measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):951–956, September 1999.

[67] H. Ishiguro and S. Tsuji. Image-based memory of environment. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'96)*, pages 634–639, 1996.

[68] H. Ishiguro, M. Yamamoto, and S. Tsuji. Omni-directional stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), February 1992.

[69] J.Minguez, L.Montano, T.Simeon, and R. Alami. Global nearness diagram navigation (gnd). In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA2001)*, Korea, 2001.

[70] M. Jogan and A. Leonardis. Robust localization using eigenspace of spinning-images. In *Proceedings of the 1st International IEEE Workshop on Omnidirectional Vision (OMNIVIS'00) at CVPR 2000*, pages 37–44, Hilton Head, SC, USA, June 2000.

[71] I. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 1986.

[72] K. Kato, S. Tsuji, and H. Ishiguro. Representing environment through target-guided navigation. In *Proceedings of the International Conference on Pattern Recognition*, pages 1794–1798, 1998.

[73] T. Kawanishi, K. Yamazawa, H. Iwasa, H. Takemura, and N. Yokoya. Generation of high resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors. In *Proceedings of the International Conference on Pattern Recognition*, pages 485–489, 1998.

[74] V. Klema and A. Laub. The singular value decomposition: Its computation and some applications. *IEEE Transactions on Automatic Control*, AC-25(2):164–176, April 1980.

[75] M. Knapek, R. Swain Oropeza, and D. Kriegman. Selecting promising landmarks. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 3771–3777, 2000.

[76] T. Kohonen. *Associative memory - a system-theoretical approach*. Springer-Verlag, Heidelberg, Germany, 1977.

[77] D. Kortenkamp and T. Weymouth. Topological mapping for mobile robots using a combination of sonar and visual sensing. In *AAAI'94*, July 1994.

[78] A. Kosaka. Purdue experiments in model-based vision for hallway navigation. In *Proceedings of Workshop on Vision for Robots at IROS '95*, pages 87–96, 1995.

[79] J. Košecká. Visually guided navigation. In *Proceedings of the 4th International Symposium on Intelligent Robotic Systems (SIRS'96)*, Lisbon, Portugal, July 1996.

[80] B. Kuipers. Modeling spatial knowledge. *Cognitive Science*, 2:129–153, 1978.

[81] B. Kuipers. The map in the head metaphor. *Environment and Behavior*, 14(2):202–220, 1982.

[82] B. Kuipers and Y. Byun. A robot exploration and mapping strategy on a semantic hierarchy of spatial representations. *Journal of Robotics and Autonomous Systems*, 8:47–63, 1991.

[83] M. Land and R.D. Fernald. The evolution of eyes. *Annual Review of Neuroscience*, 15:1–29, 1992.

[84] J. J. Leonard and H. F. Durrant-Whyte. Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation*, 7(3), June 1991.

[85] C. Lin and R. Tummala. Mobile robot navigation using artifical landmarks. *Journal of Robotic Systems*, 14(2):93–106, Feburary 1997.

[86] L. J. Lin, T. R. Hancock, and J. S. Judd. A robust landmark-based system for vehicle location using low-bandwidth vision. *Journal of Robotics and Autonomous Systems*, 25:19–32, 1998.

[87] K. Lynch. *The Image of the City*. MIT Press, 1960.

[88] S. Maeda, Y. Kuno, and Y. Shirai. Active navigation vision based on eigenspace analysis. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS'97)*, pages 1018–1023, 1997.

[89] A. Majumder, W. Seales, G. Meenakshisundaram, and H. Fuchs. Immersive teleconferencing: A new algorithm to generate seamless panoramic video imagery. In *Proceedings of the 7th ACM Conference on Multimedia*, 1999.

[90] J. Martin and J. L. Crowley. An appearance-based approach to gesture recognition. In *Proceedings of the International Conference on Image Analysis and Processing*, Florence, Italy, September 1997.

[91] M. K. Mataric. Integration of representation into goal-driven behavior-based robotics. *IEEE Transactions on Robotics and Automation*, 8(3):304–312, June 1992.

[92] Y. Matsumoto, K. Ikeda, M. Inaba, and H. Inoue. Exploration and navigation in corridor environment based on omni-view sequence. In *Proceedings of the 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1505–1510, 2000.

[93] Y. Matsumoto, M. Inaba, and H. Inoue. Visual navigation using view-sequenced route representation. In *Proceedings of the International Conference on Robotics and Automation*, pages 83–88, 1996.

[94] Y. Matsumoto, M. Inaba, and H. Inoue. Memory-based navigation using omni-view sequence. In *Proceedings of the International Conference on Field and Service Robotics*, pages 184–191, 1997.

[95] K. Miyamoto. Fish-eye lens. *Journal of the Optical Society of America*, 54(8):1060–1061, 1964.

[96] H. P. Moravec. Towards automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artifical Intelligence*, Massachusetts, USA, 1977.

[97] H. P. Moravec. The stanford cart and the cmu rover. *Proceedings of the IEEE*, 71:872–884, 1983.

[98] H. P. Moravec. Robust navigation by probabilistic volumetric sensing. Mass-market utility robots before 2010, Carnegie Mellon University, 2000.

[99] 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.

[100] H. Murase, F. Kimura, F. Yoshimura, and Y. Miyake. An improvement of the autocorrelation matirx in pattern matching method and its application to hand printed 'hiragana'. *Transactions of IEICE*, J64-D(3):276–283, 1981.

[101] H. Murase and M. Lindenbaum. Partial eigenvalue decomposition of large images using spatial temporal adaptive method. *IEEE Transactions on Image Processing*, 4(5):620–629, 1995.

[102] 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.

[103] V. Nalwa. A true omni-directional viewer. Technical report, Bell Laboratories, Holmdel, New Jersey, USA, February 1996.

[104] S. K. Nayar and S. Baker. Omnidirectional image formation. In *Proceedings of the DARPA Image Understanding Workshop*, New Orleans, LA, USA, May 1997.

[105] S. K. Nayar, S. A. Nene, and H. Murase. Subspace methods for robot vision. Technical Report CUCS-06-95, Department of Computer Science, Columbia University, New York, USA, 1995.

[106] S. K. Nayar and V. Peri. Folded catadioptric cameras. In *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, Fort Collins, June 1999.

[107] S. Nene, S. Nayar, and H. Murase. Columbia object image library (coil-20). Technical Report CUCS-005-96, Columbia University, February 1996.

[108] K. Ohba and K. Ikeuchi. Detectibility, uniqueness and reliability of eigen windows for stable verification of partially occluded objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9):1043–1048, September 1997.

[109] E. Oja. *Subspace methods for Pattern Recognition*. Research Studies Press, 1983.

[110] C. Owen and U. Nehmzow. Lankmark-based navigation for a mobile robot. In *Proceedings of the simulation of Adaptive Behaviour*. MIT Press, 1998.

[111] T. Pajdla. Localization using SVAVISCA panoramic images of agam fiducials - limits of performance. Research Report CTU–CMP–2001–11, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, March 2001.

[112] T. Pajdla and V Hlaváč. Zero phase representation of panoramic images for image based localization. In *Proceedings of the 8th International Conference on Computer Analysis of Images and Patterns*, pages 550–557, September 1999.

[113] K Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, Sixth Series 2:559–572, 1901.

[114] V. Peri and S. Nayar. Generation of perspective and panoramic video from omnidirectional video. In *Proceedings of the DARPA Image Understanding Workshop*, May 1997.

[115] C. Rasmussen and G. Hager. Robot navigation using image sequences. In *Procedings of the AAAI'96*, pages 938–943, 1996.

[116] D. W. Rees. Panoramic television viewing system. United States Patent Num. 3505465, April 1970.

[117] M. Rendas and M. Perrone. Using field subspaces for on-line survey guidance. In *Proceedings of Oceans 2000*, Providence, RI, USA, June 2000.

[118] W. Rucklidge. *Efficient Visual Recognition using the Hausdorff Distance*, volume 1173 of *Lecture Notes in Computer Science*. Springer-Verlag, 1996.

[119] A. P. Sage and J. L. Melsa. *Estimation Theory with Applications to Communications and Control*. McGraw-Hill, 1971.

[120] J. Santos-Victor and G. Sandini. Visual behaviors for docking. *Computer Vision and Image Understanding*, 67(3):223–238, September 1997.

[121] J. Santos-Victor, G. Sandini, F. Curotto, and S. Garibaldi. Divergent stereo in autonomous navigation: From bees to robots. *International Journal of Computer Vision*, 14:159–177, 1995.

[122] J. Santos-Victor, R. Vassallo, and H. J. Schneebeli. Topological maps for visual navigation. In *1st International Conference on Computer Vision Systems*, pages 21–36, Las Palmas, Canarias, January 1999.

[123] K. Sarachik. Characterising an indoor envrionment with a mobile robot and uncalibrated stereo. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 984–989, 1989.

[124] B. Schatz, S. Chameron, G. Beugnon, and T. S. Collett. The use of path integration to guide route learning ants. *Nature*, 399:769–772, 24 June 1999.

[125] B. Schiele and J. L. Crowley. Object recognition using multidimentional receptive field histograms. In *Proceedings of European Conference on Computer Vision (ECCV96)*, pages 610–619, 1996.

[126] B. Schiele and J. L. Crowley. Where to look next and what to look for. In *Proceedings of International Symposium on Intelligent Robotic Systems (SIRS'96)*, pages 139–145, 1996.

[127] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, May 1997.

[128] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. In *Proceedings of International Conference on Computer Vision (ICCV98)*, pages 230–235, 1998.

[129] A. W. Seigel and S. H. White. *The Development of spatial representations of large-scale environments*, volume 10 of *Advances in Child Development and Behavior*. Academic Press, 1975.

[130] J. Shi and C. Tomasi. Good features to track. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pages 593–600, Seattle, USA, 1994.

[131] R. Sim and G. Dudek. Mobile robot localization from learned landmarks. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'98)*, Victoria, Canada, October 1998.

[132] L. Sirovich and M. Kirby. Low dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America*, 4(3):519–524, 1987.

[133] T. Sogo, H. Ishiguro, and M. Treivedi. Real-time target localization and tracking by n-ocular stereo. In *Proceedings of the 1st International IEEE Workshop on Omnidirectional Vision at CVPR 2000*, Hilton Head, SC, USA, June 2000.

[134] P. Sturm. A method for 3d reconstruction of piecewise planar objects from single panoramic images. In *Proceedings of the 1st International IEEE Workshop on Omnidirectional Vision at CVPR 2000*, Hilton Head, SC, USA, June 2000.

[135] Svavisca. Document on specification - esprit project n. 31951 - svavisca. available at http://www.lira.dist.unige.it - SVAVISCA - GIOTTO Home Page, May 1999.

[136] T. Svoboda, T. Pajdla, and V. Hlavác. Epipolar geometry for panoramic cameras. In *Proceedings of the 6th European Conference on Computer Vision (ECCV'98)*, pages 218–231, Freiburg, Germany, July 1998.

[137] S. Thrun. Finding landmarks for indoor mobile robot navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'98)*, pages 958–963, 1998.

[138] S. Thrun and A. Bucken. Integrating grid-based and topological maps for mobile robot navigation. In *Proceedings of the 13th National Conference on Artifical Intelligence (AAAI'96)*, 1996.

[139] S. Thrun and A. Bucken. Learning maps for indoor mobile robot navigation. In *Carnegie Mellon University Technical Report CS-96-121*, 1996.

[140] E. C. Tolman. Cognitive maps in rats and men. *The Psychological Review*, 55(4):189–208, 1948. available from: http://psychclassics.yorku.ca/Tolman/Maps/maps.htm.

[141] G.J. Toomer. *Diocles on Burning Mirrors*, volume 1 of *Sources in the History of Mathematics and Physical Sciences*. Springer-Verlag, Heidelberg, Germany, 1976.

[142] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'91)*, pages 586–591, 1991.

[143] I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'00)*, pages 1023–1029, 2000.

[144] R. F. Vassallo, H. J. Schneebeli, and J. Santos-Victor. Visual navigation: Combining visual servoing and appearance based methods. In *Proceedings of the 6th International Symposium on Intelligent Robotic Systems (SIRS'98)*, pages 137–146, Edinburgh, United Kingdom, July 1998.

[145] F. Wallner. *Position estimation for a mobile robot from principal components of laser range data*. Ph.d., I.N.P de Grenoble, Grenoble, France, October 1997.

[146] F. Wallner, B. Schiele, and J. L. Crowley. Position estimation for a mobile robot from principle components of laser range data. In *Proceedings of the International Symposium on Intelligent Robotic Systems*, pages 215–224, Stockolm, Sweden, July 1997.

[147] S. Watanabe. Karhunen-loève expansion and factor analysis. In *Transactions of the 4th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, pages 635–660, Prague, Czech Republic, 1965.

[148] S. Watanabe. *Knowing and guessing - a quantitative study of inference and information*. John Wiley, New York, USA, 1969.

[149] R. Wehner and S. Wehner. Insect navigation: use of maps or ariadne's thread? *Ethology, Ecology, Evolution*, 2:27–48, 1990.

[150] L. Weiss, A. Sanderson, and C. Neuman. Dynamic sensor-based control of robots with visual feedback. *IEEE Journal of Robotics and Automation*, RA-3(5):404–417, October 1987.

[151] N. Winters, J. Gaspar, A. Bernardino, and J. Santos-Victor. Vision algorithms for omniviews cameras. In *EU IST Project: Omniviews - Delieverable DI-2*, September 2001.

[152] N. Winters, J. Gaspar, E. Grossmann, and J. Santos-Victor. Experiments in visual-based navigation with an omnidirectional camera. In *Proceedings of the IEEE ICAR 2001 Workshop: Omnidirectional Vision Applied to Robotic Orientation and Nondestructive Testing*, Budapest, Hungary, August 2001. Invited Talk.

[153] N. Winters, J. Gaspar, G. Lacey, and J. Santos-Victor. Omni-directional vision for robot navigation. In *Proceedings of the 1st International IEEE Workshop on Omnidirectional Vision (OMNIVIS'00) at CVPR 2000*, Hilton Head, SC, USA, June 2000.

[154] N. Winters and G. Lacey. Overview of tele-operation for a mobile robot. In *TMR Workshop on Computer Vision and Mobile Robots. (CVMR'98)*, Santorini, Greece, September 1999.

[155] N. Winters and J. Santos-Victor. Information sampling for vision-based robot navigation. *Journal of Robotics and Autonomous Systems*. to appear.

[156] N. Winters and J. Santos-Victor. Mobile robot navigation using omni-directional vision. In *Proceedings of the 3rd Irish Machine Vision and Image Processing Conference (IMVIP'99)*, Dublin, Ireland, September 1999.

[157] N. Winters and J. Santos-Victor. Omni-directional visual navigation. In *Proceedings of the 7th International Symposium on Intelligent Robotic Systems (SIRS'99)*, pages 109–118, Coimbra, Portugal, July 1999.

[158] N. Winters and J. Santos-Victor. Information sampling for appearance based 3d object recognition and pose estimation. In *Proceedings of the Irish Machine Vision and Image Processing Conference (IMVIP'01)*, Maynooth, Ireland, September 2001.

[159] N. Winters and J. Santos-Victor. Information sampling for optimal image data selection. In *Proceedings of the 9th International Symposium on Intelligent Robotic Systems (SIRS'01)*, Toulouse, France, July 2001.

[160] N. Winters and J. Santos-Victor. Visual attention-based robot navigation using information sampling. In *Proceedings of the 2001 International Conference on Intelligent Robots and Systems (IROS'01)*, Hawaii, USA, October 2001.

[161] R. Wood. Fish-eye views and vision underwater. *Philosophical Magazine*, 12(6):159–162, 1906.

[162] Y. Yagi. Omnidirectional sensing and its applications. *IEICE Transactions on Information and Systems*, E82-D(3):568–579, March 1999.

[163] Y. Yagi, S. Fujimura, and M. Yachida. Route representation for mobile robot navigation by omnidirectional route panorama fourier transform. In *Proceedings of the IEEE International Conference Robotics and Automation (ICRA'98)*, 1998.

[164] Y. Yagi, Y. Nishizawa, and M. Yachida. Map-based navigation for mobile robot with omnidirectional image sensor copis. *IEEE Transactions on Robotics and Automation*, 11(5):634–648, October 1995.

[165] K. Yamazawa, Y. Yagi, and M. Yachida. Omnidirectional imaging with a hyperboloid projection. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1993.

[166] E. Yeh and D. J. Kriegman. Toward selecting and recognizing natural landmarks. Technical Report 9503, Yale University, 1995.

[167] Z. Zhang and O.D. Faugeras. Building a 3d world model with a mobile robot. In *Proceedings of the 10th International Conference on Pattern Recognition*, 1990.

[168] 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.

# Appendix A

# Singular Values and Eigenvalues

## A.1 Singular Value Decomposition

The Singular Value Decomposition (SVD) [74] of an $M \times N$ matrix $A$ is a factorization of the form $A = U\Sigma V^T$. Here the columns of the $M \times M$ orthogonal matrix $U$ and of the $N \times N$ orthogonal matrix $V$ are known as the left and right *singular vectors*, respectively. The diagonal entries of the $M \times N$ matrix $\Sigma$ are non-negative and are known as the *singular values* of the matrix $A$. They are not the same thing as eigenvalues: the singular values of $A$ are the square-roots of the eigenvalues of $A^T A$. However, if $A^T A$ is symmetric and positive definite, its eigenvalues are real and non-negative. Consequently, the singular values are real and non-negative [56].

## A.2 Singular Values and Eigenvalues

We now present a proof that singular values of $A$, $\sigma(A)$ are equal to the eigenvalues of $A$, $\lambda(A)$, if $A$ is symmetric and positive definite.

Hypothesis: $A$ is symmetric and positive definite:

$$\begin{aligned} A &= A^T \\ A &= U\Sigma V^T, \text{ with } \Sigma_{ii} > 0 \end{aligned}$$

Since $A^T = A$, we have that the SVD must follow the following property:

$$\begin{aligned} A &= U\Sigma V^T \\ A^T &= V\Sigma^T U^T = A \\ \Rightarrow A &= U\Sigma U^T \end{aligned}$$

By definition, we have that $\sigma(A) = \sqrt{\lambda(A^T A)}$. Using the SVD of A, we have:

$$\begin{aligned} \lambda(A^T A) &= \lambda(U\Sigma U^T U\Sigma U^T) \\ &= \lambda(U\Sigma^2 U^T) \\ &= \lambda(\Sigma^2), \text{ because } U \text{ is unitary} \\ &= \lambda^2(\Sigma). \end{aligned}$$

Since $U$ is orthogonal, $U^T = U^{-1}$. Using the SVD of A, we then have:

$$\begin{aligned} A = U\Sigma U^T &\Leftrightarrow A = U\Sigma U^{-1} \\ &\Leftrightarrow AU = U\Sigma \end{aligned}$$

which defines $\Sigma$ as the eigenvalues of A ($\lambda(\Sigma) = \lambda(A)$). Using this result we complete the proof that we need:

$$\begin{aligned} \sigma(A) &= \sqrt{\lambda(A^T A)} \\ &= \sqrt{\lambda^2(A)} \\ &= |\lambda(A)|, \text{ and since A is positive definite} \\ &= \lambda(A) \text{ which concludes the proof.} \end{aligned}$$