ACTIVE SENSORS FOR LOCAL PLANNING IN MOBILE ROBOTICS
WORLD SCIENTIFIC SERIES IN ROBOTICS AND INTELLIGENT SYSTEMS

Editor-in-Charge:  C J Harris (University of Southampton)
Advisor: T M Husband (University of Salford)

Published:
Vol. 10: Cellular Robotics and Micro Robotic Systems
(T Fukuda and T Ueyama)
Vol. 11: Recent Trends in Mobile Robots (Ed. Y F Zheng)
Vol. 12: Intelligent Assembly Systems (Eds. M Lee and J J Rowland)
Vol. 13: Sensor Modelling, Design and Data Processing for Autonomous Navigation
(M D Adams)
Vol. 14: Intelligent Supervisory Control: A Qualitative Bond Graph Reasoning
Approach (H Wang and D A Linkens)
Vol. 15: Neural Adaptive Control Technology (Eds. R Zbikowski and K J Hunt)
Vol. 17: Applications of Neural Adaptive Control Technology (Eds. J Kalkkuhl,
K J Hunt, R Zbikowski and A Dzielinski)
Vol. 18: Soft Computing in Systems and Control Technology
(Ed. S Tzafestas)
Vol. 19: Adaptive Neural Network Control of Robotic Manipulators
(S S Ge, T H Lee and C J Harris)
Vol. 20: Obstacle Avoidance in Multi-Robot Systems: Experiments in Parallel
Genetic Algorithms (M A C Gill and A Y Zomaya)
Vol. 21: High-Level Feedback Control with Neural Networks
(Eds. F L Lewis and Y H Kim)
Vol. 22: Odour Detection by Mobile Robots
(R. Andrew Russell)
Vol. 23: Fuzzy Logic Control: Advances in Applications
(Eds. H B Verbruggen and R Babuska)
Vol. 24: Interdisciplinary Approaches to Robot Learning
(Eds. J. Demiris and A. Birk)
Vol. 25: Wavelets in Soft Computing
(M. Thuillard)
ACTIVE SENSORS FOR LOCAL PLANNING IN MOBILE ROBOTICS

PENELOE PROBERT SMITH
University of Oxford, UK
Preface

The goal of realising a machine which mimics the human ability to refine and structure behaviour in a complex, dynamic world continues to drive mobile robot research. Central to such ability is the need to gather and manipulate rich information on the surroundings. Such a grand ambition places stringent requirements on the sensing systems and on the interaction between sensor and task.

One thing which has become clear in attempts to achieve this is the need for diversity in sensing systems. The human vision system remains the inspiration for artificial analogues, but none can approach its sophistication in terms of hardware or processing. Structured light systems, which measure range directly through using a light source to probe a specific area, are a more reliable method for artificial planning. Their equivalent in sound, sonar, has increased in adaptability and reliability, driven by collaboration with bat biologists as well as from the more standard and established radar literature. Radar itself is becoming cheaper.

Given such diversity, another requirement is a structure and methodology to share and optimise information. Two important paradigms have arisen as a result. One is the idea of the logical sensor which hides the details of the physical sensing operation, so sensors may be specified in terms of task and not in terms of technology: hence a task might require, for example, a sensor to find line segments under particular conditions, rather than a particular technology such as sonar. The other is the active sensor, which abstracts and selects information according to demand - whether this is through probing the environment physically - for example through emitting radiation (the traditional active sensor) or through choice or tuning.
of algorithms. This concept is an extension of the traditional formulation of the active sensor which interacts with the environment through emitting radiation such as sound or light. By developing sensors within this framework we avoid the bottleneck of a large information repository.

Much of the work in this book is the result of research with which the editor has been associated in Oxford. It is designed both to provide an overview of the state of the art in active range and vision sensing and to suggest some new developments for future work. It describes real systems and sensors. Cross references have been included between chapters to develop and relate concepts across and within a single sensing technique.

The book starts with a brief overview of the demands for local planning, discussing the problem of finding a reliable architecture to handle complexity and adaptability. It describes the concept of the active sensor, driven by the task in hand and filtering information for that task, to provide a fast, tight sensing-planning loop. It gives an overview of common sensing technologies.

In mobile robots, a key requirement for planning is to find out where the robot is within a known region - the localisation problem. Mapping, the problem of extracting geometric or feature based information often underlies this. Reliable mapping and localisation requires robust and versatile sensors, and also a systematic method to handle the uncertainty inherent in the sensors and in the robot's own position. Chapter 2 addresses generic issues in mapping and localisation and introduces an important algorithm which is referred to many times in the book, the extended Kalman filter.

Sensors which measure range directly are particularly useful for planning. Sensors active in the traditional sense are most important here and most of the book deals with hardware and algorithms for the two most common classes of these: sonar sensors and optoelectronic sensors.

The essential factor which distinguishes the way sensors in these classes view the world is their wavelength. Whereas the data from optical sensors naturally falls into standard geometric descriptions such as lines, corners and edges, millimetre wave sensors such as sonar see the world rather differently. Part II of the book discusses millimetre wave sensors. Significant interpretation is required to extract data for comparison with a standard geometric model. In spite of this, sonar is the commonest sensor used in robotics, largely because of its low cost and easy availability. Another sensor which operates in the millimetre band is high frequency radar - more expensive but with very long range and so of great interest outdoors. Although
Preface

one of these sensors emits sound waves and the other electromagnetic waves, because of the similar wavelength their data has many similar characteristics. Chapter 3 discusses generally how these characteristics depends on both the sensor geometry (especially the antenna) and target type.

Sonar has seen particular developments in the last ten years, from a simple sensor used for obstacle avoidance to a sensor which will produce reliable and robust maps. Chapters 4 to 6 describe how this has been achieved through advances in hardware and data interpretation. Methods of modulation and signal processing drawn from underwater sonar and military radar have been applied to improve resolution and hence extend the range of environments in which sonar operates (chapter 4). Surface modelling, especially the incorporation of rough surface models, has led to better mapping and application in texture recognition (chapter 5). Drawing on analogies from biology, bio-sonar has improved efficiency through sensor placement and small sensor arrays (chapter 6). Finally the application of new processing techniques, especially morphological filtering, has led to the possibility of curve fitting, to produce information which is geometrically similar to our own perception of the world (chapter 7).

The problem with sonar is power; the maximum range is limited to around 10m or less (normally closer to 5m). Millimetre wave radar has many similar characteristics but will see over ranges huge by robot standards - over several kilometres depending on weather conditions. For this reason it is of great interest in the field, and the increasing use by the automobile industry (for automatic charging for example) means that the cost is falling, although it is still an expensive technology. Chapter 8 describes the capabilities of radar with a summary of some recent work in robotics.

Part III describes sensing at optical wavelengths. Optoelectronic sensors probe the environment using a laser or focussed light emitting diode. At their best, they provide data of high quality which is easy to interpret in terms of standard geometry. However difficulties arise from strong ambient light levels as the active light source can be swamped. A further difficulty in actually realising these systems in the laboratory is the need to scan over one or two dimensions. Unlike scanned sonar, which is compact and light, a scanning optoelectronic sensor imposes power and weight demands which place restrictions on its speed and reactivity. Because of this most applications in local planning gather only two dimensional data (often range versus orientation). Some of these issues are discussed in chapter 9, which also describes some common optical methods to measure range. Chapter
Preface

10 describes in detail a sensor based on a technology which has been of particular importance in robotics, amplitude modulated continuous wave (AMCW) operation, often known as lidar. The following chapter (chapter 11) describes the extraction of lines and curves from this and other types of optical range sensor. Chapter 12 describes active vision, in a system which allows the camera to select features of interest and to maintain these in the centre of its field of view through a multi-degree of freedom head. It is impossible to do justice to such an important subject in a book of this scope and it is hoped that this chapter, besides describing a state of the art system for mapping and localisation, will encourage the reader to pursue more specialised texts.

The final part of this book, Part IV, considers some general issues in sensor management. Chapter 13 describes a system which is showing real benefits for processing visual and infra red data. In addition it introduces the more abstract areas of adaptive sensor and knowledge representation.

The ultimate goal of autonomy remains elusive, but there are many examples of systems influenced strongly by robotics research. Bumper mounted sonar has been introduced as a parking aid in cars; radar is common not just for speed detection but for automatic charging. Surveillance systems draw on active vision to process and abstract information. The multi-agent paradigms used for routing in Internet access have their counterparts in behavioural robotics. The demand for indoor localisation has expanded into areas such as environmental monitoring as a response to the availability of GPS outdoors.

The developments described in this book are relevant to all those who are looking for new and improved ways to handle task orientated information from sensors. It is directed at a final year undergraduate or first year postgraduate level, as well as being of use as a source of ideas to researchers and interested practitioners. Inevitably it has only been able to cover some of the work going on in the field. However I have enjoyed the opportunity to put this book together and I hope that the reader will capture some of the excitement of our research and will use the bibliography as a springboard for their own further investigations.

Penelope Probert Smith
University of Oxford
Acknowledgements

My interest in robotics started when I joined Oxford thirteen years ago and I am grateful to all those who introduced me to the area, especially to Mike Brady. My greatest thanks however must go to those who have contributed to this book, both as authors and less publicly.

Foremost amongst the latter is David Witt, who offered me the use of his CTFM sonar sensor several years ago and inspired my interest in advanced sonar. I have benefited too from work by Gordon Kao, Zafiris Politis, Paul Gilkerson and Konstantinos Zografos. Others (some of whom are represented as authors) have sustained and excited my interest over the years, especially Huosheng Hu whose hardware and systems expertise made sure that we were never short of real data and situations to challenge us.

My thanks to those who have contributed to the overall publication effort, especially David Lindgren who has proved an invaluable source of knowledge on linux.

Last, but not least, my thanks go to my family for putting up with sometimes erratic hours and domestic arrangements!
This page is intentionally left blank
Contents

Preface v

Chapter 1 Introduction 3
1.1 Architectures for Planning and Perception .................. 3
1.2 Range Sensing Technologies ............................... 8
1.3 Planning Demands ........................................ 9

Chapter 2 The Mapping and Localisation Problem 13
2.1 Simultaneous Localisation and Map Building ............... 13
  2.1.1 The Map-Building Process .......................... 14
  2.1.2 The Coupling of Map Estimates ...................... 15
  2.1.3 Simultaneous Localisation and Map-Building with the EKF ....... 17

Chapter 3 Perception at Millimetre Wavelengths 21
3.1 Sensor Operation ........................................... 22
3.2 The Sensor .................................................. 24
3.3 Antenna Properties ......................................... 25
  3.3.1 The Circular Antenna ................................ 26
3.4 Altering Aperture Shape .................................... 29
  3.4.1 Antenna Arrays ....................................... 32
  3.4.2 Focused Transducers ................................ 33
3.5 Target Properties ........................................... 33
  3.5.1 Smooth Surfaces: The Specular Model ................ 34
  3.5.2 Rough Surfaces ....................................... 35
3.5.3 Scattering Cross Section ........................................ 36
3.6 Attenuation in the Transmission Medium .................. 37
  3.6.1 Beam Spreading ............................................. 38
  3.6.2 Losses ...................................................... 38
3.7 Summary .......................................................... 39

Chapter 4  Advanced Sonar: Principles of Operation and
           Interpretation .................................................. 41
  4.1 Single Return Sonar ........................................... 41
    4.1.1 Mapping and Navigation Using Single Return Sonar ... 44
      4.1.1.1 Occupancy Grid Representation ..................... 44
    4.1.2 Landmark Based Mapping ................................ 46
    4.1.3 The Geometric Target Primitives ....................... 47
  4.2 Advanced Sonar: The Sonar Signature ....................... 47
    4.2.1 Range Signature ......................................... 48
    4.2.2 Orientation Signature ................................... 50
    4.2.3 Rough Surfaces .......................................... 51
  4.3 Acquiring the Sonar Signature ................................. 51
    4.3.1 Single Frequency Sonar ................................ 52
      4.3.1.1 Improving Range Accuracy: The Correlation
              Receiver ............................................ 52
    4.3.2 Pulse Compression Sonar ................................ 54
    4.3.3 Continuous Wave Frequency Modulated Sonar ........... 56
    4.3.4 Doppler Effects ......................................... 60
  4.4 Summary ........................................................ 60

Chapter 5  Smooth and Rough Target Modelling:
           Examples in Mapping and Texture Classification ....... 61
  5.1 Power Received by the Transducer .......................... 61
  5.2 Smooth Surface Model ....................................... 62
    5.2.1 Backscattering Coefficient ............................ 62
    5.2.2 The Target Geometry Coefficient ...................... 63
    5.2.3 Mapping Experiments .................................... 63
      5.2.3.1 Finding the Position of Each Feature ............. 64
      5.2.3.2 Finding Geometric Type ............................ 65
      5.2.3.3 Data Integration .................................. 65
  5.3 Rough Surface Planar Models ................................ 68
    5.3.1 Backscattering Coefficient of Rough Surface ........ 69
Contents

5.3.1.1 Finding Position of Rough Surfaces .................................. 70
5.4 Mapping Heterogeneous Environments ........................................ 72
5.5 Texture: Classifying Surfaces .................................................. 72
5.5.1 Reflections from Real Surfaces ............................................. 73
5.5.2 Pathways Classification ....................................................... 75
5.5.3 Finding Suitable Features .................................................... 76
5.5.4 Remarks ............................................................................. 77
5.6 Summary ............................................................................... 77

Chapter 6 Sonar Systems: A Biological Perspective 79
6.1 Introduction ........................................................................... 79
6.2 Echo Formation ...................................................................... 81
6.2.1 Transformations ................................................................... 82
6.2.2 Reflection ........................................................................... 84
6.2.2.1 Reflections from a Planar Reflector ................................... 84
6.2.2.2 Reflections from a Corner ............................................... 85
6.2.2.3 Reflections from an Edge ............................................... 86
6.3 Monaural Sensing .................................................................... 86
6.3.1 Inverting the Echo Formation Process ................................. 87
6.3.2 Extraction of Information: Cochlear Processing .................. 87
6.4 Multi-Aural Sensing ............................................................... 88
6.4.1 Echo Amplitude and Echo Arrival Time: Two transmitters, Two receivers .................................. 89
6.4.1.1 Sensor Setup ................................................................... 89
6.4.1.2 Localisation of Planes and Corners ................................. 90
6.4.1.3 Recognition of Planes and Corners ................................. 91
6.4.2 Echo Arrival Time Information: Two Transmitters, Two Receivers ............................................ 93
6.4.2.1 Sensor Setup ................................................................... 94
6.4.2.2 Localisation of Edges and Planes/Corners ....................... 94
6.4.2.3 Recognition of Edges, Planes and Corners ....................... 95
6.4.3 Echo Arrival Time Information: One Transmitter, Three Receivers ................................................ 97
6.4.3.1 Sensor Setup ................................................................... 97
6.4.3.2 Localisation of Edges and Planes/Corners ....................... 98
6.4.3.3 Recognition of Edges, Planes and Corners ....................... 99
6.4.3.4 Localisation of Curved Reflectors ................................... 101
6.4.4 One Transmitter, Two Receivers: 3 Dimensional World
Model .................................................. 103
   6.4.4.1 Sensor Setup .................................. 104
   6.4.4.2 Localisation of a Point-Like Reflector in 3D . 105

6.5 Summary ............................................. 109

Chapter 7 Map Building from Range Data Using
Mathematical Morphology .............................. 111

7.1 Introduction ......................................... 111
7.2 Basics of Sonar Sensing ............................ 114
7.3 Processing of the Sonar Data ..................... 115
   7.3.1 Morphological Processing .................... 117
   7.3.2 Curve Fitting .................................. 119
   7.3.3 Simulation Results ............................ 121
      7.3.3.1 Linear Arrays ............................ 121
      7.3.3.2 Circular Arrays .......................... 122
      7.3.3.3 Arbitrarily-Distributed Sensors ......... 122
7.4 Experimental Verification .......................... 125
   7.4.1 System Description ............................ 125
   7.4.2 Experimental Results ........................ 128
   7.4.3 Computational Cost of the Method .......... 133
7.5 Discussion and Conclusions ........................ 133

Chapter 8 Millimetre Wave Radar for Robotics .... 137

8.1 Background .......................................... 137
8.2 When to Use Millimetre Wave Radar ............ 138
8.3 Millimetre Wave Radar Principles ............... 140
   8.3.1 Range Resolution ............................. 140
   8.3.2 Pulse Compression ............................ 141
   8.3.3 Stepped Frequency ............................ 142
   8.3.4 Frequency Modulated Continuous Wave ...... 143
   8.3.5 Angular Resolution and Antennas .......... 146
   8.3.6 Scanning and Imaging ....................... 148
      8.3.6.1 Mechanical Scanning ..................... 148
      8.3.6.2 Electronic Scanning ...................... 148
      8.3.6.3 Image Representation .................... 149
8.4 Review of Work Done in the Field ............... 151
   8.4.1 Indoor Applications ......................... 151
8.4.1.1 Technische Universität München .............. 151
8.4.1.2 St. Petersburg State Technical University .... 153
8.4.2 Outdoor Applications .......................... 153
8.4.2.1 Robotics Institute: Carnegie Mellon University 153
8.4.2.2 Helsinki University of Technology ............ 154
8.4.2.3 Australian Centre for Field Robotics: Sydney University ....... 154
8.5 Airborne Radar Systems .......................... 156
8.5.1 Imaging Range and Resolution .................. 156
8.5.2 Results ..................................... 158
8.6 Waypoint Navigation Process ...................... 159
8.6.1 Navigation Error Estimation ...................... 161
8.6.2 Results ..................................... 161
8.7 Summary ...................................... 162

Chapter 9 Optoelectronic Range Sensors 165
9.1 Introduction ...................................... 165
9.2 Range-Finders .................................... 165
  9.2.1 Introduction .................................. 165
9.3 Radiometric Design ............................... 166
  9.3.1 Specular Reflection ........................... 168
  9.3.2 Diffuse Reflection ............................. 171
  9.3.3 The Emitter and Detector ...................... 172
  9.3.4 Optical Geometry ............................. 174
9.4 Ranging Sensors .................................. 177
  9.4.1 Triangulation .................................. 177
  9.4.2 Lidar ......................................... 180
    9.4.2.1 Pulsed Modulation ......................... 181
    9.4.2.2 Amplitude Modulation Continuous Wave .... 182
    9.4.2.3 Frequency Modulation Continuous Wave .... 184
9.5 Scanning Range-Finders ........................... 186
  9.5.1 Introduction .................................. 186
  9.5.2 Scanning Methods ............................ 186
    9.5.2.1 Holographic Scanners ...................... 187
    9.5.2.2 Acousto-Optic Scanners .................... 187
  9.5.3 Some Scanning Sensors ......................... 188
    9.5.3.1 The Sick Sensor: Pulsed Lidar .............. 188
    9.5.3.2 AMCW Lidar Sensors ....................... 188
12.1.3 A Robot Platform with Active Vision ........................................... 242
12.2 Scene Features .................................................................................. 244
  12.2.1 Detecting Features ........................................................................ 244
  12.2.2 Searching for and Matching Features ............................................. 247
  12.2.3 Other Feature Types ....................................................................... 249
12.3 Fixation ............................................................................................... 251
  12.3.1 Acquiring Features .......................................................................... 251
  12.3.2 The Accuracy of Fixated Measurements .......................................... 252
12.4 Localisation and Map-Building ........................................................... 254
  12.4.1 An Extended Experiment ................................................................. 254
12.5 Continuous Feature Tracking ............................................................... 259
12.6 A Fixation Strategy for Localisation ....................................................... 261
  12.6.1 Choosing from Known Features ..................................................... 262
  12.6.2 Experiments .................................................................................. 263
12.7 Steering Control and Context-Based Navigation .................................... 266
  12.7.1 Steering a Twisting Course ............................................................ 266
12.8 Summary ............................................................................................. 269

Chapter 13 Strategies for Active Sensor Management ............................... 271
13.1 Introduction ......................................................................................... 271
13.2 Simple Signal Processing Tools ........................................................... 275
13.3 Reconfigurable Sensors and Signal Processing Tools ............................ 278
13.4 A Sensor-Centred Image Segmentation Algorithm ................................ 282
13.5 Signal Processing Tool Selection Strategies ......................................... 284
13.6 Dynamic Signal Processing Tool Scheduling ....................................... 287
13.7 Conclusions ....................................................................................... 289

Bibliography ............................................................................................ 291

Appendix A: Contact Details of Authors .................................................. 307

Index .......................................................................................................... 311
PART I

GENERIC ISSUES
This page is intentionally left blank
Chapter 1

Introduction

Penelope Probert Smith

Research into mobile robotics is concerned fundamentally with complexity and change. Apposite and timely sensing is crucial.

The fundamental aim is to provide complex systems with the ability to react and to adapt to diverse environments autonomously. A mobile robot has four needs above all:

- The ability to perceive the environment, and to deliberate about its own relationship to the environment
- The ability to reason spatially to plan, for local route finding and to fulfill a task or mission
- A reliable software architecture which provides rapid communication between essential processes
- Good hardware and locomotion control

Complexity is the issue which drives mobile robot research. The complexities of the interaction of software and hardware, the multi-level reasoning which provides fast reactive capability but can also extracts efficient planning in complex environments; these are the sorts of issue which provide the excitement and challenges.

This book deals both with the technologies of sensing and with the structure of sensing systems. Sensing cannot be considered in isolation. Sensing both serves and directs other functionality in the robot. We must take a holistic view of robotics, viewing sensing within the whole system.

1.1 Architectures for Planning and Perception

Early work in robotics failed to do this and separated out the task of sensing from planning. Its aim was to optimise performance on a global scale and for this it needed information to be as complete as possible. The sensing
goal was completeness: to build up a full model of the world. The model was then made available to all planning and control tasks.

The "world model" was usually described geometrically, in a reference frame external to the robot and to any particular task. In this way it was available to many tasks. The world model was kept in a global data structure sometimes called a blackboard [Harmon 86; Nii 86], which could be updated by some processes and read by all (figure 1.1). The blackboard acted as a data repository for all the sensing processes. Planning was strictly hierarchical, with all planning processes having access to the same data. With full knowledge of the world around it, the robot, placed somewhere within this map, could search for an optimal plan.

![Fig. 1.1 The blackboard architecture](image)

The success of the plan depends on the integrity of the map. Stereo vision was typically used as the major sensing modality, since it can in theory build up a full 3-dimensional model of the world around. Typical processing moves from edge detection, to correspondence between the two
Architectures for Planning and Perception

Cameras, to feature representation. Vision was sometimes augmented by a laser stripe range-finder. Because data was held in a global representation, it was difficult to include data which is naturally extracted in a different representation, such as that from sonar.

The method provided impractical, largely for two reasons.

- The first was uncertainty in sensing. For good performance, stereo vision relies on good lighting, good calibration. Occlusion means not only that an image is obscured in one camera, but that the correspondence between camera images may fail. It was impossible to hold data in sufficient detail, of suitable format and reliability to suit all tasks.
- The second was responsivity. The logical and physical separation between the sensors and the robot control leads to poor reactivity to external changes. The time for deliberation between sensing and action was too large - both because of the labour in processing a complete map, the size of the data structure called the blackboard, and because of the need for many processes to access the world model. The ability to react quickly to changes in the environment was lost.

There are various ways in which this architecture can be made more practical - for example using a distributed blackboard, including sensors for obstacle avoidance which communicate through interrupts. However an alternative paradigm was introduced which threw away the idea of completeness on a global scale to emphasise local reactivity. The subsumption architecture [Brooks 86] changed the relationship between sensing and planning. It abandoned any idea of an information repository. It followed a biological analogy, from observing that a set of apparently individually simple behaviours can result in significant achievement - consider, for example, the constructions created by a colony of ants! The subsumption architecture typically put together a set of behaviours such as “avoid obstacle”, “follow wall”, “go straight ahead”. Tasks were designed to operate independently, each having access to a set of simple sensors which it directed as required (figure 1.2). A hierarchy between the layers determined precedence when there was conflict.

Practical problems with this approach may arise from complexity of communications. More fundamentally the lack of consistency between the
perception of each level may lead to cyclic behaviour. In addition it has not been proved possible yet to meet task directed planning at a high level. The main interest is to investigate the synthesis of biological systems. The field of evolutionary robotics examines the synthesis of complex behaviours from a number of simple functions served by simple sensors.

Evolutionary robotics is an example of “bottom up” design - create some simple functions and see how they combine. The architectures associated with the blackboard are “top down” - specify the top level requirements and design a system to fulfill them. This of course is the approach used in engineering systems, and its advantage is that behaviour is predictable and purposeful.

The best practice in robotic systems uses top down design, but draws from the subsumption architecture the idea of sensors designed to serve specific tasks. Emphasis is on allowing the robot to be reactive - to react rapidly to new events - at the local level, but deliberative at task planning levels. Sensors are active participants in decision making and planning. Rather than providing as much information as possible, in some generic
format, the sensor attempts to provide information according to the need of a planning task.

The concept of an active sensor in robotics is of the sensor as participant. The sensor contains not just hardware, but reasoning too. The architecture is decentralised, with the sensor itself containing not just processing algorithms but also a decision process (figure 1.3). The sensor may choose whether to take part in a task, which parts of the environment to examine, which information to obtain. By concentrating on the provision of timely data, the active sensor can provide rapid response to new environments and unexpected changes.

This definition of an active sensor includes the type of sensor traditionally deemed active - those which probe particular parts of the environment with radiation (sound or electromagnetic waves). Sensors active in this sense are especially important for local planning where fast reaction to change is needed, since they measure range directly.

Good hardware and basic processing are essential in the sensor. Much of
this book is concerned with describing hardware and algorithms. Because of their importance, sonar and opto-electronic range sensors are discussed in special detail.

1.2 Range Sensing Technologies

Sensors for local planning need to return information primarily on range. The commonest technology to measure range directly uses echo detection. The earliest type of sensor of this type was sonar, developed in the first world war to determine the position of the sea floor for submarine navigation. Sonar is still the main sensor in use for underwater robots. Sound has low attenuation in water and at low enough frequencies will propagate many miles. In mobile robotics low cost air-borne sonar devices have been popular for many years for ranging and obstacle avoidance. Their main drawback is that they are limited by current technology to a range of between 5m and 10m. Underwater robotics uses sonar with frequency between a few hundred kHz and 2MHz. The higher frequencies have better resolution; the lower ones travel further.

Conventionally the technology which complements sonar in air is radar, which was developed most fully in the Second World War. Early systems used valve technology to produce a signal at hundreds of MHz to a few GHz, with frequency modulation and pulse compression to improve resolution. Sonar analogues of these methods are discussed in chapter 4. Millimetre wave radar systems use frequencies at around 90-100GHz. A few experiment on using this type of radar for outdoor mobile robots have been reported, but the benefits from the technology are far from proven.

Another type of active range sensor, developed more recently, uses optical frequencies. Optical range sensors have become more popular with the widespread availability of laser and light emitting diodes. Although methods based on imaging following structured illumination of the scene have been available for some time, their use has been confined largely to manufacturing inspection. The maximum range is low and limited by ambient light levels. A number of range sensors based on time of flight measurement were developed by mobile robotics groups over the last decade and a half but only recently have suitable commercial models become available.

All these sensors rely on the propagation of waves and the detection of echos from the transmitted wavefront. We can understand the broad dif-
ferences between the different types of sensing technologies, by considering the properties of these waves. In spite of the differences in frequency and type (transverse or longitudinal), they can all, electromagnetic and sound, be described by a similar mathematical expression. Difference in how they perceive a scene arises largely because of differences in wavelength.

Some typical systems used in mobile robots are summarised in Table 1.1. The figures given are all for air-borne sensors except where noted. From this table we see that sonar and radar operate at broadly similar wavelengths, whereas light has a wavelength several orders of magnitude less.

<table>
<thead>
<tr>
<th>type of system</th>
<th>velocity of wave</th>
<th>wavelength</th>
</tr>
</thead>
<tbody>
<tr>
<td>air-borne sonar (45kHz)</td>
<td>340msec⁻¹</td>
<td>7.56mm</td>
</tr>
<tr>
<td>underwater sonar (200kHz)</td>
<td>1470msec⁻¹</td>
<td>7.35mm</td>
</tr>
<tr>
<td>94GHz radar</td>
<td>3 × 10⁸msec⁻¹</td>
<td>3.2mm</td>
</tr>
<tr>
<td>visible (red) optical sensor</td>
<td>3 × 10⁸msec⁻¹</td>
<td>0.8μm</td>
</tr>
<tr>
<td>near infra-red optical sensor</td>
<td>3 × 10⁸msec⁻¹</td>
<td>1.55μm</td>
</tr>
</tbody>
</table>

Table 1.1 A comparison of some common sensing technologies

1.3 Planning Demands

To finish this chapter we consider some sensing needs of the local planner. The key requirement is to plan a route from local knowledge and to oversee the movement of the robot along that route. To do this it needs both to produce a local map of the immediate area and to keep track of where the robot is in that map. This is the process of joint mapping and localisation.

Feature recognition is key to both mapping and localisation. Success in planning depends on the number of features which can be detected and on their reliability. It is governed by:

- The type of feature which can be detected by the sensor. Whereas for an opto-electronic range sensor a complete line model may be built up, other sensors may only pick up point features.
- The success of the correspondence process - the process by which a feature measured from one viewpoint can be matched to the same
Introduction

feature from another (for example from a new robot position)

• The number of features which can be measured. Most active sensors have a restricted field of view and have to be rotated mechanically to gain a new viewpoint. This makes movement slow so attention is normally focussed on just a few positions.

Feature recognition alone however is not enough. Crucial to the whole process of mapping and localisation is the placing of features relative to one another and to the robot as it moves around in the environment. This requires correct handling of error, both in the sensors and in the robot itself. A successful strategy requires a method which incorporates these errors as readings are built up over time. The most common algorithm to use is one based on the extended Kalman filter, the EKF. The EKF is a non-linear extension to the Kalman filter.

The EKF provides an estimate of system state. What is included as state depends on the application. In the mapping context, the state is likely to include both feature and robot position (x-y position, bearing, orientation) and possibly robot velocity. An uncertainty is associated with each of these, related to errors in the measurement and model.

The EKF updates the state between two time instants in a two part process:

• First it uses a model to predicts how the state varies from one instant to the next

• Second it takes account of any measurements available.

A covariance is associated with each of these. The measurement and prediction, together with their covariances, are then used to determine an estimate of state, together with its covariance.

The EKF may be used in many ways. In chapter 11 we describe its use for extracting a line segment. It is commonly used in sensor integration, either using readings from the same sensor or from ones different. In chapter 5 it is used for mapping and in chapter 12 it is used in a joint mapping and localisation problem.

Joint mapping and localisation is central to the local planner and one of the most common operations undertaken in robots which are task orientated. The process places particular demands on the sensor processing because typically it takes place over some time and over a considerable distance moved by the robot. Without a proper formulation of error the map
may become inconsistent. The next chapter is devoted to this problem.
This page is intentionally left blank
Chapter 2

The Mapping and Localisation Problem

Andrew J. Davison

2.1 Simultaneous Localisation and Map Building

When a robot needs to move *repeatably* in surroundings about which it has little or no prior knowledge, calculating ego-motion using just *dead-reckoning* (where for instance odometry counts the number of turns of each of its wheels), is not sufficient: estimates based solely on such measurements of relative motion will have errors which compound over time to drift steadily from ground truth. It is necessary that the robot uses its outward-looking sensors to identify landmarks in the surroundings, and then use measurements of the relative positions of these landmarks from future points on its movement path to lock down its localisation estimates — essentially, it must make a map of features in the scene and then estimate its location relative to this map.

Anything whose relative location to a robot can be repeatably measured using a sensor can be thought of as a "feature" which can be put into a map. Typically, sensors such as sonar, vision or laser range-finders can be mounted on robots and used to identify geometrical features such as points, lines and planes in a scene. In general, a combination of sensors making measurements of various feature types can be used together in map-building, all contributing to localisation. This combination of capabilities is sometimes referred to as sensor fusion.

The most important thing to consider when formulating a map-building algorithm is that:

- All sensor measurements are uncertain.

Maps must reflect this fact if they are to be useful, and be formulated in a probabilistic fashion. Two distinct approaches to map-building are possible, depending on the required application. One is to build a map based on data
acquired during a preliminary guided visit to an environment, processing all the measurements obtained afterwards and offline to produce a map for future use. Successful batch methods of this type in robotics (e.g. [Thrun et al. 98]) share many similarities with state-of-the-art “structure from motion” techniques in computer vision [Torr et al. 98; Pollefeys et al. 98]. Algorithms to build maps in this way are able to make use of sophisticated optimisation algorithms to make the best use of all the data.

However, in robot applications we are often not able to afford the luxury of pre-mapping all areas which the robot will be required to visit, and therefore must consider the more challenging problem of sequential localisation and map-building. If a robot is to enter an unknown area and then proceed immediately to move around it and react to new information as it arrives, batch algorithms are unsuitable due to their large computational complexity. Speaking more specifically, in sequential map-building we are limited by the fact that as each new piece of data arrives, it must be possible to incorporate the new information into the map in an amount of processing time bounded by a constant, since the robot must take action and the next piece of data will soon be arriving. This in turn requires that our representation of all knowledge obtained up to the current time must be represented by a bounded amount of data: that amount cannot grow with time.

Sequential map-building is therefore the process of propagating through time a probabilistic estimate of the current state of a map and the robot’s location relative to it. In the following, we will look at some of the general properties of the map-building process, and then discuss in more detail the most common approach taken in sequential map-building, using the Extended Kalman Filter.

2.1.1 The Map-Building Process

A map which is made by a traveller or robot who does not have some external measure of ego-motion is fundamentally limited in its accuracy. The problem is caused by the compound errors of successive measurements. Consider, for example, a human given the task of drawing a very long, straight line on the ground, but equipped with only a 30cm ruler, and unable to use any external references such as a compass or the bearing of the sun. The first few metres would be easy, since it would be possible to look back to the start of the line when aligning the ruler to draw a new
section. Once this had gone out of view, though, only the recently drawn nearby segment would be available for reference. Any small error in the alignment of this segment would lead to a misalignment of new additions. At a large distance from the starting point, the cumulative uncertainty will be great, and it will be impossible to say with any certainty whether the parts of line currently being drawn were parallel to the original direction. Changing the measurement process could improve matters: if, for instance, flags could be placed at regular intervals along the line which were visible from a long way, then correct alignment could be better achieved over longer distances. However, eventually the original flags would disappear from view and errors would accumulate — just at a slower rate than before.

Something similar will happen in a robot map-building system, where at a certain time measurements can be made of only a certain set of features which are visible from the current position — probably these will in general be those that are nearby, but there are usually other criteria such as occlusion or maximum viewing angle. It will be possible to be confident about the robot’s position relative to the features which can currently be seen, but decreasingly so as features which have been measured in the more distant past are considered. A properly formulated map-building algorithm should reflect this if the maps generated are to be consistent and useful for extended periods of navigation.

2.1.2 The Coupling of Map Estimates

Autonomous map-building is a process which must be carefully undertaken, since the processes of building a map of an area and calculating location relative to that map are inherently coupled. Many early approaches [Durrant-Whyte 94; Harris 92] to online map-building took simple approaches to representing the state and its uncertainty; the locations of the moving robot in the world and features were stored and updated independently, perhaps using multiple Kalman Filters. However, if any type of long-term motion is attempted, these methods prove to be deficient: though they produce good estimates of instantaneous motion, they do not take account of the interdependence of the estimates of different quantities, and maps are seen to drift away from ground truth in a systematic way, as can be seen in the experiments of the authors referenced above. They are not able to produce sensible estimates for long runs where previously seen features may be revisited after periods of neglect, an action that allows drifting estimates to
The Mapping and Localisation Problem

(1) Initialise feature A. (2) Drive forward. (3) Initialise B. and C.
(4) Drive back. (5) Re-measure A. (6) Re-measure B.

Fig. 2.1 Six steps in a example of sequential map-building, where a robot moving in two dimensions is assumed to have a fairly accurate sensor allowing it to detect the relative location of point features, and less accurate odometry for dead-reckoning motion estimation. Black points are the true locations of environmental features, and grey areas represent uncertain estimates of the feature and robot positions.

be corrected.

To attempt to give a flavour of the interdependence of estimates in sequential map-building, and emphasise that it is important to estimate robot and feature positions together, steps from a simple scenario are depicted in Figure 2.1. The sequence of robot behaviour here is not intended to be optimal; the point is that a map-building algorithm should be able to cope with arbitrary actions and make use of all the information it obtains.

In (1), a robot is dropped into an environment of which it has no prior knowledge. Defining a coordinate frame at this starting position, it uses a sensor to identify feature A and measure its position. The sensor is quite accurate, but there is some uncertainty in this measurement which transposes into the small grey area representing the uncertainty in the estimate of the feature's position.

The robot drives forward in (2), during this time making an estimate of its motion using dead-reckoning (for instance counting the turns of its wheels). This type of motion estimation is notoriously inaccurate and
causes motion uncertainties which grow without bound over time, and this is reflected in the large uncertainty region around the robot representing its estimate of its position. In (3), the robot makes initial measurements of features B and C. Since the robot's own position estimate is uncertain at this time, its estimates of the locations of B and C have large uncertainty regions, equivalent to the robot position uncertainty plus the smaller sensor measurement uncertainty. However, although it cannot be represented in the diagram, the estimates in the locations of the robot, B and C are all coupled at this point. Their relative positions are quite well known; what is uncertain is the position of the group as a whole.

The robot turns and drives back to near its starting position in (4). During this motion its estimate of its own position, again updated with dead-reckoning, grows even more uncertain. In (5) though, re-measuring feature A, whose absolute location is well known, allows the robot dramatically to improve its position estimate. The important thing to notice is that this measurement also improves the estimate of the locations of features B and C. Although the robot had driven farther since first measuring them, estimates of these feature positions were still partially coupled to the robot state, so improving the robot estimate also upgrades the feature estimates. The feature estimates are further improved in (6), where the robot directly re-measures feature B. This measurement, while of course improving the estimate of B, also improves C due to their interdependence (the relative locations of B and C are well known).

At this stage, all estimates are quite good and the robot has built a useful map. It is important to understand that this has happened with quite a small number of measurements because use has been made of the coupling between estimates.

2.1.3 Simultaneous Localisation and Map-Building with the EKF

As we have seen, sequential localisation and map-building must be treated as a statistical problem, and in the broadest language its solution involves the propagation through time of a multi-dimensional probability distribution representing current knowledge about the state of the robot and map features, estimates of which will generally be strongly coupled. Representing general probability distributions in multiple dimensions is a difficult task: due to non-linearity in the movement and measurement processes of
most robots, these distributions will have complex shapes which are not easily parameterised. One approach which has recently achieved great success in computer vision research is to represent probability distributions by a population of discrete samples or "particles" [Isard and Blake 96] — this method has the advantage of being able to represent any shape of distribution, including those which have multiple peaks. However, particle methods are computationally expensive, and in particular their performance scales badly with the number of dimensions of parameter space, and are thus currently unsuited to the map-building problem with its large number of unknown parameters.

More feasible is to consider an approximation to the shape of probability distributions in such a way which makes the computation tractable. The Extended Kalman Filter is such a method, forcing a gaussian shape on the probability density of all estimated quantities. However, this is generally a good approximation to make in many systems, and the EKF has been repeatedly proven in robot localisation algorithms. Called "Stochastic Mapping" in its first correctly-formulated application to robot map-building [Smith et al 87], the EKF has been implemented successfully in different scenarios by other researchers [Castellanos 98; Chong and Kleeman 99a; Davison 98; Davison and Murray 98; Durrant-Whyte et al 99; Kwon and Lee 99]. The key to these approaches is using a single state vector to store together estimates of the robot position and those of feature positions, and subsequently the ability to correctly propagate all the coupling between estimates which arises in map-building.

Current estimates of the locations of the robot and the scene features which are known about are stored in the system state vector $\hat{x}$, and the uncertainty of the estimates in the covariance matrix $P$. These are partitioned as follows:

$$
\hat{x} = \begin{pmatrix} \hat{x}_v \\ \hat{y}_1 \\ \hat{y}_2 \\ \vdots \end{pmatrix}, \quad P = \begin{bmatrix} P_{xx} & P_{xy_1} & P_{xy_2} & \cdots \\ P_{yx_1} & P_{y_1y_1} & P_{y_1y_2} & \cdots \\ P_{yx_2} & P_{y_2y_1} & P_{y_2y_2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.
$$

(2.1)

$\hat{x}_v$ is a vector containing the robot state estimate, and $\hat{y}_i$ the estimated state of the $i$th feature. The number of parameters in each of these vectors depends on the specifics of the robot system under consideration and what kind of features it is observing. $P$ is square and symmetric, with width and
height equal to the size of \( \mathbf{x} \). \( \hat{\mathbf{x}} \) and \( \mathbf{P} \) will change in size as features are added or deleted from the map.

If the robot starts with no knowledge about its surroundings, initialisation of the map takes place by zeroing the robot state and covariance (i.e. defining the world coordinate frame to be at the robot’s start position), and with no features in the state vector. Alternatively, if some features are known as prior knowledge, they can be inserted into the map straight away and the uncertainty in the robot’s position relative to them used to define an initial \( P_{xx} \).

The map data is then updated repeatedly in two steps:

1. Prediction, or the process stage of the filter. When the robot moves, a motion model provides a new estimate of its new position, and also a new uncertainty in its location. In the prediction step, positional uncertainty always increases.

2. Measurement update, when measurements of one or more features are incorporated leads to an overall decrease in the uncertainty in the map.

These two steps can be thought of as blind movement followed by measurement. The full uncertainty information contained in the map allows many intelligent choices to be made about things such as which features to make measurements of, where to search for feature correspondences and how the robot should move to remain safely close to a given path.

The current challenge of map-building research is how to deal with the computational complexity of stochastic localisation and mapping (SLAM) algorithms for real-time applications. Although the EKF is a relatively efficient algorithm, when maps grow very large the coupling between all estimates means that performance will tail off and recently many approaches have been proposed for increasing efficiency, including splitting regions into submaps [Chong and Kleeman 99a; Leonard and Feder 99] and the postponement of calculations [Davison 98].

An example of the mapping and localisation algorithm in practice is shown in chapter 12.
PART II

MILLIMETRE WAVE SENSORS
Chapter 3
Perception at Millimetre Wavelengths

Penelope Probert Smith

The millimetre wave sensor most common in mobile robotics is sonar. It normally operates at around 45kHz (wavelength about 7mm), a frequency which provides a range of about 5m. High frequency radar has some similar characteristics. Radar has the advantage of far lower attenuation in air, but is expensive. Frequencies of 77GHz and 94GHz, for which the wavelength is a few millimetres, has been investigated for outdoor navigation [Clark and Durrant-Whyte 98a; Boehnke 98]. However systems suitable for robotics are less advanced than sonar systems, partly owing to their expense. Therefore most of the discussion in this chapter relates to sonar.

Sensors with wavelengths in the millimetre range view the world rather differently than those which use light. The differences arise because wavelength is comparable to the dimensions both of the transducer itself and of variability in the surface of typical reflectors. Two major effects result.

- An interference pattern results from the radiation across the transducer aperture, which leads to relatively wide beams with peaks and troughs in power as angle varies. This has consequences both in scene interpretation and in resolution:

  1. As the beam becomes wider the angular resolution of a single reading decreases. The radiation typically extends over a fan of 20 degrees or more. This is useful for providing a sensor for obstacle avoidance, but means that a single reading provides poor angular resolution. However as we see later, a wide beam is actually necessary given the reflective properties of surfaces, and can be exploited in systems using multiple transducers (see chapter 6).

  2. Wide beams have significant losses from beam spreading, so the signal at the receiver is reduced. The consequent reduction in signal to noise ratio leads to deterioration in range
resolution.

- Many surfaces, which to our eyes look rough, appear smooth and mirror-line at millimetre wavelengths. Interpretation is needed to relate the signals received to the actual geometry of a feature. In addition the amount of radiation returned from more complex objects is related more to the angle and geometry of their surfaces than to their physical size.

There are also factors directly related to wavelength which affect time resolution. We consider these in chapter 4.

### 3.1 Sensor Operation

The millimetre wave sensors we are considering in this chapter work through echo detection. The following stages take place in such a sensor:

- A transducer converts an electrical signal into another form of energy (sound or electromagnetic)
- The energy is launched as a wave across an aperture or antenna
- It propagates through the atmosphere until it meets a reflector
- A proportion is backscattered to the receiver and gathered by an antenna
- The energy is converted back into an electrical signal and stored in a computer

This process is often described in terms of power. The radar equation relates the power received and transmitted:

\[
I_r = I_t G_t A_r \sigma L^\text{power}_{\text{at}}
\]  

(3.1)

In this equation, \( I_r \) and \( I_t \) are, respectively, the transmitted and received power or intensity. \( \sigma \) is a quantity known as the scattering cross section of the object, which describes how much power is scattered back to the receiver when power is incident from the transmitter. \( L^\text{power}_{\text{at}} \) describes attenuation in the atmosphere from loss and spreading. \( G_t \) is the gain of the transmitting antenna in the direction of the target, \( \theta \), and is related to its angular selectivity. \( A_r \) is the effective area of the antenna, and is defined so that the power received is the product of the power intensity per unit
area and $A_r$. For apertures $A_r$ is closely related to physical area. In general is it related to the gain $G_r$ of the same antenna in transmission through a thermodynamic relationship:

$$G_r = \frac{4\pi A_r}{\lambda^2}$$

(3.2)

The radar equation, as its name suggests, is relevant in radar and communications technologies. It usually assumes that the antenna is aligned so the gain is maximum.

In radar the signal at the receiver (voltage or current) varies linearly with power for many receiver technologies and so it is natural to think in terms of power transfer. However for sonar it is more physical to think about the propagation of the pressure wave since for most technologies the signal generated at the receiver is more closely proportional to pressure than to intensity (proportional to the square of pressure). For sonar therefore we write a modified equation in terms of amplitude:

$$P_r = K P_t \Psi_t(\theta, \phi) \Psi_r(\theta, \phi) \Gamma L_{at}^{\text{pressure}}$$

(3.3)

where $P_r$ and $P_t$ are the amplitudes of the pressure waves at transmitter and receiver, $\Gamma$ represents the proportion of power incident on a reflector scattered to the receiver, $\Psi_r$ and $\Psi_t$ are the angular radiation patterns (beam patterns) of the transducer and receiver and $K$ is a constant. $L_{at}^{\text{pressure}}$ describes attenuation in the atmosphere from loss and spreading.

The beam pattern and the gain are closely related. In the direction of maximum:

$$G = kD$$

(3.4)

where $k$ is an efficiency factor and we define $D$, the directivity:

$$D = \frac{4\pi |\Psi_r^2(\theta, \phi)|_{\text{max}}}{\int_4 \int_\pi |\Psi_r^2(\theta, \phi)|d\Omega}$$

(3.5)

where the integral is over the solid angle $\Omega$. [Kraus 84]

Both these equations ignore any modification of the signal by the transducers except for spatial effects resulting from the antenna (chapters 4 and 6 discuss other effects).

In the rest of this chapter we investigate the parameters in equations 3.1 and 3.3 more closely.
3.2 The Sensor

In concept the sensor consists of two parts: a transducer to produce wave energy, and an aperture or antenna to radiate or receive such energy. However these may be integrated into a single component.

Sonar sensors designed to operate in air normally use electrostatic transducers (in contrast, sensors for operation in liquids are normally based on piezo-electric crystals). The Polaroid transducer, a common commercial robot sensor, combines the transducer with a circular antenna. A schematic diagram is given in figure 3.1. We can think of the sensor as a capacitance consisting of a metal membrane and a metal backplate, which stores charge when a voltage is applied between them. To operate the sensor as a transmitter a radio frequency voltage drive signal is applied, of the same frequency as the ultrasound signal desired. A changing electrostatic force between the membrane and backplate results, causing the membrane to vibrate, and launching an ultrasound wave across the circular aperture. When used as a receiver the membrane is biased at about 200V. As the pressure wave hits the aperture, the foil is deflected and modulates the capacitance. The resulting modulation in bias voltage is approximately proportional to the amplitude of the pressure wave (using small signal approximations).

![Fig. 3.1 A schematic diagram showing the operation of a typical air-coupled sensor](image)

In millimetre wave radar the transducers in transmitter and receiver may be semiconductor devices, either transistors or diodes. Microwave transistors operate similarly to lower frequency devices but are designed to have very low parasitic capacitances and fast transit times. The IMPATT diode uses properties of the avalanche breakdown region of semiconductors
3.3 Antenna Properties

The function of the antenna is to receive or radiate power. We can view an antenna as being made up of many tiny radiating elements, each of which radiates a wave. For sonar this is a pressure wave; for radar an electromagnetic wave. Interference between elements on the antenna means that the total wave generated is not isotropic but has a distinctive angular pattern, known as the radiation pattern. The variation in power transmitted with angle which this pattern describes gives rise to the beam pattern or directional selectivity of the antenna.

Typically the pattern can be divided into two regions. In the first, known as the near field, the beam does not diverge significantly. However the amplitude of the wave decays according to the second power of distance so energy extends only a few wavelengths from the antenna. The second part, which is of more interest in sonar or radar, is known as the far field. The amplitude exhibits an inverse law with distance (the power the familiar inverse square law) and the beam is divergent. All the discussion which follows relates to the far field.

The shape of the radiation pattern depends on the shape and size of the antenna. The natural world provides some examples. Compare the shape of the ear (sounds has long wavelength) with the geometrical simplicity of the eye (light has short wavelength). The folds and flaps in our ears provide both direction and frequency filtering at these longer wavelengths. Bats are particularly interesting. Since ultrasound is their main sensor for navigation and capturing food, they have evolved with very specialised auditory systems. Both their ears, which act as receivers, and their noses, which act as transmitters, are shaped and provided with precise movements to optimise the task of detecting and capturing very small prey in particular environments. Figure 3.2 shows a few examples.

It is not possible to imitate such complexity easily and typical ultrasound transducers are simply circular. The Polaroid transducer for example is shown in figure 3.3. The aperture has a width of about 40mm and
Perception at Millimetre Wavelengths

Fig. 3.2 Three species of bat, showing some types of facial shapes which provide highly directional sonar. The horseshoe bat (centre), so called because of the shape of the growth on its nose, moves this growth as it flies to direct the ultrasound emitted through the nose. The sword bat too has a complex nose and ears whereas the tent bat uses size of ear to make up for the simpler shape. All bats flap their ears too whilst flying to locate sound better

is designed to operate between 45 kHz and 90kHz (wavelengths between about 7.5 and 3.8 mm). Most of the power is concentrated within a central lobe, which extends over an angle of about 26° to 13°, for this range of wavelength. Circular apertures are used in radar as well, especially when they are fabricated as patch antenna, a flat antenna etched onto microstrip. A more focused beam is produced by a parabolic antenna, also shown in figure 3.3. This example has diameter 100mm, and is designed to operate at 94GHz (wavelength about 3mm). The width of the centre lobe is similar for a parabola and a circular dish of the same dimensions, but the parabola has smaller side-lobes: for example the peak power in the first side-lobe is less by about 4dB.

3.3.1 The Circular Antenna

The circular antenna is the most important in robotics largely because of ease of fabrication. The interference pattern which arises from radiation at small adjacent elements across the antenna results in a far field radiation pattern as follows (considering just one dimension):
Fig. 3.3 On the left, a Polaroid transducer taken from a robot, on a circular mount. The round aperture in the holder is the transducer itself and the electronics is in the box. On the right, a commercial 94GHz dish. Note the waveguide feed at the centre.

\[ \Psi(\theta) = \Psi_0 \frac{|J_1(ka \sin \theta)|}{ka \sin \theta} \]  

(3.6)

In this expression the antenna is of radius \( a \) and \( k = 2\pi/\lambda \), where \( \lambda \) is the wavelength of the pressure wave radiated. \( \Psi(\theta) \) describes the variation of pressure transmitted with angle \( \theta \) and \( \Psi_0 \) is a constant. \( J_1 \) is a Bessel function of the first kind. Bessel functions are a set of functions which describe several wave phenomena. The magnitude of the Bessel function \( J_1 \) together with a typical radiation pattern is shown in figure 3.4. The oscillatory nature of the Bessel function results in a central lobe and a set of side-lobes in the radiation pattern. The central lobe extends to an angle:

\[ \theta \approx \sin \theta = \frac{0.61\lambda}{a} \]  

(3.7)

Because Bessel functions are cumbersome to work with algebraically, equation 3.6 is often approximated by an exponential expression

\[ \Psi_t(\theta) = \Psi_0 \exp\left(-\frac{2\theta^2}{\theta_0^2}\right) \]  

(3.8)
Various values of $\theta_0$ can be chosen, depending on the part of the pattern which should be matched. In general the extent of the first lobe is the most important parameter affecting the variation of the signal received with angle. Typically we choose $\theta_0$ to match this region of the pattern [Kuc and Viard 87]:

$$\theta_0 = \arcsin\left\{0.61 \frac{\lambda}{a}\right\}$$  \hspace{1cm} (3.9)

From this equation and figure 3.4 we note two points:

1. The extent of the central lobe depends on the ratio of characteristic dimension (in this case the radius) to wavelength. Large antenna provide more directed beams.
2. The radiation pattern is oscillatory and has significant side-lobes. In real antenna side-lobes may be reduced by non-uniform drive at the antenna or by using more than one frequency.

Power reflected from the side-lobes, especially near the peaks of the first, may contribute to the echo received, sometimes suggesting that there is an object in front of the transducer whereas in fact it is well off centre.
When the same antenna is used as a receiver it is again directive, picking up radiation mainly along the line of sight for a circular aperture. We can use the reciprocity theorem from antenna theory [Skolnik 81] to show that the pattern in receiving and transmitting is identical. Hence, if we define \( \Psi_t(\theta) \) as the directivity in transmission and \( \Psi_r(\theta) \) as the same in receiving, then

\[
\Psi_r(\theta) = \Psi_t(\theta)
\]  \hspace{1cm} (3.10)

Again it can be helpful to think of interference as the cause of the receiver directivity. The wavefront is curved but the antenna surface is approximately flat, so different components of the wave excite elements on the antenna with different phase.

Mathematically, therefore, when a single antenna is used in transmission and receiving, the joint transmitter-receiver pattern is simply the product of the far field transmitter and receiver patterns. * Using the exponential approximation:

\[
\Psi_{t\rightarrow r}(\theta) = \Psi_o^2 e^{\frac{4\theta^2}{\theta_o^2}}
\]  \hspace{1cm} (3.11)

3.4 Altering Aperture Shape

Changing the size of the aperture alters the radiation pattern. Another way to alter it is by using differently shaped aperture.

It can be shown that the far field pattern is simply the Fourier transform of the aperture illumination. Therefore, for example, a narrow slit aperture illuminated uniformly (i.e. a rectangular pulse of illumination) has a far

*It is sometimes more intuitive to think of the joint pattern arising in terms of the excitation at the transmitter and receiver. Each element in the receiver responds to each element in the transmitter, with a phase factor depending on the distance travelled. Integrating all these element across the transmitter and receiver results in the convolution of the excitation pattern at the transmitter with the response to the incoming wave across the receiver. Because the far field pattern and the antenna excitation are a Fourier transform pair (as discussed in the next section), this is the same as the product of the radiation patterns in the far fields.
field pattern:

\[ \Psi(\theta) = \Psi_o \left| \frac{\sin(ka \sin \theta)}{ka \sin \theta} \right| \]  

(3.12)

This is similar to the circular pattern (equation 3.6), but replacing the Bessel function with a sinusoid. Rectangular apertures are less common because of the difficulty of providing uniform excitation: effects at corners modify the pattern.

We can use intuition about the Fourier transform to decide how we might shape the aperture for particular applications. Poor localisation at the antenna (i.e. a large antenna) implies good localisation in the far field (i.e. a narrow beam). Figure 3.5 shows the antenna in the KASPA sensor which was developed as a mobility aid for the blind. It is designed to be worn as a pair of spectacles. The central transmitter is narrow in the horizontal direction, broad in the vertical direction. Therefore it produces a horizontal fan of power, to provide broad coverage of the area ahead. Reflections are detected in three receivers. The central one (just above the transmitter) provides good localisation in the horizontal plane and therefore information on reflectors immediately ahead. The two side receivers determine the angular offset of objects outside this central region by a stereo effect. The signals from the side receivers, mixed with the central receiver, are fed into each ear to provide the direction of nearby objects. Range information is provided at the same time as an audible frequency through using CTFM modulation (we will see how this works in the next chapter).

Fig. 3.5 The KASPA sensor. The lower central aperture is the transmitter, the other three receive. The two side apertures are small so they receive power over a significant angle; the upper one receives power from the central region
The position of the receivers affect the amount of power they receive. Even given the relatively wide radiation pattern of sonar, a lot of power is lost if the transmitter and receiver are not coincident (as they are in the Polaroid sensor). The broad radiation pattern in the vertical plane for the upper central receiver is needed to make up for the vertical offset. The side sensors are angled towards the centre to compensate for their horizontal offset.

Achieving efficient power collection is a problem in stereo ultrasound. For sensitivity to direction, a large displacement between receivers is required. However too large a displacement causes problems in correspondence as the receivers fail to see the same object.

Joint radiation patterns for the KASPA sensor, from transmitter to receiver, are shown in figures 3.6 and 3.7. Note the offset in the direction of peak power which results from the offset of the side receivers and how the extent of the central lobe depends on the size of the receivers.

![KASPA: transmitter to central receiver](image)

Fig. 3.6 Directivity pattern from the KASPA sensor, from transmitter to the main receiver. The simulated pattern shows the results from a model which approximated the central transmitter and receiver by rectangular shapes. Note that because the KASPA sensor operates using a variable frequency the side lobes are reduced significantly. The real apertures are rounded with the result that the side-lobes disappear entirely in the measured data. The discontinuities in the modelled data are an artifact of the simulation.
3.4.1 Antenna Arrays

Another way to alter the beam shape is to use an array of antenna. The advantage is a narrower central beam, but the downside is that the side lobes become more significant. Arrays are not used often in robot guidance but are common in radar and in underwater sonar. A successful implementation in air was reported in [Nagi et al 92]. Patch antenna arrays can be produced particularly cheaply. Beam steering is normally achieved in the receiver by software.

The stereo processing in the KASPA sensor is an implementation of differential sonar, which uses just two transducers, placed a few centimetres apart, to determine the orientation of a reflector. A large spacing allows the difference in range to be used to determine angle, but because of problems with correspondence narrow spacing is often preferred. In this case it is usually better to use amplitude difference to find angle, although problems with saturation limit the use of this technique. In radar, the equivalent technique (known as monopulse radar since direction can be established from a single pulse transmitted) is common for tracking rapidly moving objects such as aircraft.

In chapter 6 the advantage of using small arrays to recognise geometrical shape is discussed.
3.4.2 *Focused Transducers*

For optical sensors, lenses are used in transmission and receiving. Even for a semiconductor laser source, focus may be improved in transmission by including a lens. However, a lens only focuses power at a particular distance; beyond this the beam diverges again. For light emission, lenses in the transmission path are most useful when the beam needs to be shaped; for example, a cylindrical lens produces a line of light from a point source. At the receiver, a lens is needed to gather power over a wide area and to focus it onto the detector.

Millimetre waves can be focused in a similar way with an external lens. Lenses are formed of a material of a different impedance (acoustic or electrical) than the surrounding air. For electromagnetic waves lenses are made of dielectric materials. For sonar, they are made of materials of different density or elasticity, such as perspex. Sometimes a lens is useful in transmission, especially when power needs to be concentrated at a fixed distance, but focusing is normally confined to near field operation in applications such as medical imaging where depth is well defined. In robotics the variation in depth means that focused ultrasound is rare. In the receiver the far field radiation pattern is such that the receiver acquires power over a wide angle without a lens.

3.5 *Target Properties*

So far we have seen how the sensor itself affects the way in which the environment is perceived. We also need to understand how features in the environment reflect the millimetre wave power.

To interpret optical images we use two models of reflection. The first describes reflection at a smooth surface: light is reflected so that the angles of incidence and reflection are equal. The second describes reflection at a diffuse surface: light is reflected isotropically, with a power density depending on the cosine of the angle of incidence [Klein and Furtak 86]. Because we see the environment from reflected ambient light the geometric models on which we base our own navigation and perception match these.

However surfaces behave differently at millimetre wavelengths. The reflection from a target depends not only on the material it is made of, but more critically on its shape and the way it is presented to the sonar. Small changes in the orientation of the target to the incident power may have a
very large effect on the power collected by the receiver. Sharp edges and protruding objects reflect much more power than smooth surfaces.

3.5.1 Smooth Surfaces: The Specular Model

Surfaces exhibiting smooth reflection, such as mirrors, are rare for light. We see most objects, even at oblique angles of incidence, because they act as diffuse reflectors. This is because smooth reflections occur only when the variation in surface properties is comparable to the wavelength. For millimetre waves the long wavelength means that many surfaces appear completely smooth. These include most internal walls, some brick walls and some types of flooring.

Reflections from many surfaces therefore occur only from power incident at a single point on the surface, from which the angles to the transmitter and receiver are the same (angles of incidence and of reflection are equal). We can usually assume the transmitter and receiver to be coincident on the scale of the environment, so a reflection from a wall occurs only at the point normal to the transmitter/receiver pair. As the sensor is rotated in front of a wall the range therefore remains constant. This is illustrated in figure 3.8. A smooth plane surface therefore appears curved. Other smooth shapes, such as cylinders and edges produce similar regions. A cylinder for example always returns the power directed towards the centre.

Durrant Whyte and Leonard [Leonard and Durrant Whyte 92] coined the term “region of constant depth” or RCD for such regions. By analysing the RCD resulting from a surface or an edge, the orientation and position of the reflector can be determined. A model of the movement of the RCD with sensor movement was developed to distinguish primitive targets, in particular edges and planes.

Amplitude is very sensitive to orientation. Because the only power reflected is that from a particular point, the amplitude received varies according to the directivity of the sensor in that direction, $\Psi_{t-r}$ (equation 3.11). The variation of amplitude with angle can be used to determine orientation, since the maximum in amplitude occurs when the sensor points normally to the plane. However saturation is often a problem when the sensor is pointing straight at a surface.
3.5.2 Rough Surfaces

For light, the alternative to the smooth model is the diffuse model. The wave-front is re-radiated isotropically from each point on the target. However a fully rough surface does not occur commonly at millimetre wavelengths. Various physical models have been developed, especially in undersea sonar, to determine how the power received depends on factors related to surface variation [Voronovich 99]. In the robotics literature a simple statistical model, which assumes that the surface height follows a Gaussian distribution, has been adopted [Bozma and Kuc 94; Politis and Probert Smith 01]. This model, combined with the radiation pattern, can be used to determine the sound intensity back-scattered at any angle in terms of the standard deviation in surface height and the correlation distance in the surface. However most rough surfaces also show reflections from individual patches, which result in spikes in amplitude.

Many surfaces are neither completely smooth or completely rough. We can often identify two components of reflected power. Figure 3.9 shows reflections measured from two flooring surfaces, one smooth and the other textured. The closer reflection is at the distance along the normal from the sensor to the surface, and corresponds to the smooth reflection. It is well localised, and may arise from power in a side-lobe. The further reflection, which corresponds to the rough component and is seen only in the figure on
the right, is centred about the line of sight of the sensor, where the transmitted power is maximum. The magnitude to the power backscattered from a particular point is less, but reflections arise over an angle corresponding to the central lobe. Therefore it extends over a significant range.

Note that the overall shape depends on the directivity of the antenna as well as the surface properties.

![Graph](image)

**Fig. 3.9** On the left, a reflection from a smooth plane, viewed obliquely, showing only a specular component in the direction of normal incidence. On the right, one from a rough surface, showing components centred both in the direction of normal incidence and along the line of sight. The smooth lines show the fit to a model of the reflected wave [Politis and Probert Smith 01]

### 3.5.3 Scattering Cross Section

So far we have dealt mainly with planar surfaces. For targets which are small compared with wavelength planar analogies cannot be used and we need to talk about scattering rather than reflection. For very small targets (comparable with the wavelength or less), scattering results both in backscattered power and in power propagated past the target, with a shadow region in front. Raindrops are examples of such targets, and contribute significantly to scattering in radar.

In general we describe targets using the scattering cross section, $\sigma$ (see equation 3.1). Because it depends so much on shape, the target cross section does not necessarily bear a simple relationship to the physical area of the target. Specular reflection may mean that there is no radiation back-scattered towards the receiver. In more complex targets, most radiation is
returned from edges and discontinuities.

The target cross section also varies with frequency. The measured cross-section of a man at some conventional radar frequencies has been reported as follows [Skolnik 81]:

<table>
<thead>
<tr>
<th>Frequency GHz</th>
<th>Wavelength mm</th>
<th>Radar Cross Section m²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.120</td>
<td>268</td>
<td>0.098-0.997</td>
</tr>
<tr>
<td>2.890</td>
<td>103</td>
<td>0.140-1.05</td>
</tr>
<tr>
<td>4.800</td>
<td>625</td>
<td>0.368-1.88</td>
</tr>
<tr>
<td>9.375</td>
<td>32</td>
<td>0.495-1.22</td>
</tr>
</tbody>
</table>

These wavelengths are all much longer than those common in robotic sensors, but the figures illustrate how the cross section depends on factors other than geometric properties. It is also clear the cross-section for this particular target (which is diffuse) becomes closer to the physical area as wavelength decreases.

It is interesting too to compare some radar cross sections of different objects (taken at around 100mm):

<table>
<thead>
<tr>
<th>Object</th>
<th>Radar Cross Section m²</th>
</tr>
</thead>
<tbody>
<tr>
<td>single engine aircraft</td>
<td>1</td>
</tr>
<tr>
<td>jumbo jet</td>
<td>100</td>
</tr>
<tr>
<td>small open boat</td>
<td>0.02</td>
</tr>
<tr>
<td>cabin cruiser</td>
<td>10</td>
</tr>
<tr>
<td>car</td>
<td>100</td>
</tr>
<tr>
<td>bicycle</td>
<td>2</td>
</tr>
<tr>
<td>man</td>
<td>1</td>
</tr>
<tr>
<td>bird</td>
<td>0.01</td>
</tr>
</tbody>
</table>

The cross section is related more to shape and material than to size. Back-scattered reflections from the aircraft arise mainly from the engines and other protrusions; hence the single engine aircraft and the man backscatter similar amounts of power in spite of the differences in area.

### 3.6 Attenuation in the Transmission Medium

Returning to equations 3.1 and 3.3 we can redefine \( \sigma \) and \( \Gamma \) in terms of a quantity related to surface type and one related to shape.
Referring particularly to equation 3.3 we define:

\[ \Gamma = \rho \mathcal{G} \]  

(3.13)

where \( \rho \) is the scattering coefficient related to the surface type and \( \mathcal{G} \) relates to shape. Because both may affect the type of scattering and hence the form of beam spreading, they may be functions not only of angle but of distance as well. \( \mathcal{G} \) is defined so that it is unity for a planar reflector.

To examine the effects of the transmission medium more carefully we take the losses in two parts, first that due to spreading and then that due to loss. To remove the dependence on target from the spreading term, we define the spreading term for a planar reflector.

### 3.6.1 Beam Spreading

The type of spreading depends on whether the target is diffuse or specular. Different assumptions are normally used for sonar and radar.

- For a diffuse reflector the power is scattered isotropically. The amplitude of the vibration decays as \( 1/r \) on both forward and reverse journeys. Hence amplitude decays as \( 1/r^2 \), when both journeys are included; power as \( 1/r^4 \). Spreading loss of this type is typically assumed in conventional radar.

- For a specular reflector the echo formation can be modelled using a ray model (chapter 6). The decay in amplitude depends on both distance and how the reflector scatters the signal, as described in section 6.2.2. This model is normally assumed for sonar (equation 3.3). Measurements then suggest that for a planar surface we can define the amplitude loss as \( 1/2r \) for a total path length \( 2r \).

### 3.6.2 Losses

Beam spreading is one cause of attenuation in the transmission medium. However this is not a lossy mechanism; it is caused by scattering. Loss occurs too. The power suffers exponential attenuation, caused largely by molecular absorption.

We can describe an attenuation constant \( \alpha \) such that the amplitude attenuation is \( e^{-2\alpha r} \) (power attenuation \( e^{-4\alpha r} \)) where \( 2r \) is the total path length. The attenuation constant \( \alpha \) increases with decreasing wavelength.
It can be represented approximately over the range of frequencies of interest by an inverse relationship: \( \alpha \approx 1/\lambda \).

In air borne sonar, at a wavelength of around 45kHz, \( \alpha \) is about 0.035Np/m. \(^\dagger\) It varies significantly with temperature and humidity.

In water borne sonar the attenuation is low. In air though the higher attenuation is the main factor limiting the signal to noise ratio. A long wavelength is needed if the wave has to propagate any distance. Bats for example use wavelength in the range of 3-6mm to detect prey a metre or so away; longer wavelength would compromise resolution. Whales which may communicate over many miles use low frequency sound with wavelength several orders of magnitude lower.

For radar, distinct peaks and troughs result in the attenuation-frequency curve depending on molecular resonances (chapter 8). Water vapour is the dominant factor affecting attenuation. Frequency bands are selected which lie within the regions of low attenuation when long distances need to be covered. Frequency use is controlled closely by the telecommunications regulating bodies. For robotics, for which distances are normally small, other frequencies may be feasible technically, but there are few products available outside the commercial frequencies. The resolution is inadequate for feature localisation at longer wavelengths.

3.7 Summary

This chapter has provided an introduction to the way in which the world is perceived using millimetre wave sensors. The discussion is applicable to both sonar and millimetre wave radar. We have introduced the concept of beam directivity and shown how it depends on the size and shape of the antenna. We have discussed target properties and shown how reflections from simple shapes relate to position and orientation.

\(^\dagger\)\( \alpha \) is often expressed as dB: \( \alpha_{dB} = 20\log_{10}\alpha \).
This page is intentionally left blank
Chapter 4

Advanced Sonar: Principles of Operation and Interpretation

Penelope Probert Smith

The last chapter dwelt on spatial effects arising from wavelength. This chapter examines recent developments in sonar, especially those based on the sonar signature. The sonar signature provides information on range and amplitude from all echoes along the line of sight. We show in this chapter how reliability and efficiency can be improved through using transmitter modulation similar to that observed in nature. Radar has used similar techniques for many years.

We start by discussing the traditional pulsed sonar, which returns only the distance to the nearest object. We call this “single return sonar”. We then extend the discussion to more advanced sonar. We describe three ways to measure the sonar signature.

An important area of advanced sonar which we omit in the chapter is the development of small sonar arrays. These are covered in chapter 6.

4.1 Single Return Sonar

The majority of sonar systems used in robotics do not return the signature but simply the distance to the nearest object. The transmitter is pulsed and emits a short burst of sound at a single frequency. The distance to the nearest reflector is determined by measuring the time between the point at which the pulse is transmitted and that at which the signal received is above a threshold. A schematic diagram for such a system, as implemented in the Polaroid sensor, is shown in figure 4.1.

This type of system throws away a lot of information, since it ignores all reflections beyond the first. It also ignores the information included in the amplitude of the echo, which contains information on the target and its orientation. It is noisy, especially sensitive to noise spikes. In addition it can only be calibrated for reflectors of a particular strength, as we see in
Fig. 4.1  The principles of first return sonar using a threshold detector, as used in the Polaroid and similar sensors. On a “start” signal a pulse consisting of a number of cycles of a single frequency is transmitted, and a counter is started. The counter stops when the signal received exceeds a preset threshold. The time of flight can be read directly from the counter and converted to distance. The system could be modified to use a threshold which decreases with time over the receiving period to account for the attenuation of the sonar signal with range

figure 4.2.

A method to remove the bias problem was described by Barshan and Kuc[Barshan and Kuc 90]. The leading edge of the echo is fitted to a quadratic curve. The parameters of this curve are used to estimate the time delay of the pulse.

In spite of the problems, this type of transducer has widespread application. It is particularly suitable for obstacle avoidance, which requires only approximate spatial resolution. Good results are given for example when the main obstacle is people, which act as good reflectors from all directions. However problems arise from specularity for many objects. More reliable results can be achieved through using a large number of measurements in time or space so invalid data can be filtered out or ignored. Statistical methods of integration also improve performance, for example using the Bayes update rule which can incorporate a statistical model of sensor failure as well as contextual knowledge [Hu et al 95].

A number of mobile robots sold commercially for research include a ring of such sensors. A typical configuration is shown in figure 4.3. To
Fig. 4.2 The envelope of a perfect rectangular pulse (solid line) is modified by the limited bandwidth of a real transducer, which here acts as a band pass filter (upper dotted line). The instant at which a threshold voltage of 0.25V is reached depends on the transducer characteristic. (In fact the effect is exaggerated in this diagram for clarity). Also shown is the pulse from a poorer reflector, arriving at the same time. The times at which the two real signals cross the threshold are 935 and 1023μs respectively, illustrating the range biasing caused by variable amplitude of real signals.

Fig. 4.3 Examples of a sonar ring on a Real World Interface platform. You can see the circular transducers just below the top of the main platform.

provide good coverage, the transducers may be placed at intervals as small as 15° (the optimum interval is governed by the beam-width).

Cross talk between sensors is a problem in using a close spaced array efficiently. Improvements can be made by sequencing the firing times carefully.
4.1.1 Mapping and Navigation using Single Return Sonar

Sonar has natural advantages for mapping and navigation - it is robust to different light levels, almost immune to background noise, (given appropriate drive signals and filtering) and sensors are cheap. Sonar sensors are often included on platforms to provide a bumper for obstacle avoidance.

There are two main difficulties: the wide beam makes angular resolution of a single reading poor and target properties, especially specularity, cause erroneous readings. The wide beam means that even tasks such as finding an open door are a problem since reflections come back from the surround. Specularity (section 3.5.1) may mean that a target is not seen or it may cause occasional readings which are grossly incorrect, resulting from several echoes around the environment (like going into a hall of mirrors). These problems mean that it is almost impossible to extract line segments directly [Crowley 85; Crowley 89] to match external plans or maps. Other methods must be used.

4.1.1.1 Occupancy Grid Representation

The brittleness of line segment-based feature extraction was a motivating force behind the occupancy grid representation for sonar maps [Beckerman and Oblow 90; Elfes 87; Gilbreath and Everett 88; Noborio et al 90; Zelinsky 88].

Elfes [Elfes 87] was one of the first to establish the occupancy grid representation for sonar, using first return sonar. The occupancy grid is a simple way of representing the world by dividing the environment into a finely spaced grid of squares in a plan representation. Each square is marked as empty or occupied, depending on results from the sensor. An example of a grid is shown in figure 4.4. This shows a coarse grid of a rectangular environment with three obstacles in it. (In practice the grid would be much finer.) Following the first scan of a sonar sensor some squares are marked as occupied. Note (i) that only part of the smooth boundaries are seen as the reflections outside this range are specular and no radiation is returned to the sonar and (ii) that obstacles are extended beyond their physical boundaries owing to the width of the beam. As the robot moves around the sensor sees the world from different viewpoints and gradually a reliable...
Fig. 4.4 The first stage of building up an occupancy grid. A rectangular environment is bounded by smooth walls and contains of two smooth objects (lighter shading) and a rough one (dark shading). A sonar device rotates on the robot, marked with a blob. As the sensor rotates, it assumes that any echo returned is along its line of sight, and if the range of the echo is within certain bounds it marks the equivalent square with a cross (echoes may result too from multiple reflections but most will be at illegal ranges). The figure shows the result after a single complete (360°) scan.

representation is built up. The representation can be improved by marking each square with the probability of occupancy rather than just a binary value to allow for uncertainty. The map can be used as input to planning algorithms.

Dense coverage of the space is required for a reliable map. One difficulty in acquiring enough readings is the time it takes because of the relatively slow speed of sound. Time may be reduced using an array and acquiring data in parallel, but the wide beam may lead to interference between adjacent sensors. A more fundamental problem is uncertainty: the uncertainty in the earlier parts of the map increases as the robot moves owing to increasing uncertainty in relative position. An algorithm in which, over time, the probability of occupancy reverts to a don't know state is sometimes used to handle this.

Using an environment made up of cylindrical objects, Borenstein and Koren have demonstrated reliable mapping and navigation based on the
occupancy grid and a potential field algorithm [Borenstein and Kuc 91]. * Dense, fast measurements are taken using an array of 24 ultrasonic sensors. Because of the density of measurement, and the change in orientation to reflectors as the robot moves, specularity is not a problem as at least some readings will be taken in directions where there is some back-scattered power.

4.1.2 Landmark Based Mapping

The occupancy grid is tolerant both to specularity and to the poor angular resolution of sonar because of the density of the readings taken. However the resulting map cannot be correlated easily with most external representations, which are based on geometry or landmarks. In addition, the map is prone to long term drift since the form of data representation means that it is difficult to correlate new readings with those already processed. Wijk and Christiansen [Wijk and Christensen 00] examine a landmark based method. A landmark is defined as a significant reflector in the environment, and can normally be correlated with particular physical features. Their method is based on the fact that adjacent sensors in a sonar array normally have enough overlap to see the same object. Triangulation, based on the measurements from pairs of adjacent sensors, is used to find the strong reflectors which are detected by both sensors in the pair. A voting scheme determines the validity of each feature as a landmark. Navigation of a robot in a large indoor area is demonstrated. Again dense measurements are used to overcome some of the basic difficulties in sonar. The system is demonstrated on the Nomad platform with a polar array of 16 sensors each with 25° beam-width. Henderson et al [Henderson et al 98] also use pairs of transducers to extract information on walls, but through combining their individual estimates of range to determine orientation.

*The potential field algorithm borrows ideas from electrostatics to find a path as follows. The robot is assigned a positive charge. Each obstacle observed is also assigned a positive charge but the target location is assigned a negative charge. The total electrostatic force on the robot is then determined in both magnitude and direction, and the robot moved according to this force until it reaches the target. Problems may arise owing to false maxima and a random element is often introduced into the robot's movement to overcome this.
4.1.3 The Geometric Target Primitives

A more efficient representation is based on metric information, in which a geometric description is built from line and point primitives. Standard feature primitives are built up from these: planes (walls), corners and edges.

Within a room, for example, a line results from a planar segment such as a wall, walls meet in a corner (viewed from the inside) and an edge might occur from the outside corner of a cabinet. The primitives are useful in themselves as landmarks but can be used together to create higher order representations (e.g. the presence of two corners with some line segments in between confirms a wall).

Chapter 6 describes how two transducers can be used together to determine primitive type and position [Barshan and Kuc 90; Audenaert et al 93]. Kleeman and Kuc [Kleeman and Kuc 95] showed that two transmitters were required to produce this information from a single measurement, and Hong and Kleeman extended the models into three dimensions [Hong and Kleeman 98]. Leonard and Durrant Whyte [Leonard and Durrant Whyte 91] took an alternative approach, using the region of constant range or depth which occurs as the beam from a single sensor is swept across a single feature. The centre of each region, or RCD, is tracked as the robot moves and the movement used to identify the primitive. The method is shown to be accurate and reliable but required very dense measurements: a scan increment of 1 degree or less with readings taken each centimetre.

More recently further work using image geometry has been presented by Barshan [Baskent and Barshan 99], who has demonstrated a morphological filtering approach, again using the Nomad mobile robot. Morphological filters are popular in image processing for tasks such as edge detection, enhancement and smoothing. They can be used at different scales to ignore certain discontinuities (see chapter 7). Barshan demonstrates their use to remove occasional specular reading in the extraction of walls and corners as the platform moves, and to find the edges of a doorway.

4.2 Advanced Sonar: The Sonar Signature

To use single return sonar successfully dense measurements are needed together with models based on certain assumptions about the environment (normally that they are specular). Even then reliability is poor. The most important recent advances in the use of sonar have been achieved by a
combination of hardware and software development. In particular they are based on algorithms which use the sonar signature.

4.2.1 Range Signature

The range signature is a plot of amplitude against range for all reflectors along the line of sight. Because the beam is wide, occlusion is less significant for sonar than it is for vision, and a number of reflections can normally be identified. Figure 4.5 shows typical range signatures. On the left, we see the signature when the sensor was pointed at a wall behind a wire netting fence. Both wall and fence are clearly visible, the fence at about 1m, the wall at 2m. In spite of the fact that it is a lot closer the signal returned from the fence is the smaller since the netting is a poor reflector. On the right we see the pattern of reflections from a flight of steps, where the reflections from individual steps are prominent. Fourier analysis reveals the periodicity in this pattern.

![Range Signature](image)

Fig. 4.5 On the left: the range signature from a street scene, consisting of a wire netting fence in front of a wall. On the right, the signature from a flight of steps. (Data taken with CTFM sensor, described in section 4.3.3)

Smooth surfaces return just a specular component of reflection, which is represented by a vertical spike in the range signature. In practice some spread occurs because of quantisation effects in acquiring the signal. However it occurs too because surfaces are not perfectly smooth. Because a significant area of the surface is insonified, the sound returned does not arrive at a single instant, since some parts of the sound have travelled further
than others. The projection of the beam cross-section (which is circular) onto a plane illuminates an ellipse. The size of this ellipse depends on its distance and orientation. Figure 4.6 shows experimentally how the resulting spread in time of flight is seen in the range signature.

![Diagram showing FM sonar images of surfaces.](image)

Fig. 4.6  FM sonar images of surfaces. Plot (a) is the image of a smooth surface at zero bearing and plot (b) is the image of a smooth surface obtained at the same distance but a bearing angle of about 28° (the second small peak in plot (b) is due to a further reflector in the line of sight). The basic shape of a smooth surface image does not vary significantly, but there is significant change in the amplitude of the peak with angle of incidence (arising from the angular variation of the radiation pattern). Plot (c) is the image of a rough surface at zero bearing angle and plot (d) is the image of a rough surface obtained at the same position but bearing angle of about 28°. It can be seen that the basic shape for the rough surface varies significantly with bearing angle.

Kuc and Siegel attempted to use this variation in shape to determine the orientation of planar surfaces, defining a set of matched filters (see section 4.3.1.1), each designed to match a plane at a different orientation. The incoming signal was filtered with each in turn, and the filter which resulted in the largest output assumed to be modelling the correct orientation. Kleeman and Yata [Yata et al 98] describe a similar technique to measure orientation using just a single reading. Improvement in range accuracy was shown by Kuc and Kleeman [Kleeman and Kuc 95] through
using the predicted shape of the pulse to improve estimates of arrival time in a correlation receiver (section 4.3.1.1).

A difficulty with the technique is that it is very sensitive to a good model of the surface. Effects from changes in orientation cannot be separated out from the effects from changes in roughness.

A number of workers have used the sonar signature to classify objects more generally, either to determine orientation or to recognise a particular object. Various formats of echo representation (time domain, frequency domain and time-frequency) have been used as inputs to neural networks [Dror et al 95; Harper and McKerrow 95a; Dror et al 96] in applications varying from face to plant recognition. Politis has abstracted features based on statistical properties of the signature to recognise the configuration of office chairs, and also textured pathways [Politis and Probert Smith 01] (see chapter 5). Kuc [Kuc 97a; Kuc 97b] has used correlation techniques to classify objects such as washers and paper clips.

### 4.2.2 Orientation Signature

Another useful way of presenting information is through the orientation signature, which is a plot of the amplitude of the reflection from a single target as the orientation of the sensor to the target varies. It is obtained by extracting the amplitude of the reflection from a single target over a set of range signatures as the sensor is rotated about a point. For reflectors with reasonable separation (more than a few centimetres), it is normally easy to identify the echo resulting from the same reflector between the different orientations.

The main advantage gained from the orientation signature is being able to determine orientation to the target accurately. The maximum amplitude arises at the point at which the target lies along the centre line or \textit{bore-sight} of the transducer beam. The signature can be used too to guide target recognition, particularly between line (i.e. planar) and point (corner, edge) targets. Although in theory the orientation signature is identical for smooth planes and corners, [Barshan and Kuc 90], in practice it does provide some discrimination [Kao and Probert 00] (chapter 5).

Saturation may more generally be a difficulty in exploiting the sonar signature. Sonar is very sensitive to angle and to get enough power to see over 5m, a typical maximum range, the transducer will saturate if pointed at normal incidence to a wall closer than 1-2m. This problem has prevented
amplitude data being used much in the past. It can be overcome in processing through software models [Politis and Probert 98] or, better, prevented by using a drive with variable amplitude at the transmitter.

4.2.3 Rough Surfaces

Rough surfaces are not homogeneous, and closer examination of the sonar signature suggests that the echo is dominated by discrete randomly placed reflectors on the surface. These are probably elements at an angle close to normal of the incident beam. This hypothesis is supported by figure 4.7, which shows how small elements at constant range make up the range signature for a rough surface. It also explains the high density of spikes in figure 4.8, which shows the reflections resulting from a thick carpet, fitted against a theoretical approximation which matches energy [Politis and Probert Smith 01].

Fig. 4.7 Variation of range with angle for a plane target parallel to the transducers at zero bearing. Range is 600mm

4.3 Acquiring the Sonar Signature

Three technologies are commonly used to obtain the sonar signature:

- Pulsed single frequency sonar. No changes are needed in the transmitter. In the receiver the threshold detector is replaced by a data acquisition module which samples the complete echo signal.
- Pulse compression sonar. This works in a similar way but the pulse transmitted is frequency modulated. The technique is common in radar and mimics the call of many species of bat. With a suitable
receiver, pulse compression sonar provides excellent resolution.

- Continuous wave frequency modulated (CWFM) sonar (often called continuous time frequency modulated or CTFM sonar). This type of modulation is observed in nature too. The pulse transmitted is frequency modulated again, but this time extends over most of the repetition period. Range is determined from the frequency difference between the pulses transmitted and received. The technique is the most reliable in unknown environments since the sensors emit more power and an averaging mechanism is built into the processing - but at the expense of more sophisticated hardware design.

The different transmitter signals are summarised in figure 4.9.

4.3.1 **Single Frequency Sonar**

Improvements to pulsed, single frequency, sonar from acquiring the complete echo signal were first demonstrated by Kuc [Kuc and Siegel 87].

4.3.1.1 **Improving Range Accuracy: The Correlation Receiver**

The range signature data can be used to improve range resolution through correlation at the receiver. The correlation receiver needs an internal model of the signal expected. We call this model the reference signal. Commonly
the reference signal is taken to be identical to the signal transmitted.

The receiver picks out the time of arrival of the echo pulse. A schematic diagram is shown in figure 4.10.

The mathematical operation performed by a correlation receiver is as follows:

$$\int_{-\infty}^{\infty} x_\circ(t + \tau)x(t)d\tau$$  \hspace{1cm} (4.1)$$

where \(x(t)\) is a signal detected at time \(t\) and \(x_\circ(t + \tau)\) is the reference signal, delayed by \(t = \tau\). Optimum results are obtained when \(x_\circ\) is identical to \(x\) apart from a time delay. The integral is maximum when the signals \(x_\circ\) and \(x\) coincide completely; i.e. when \(\tau = 0\).

For real signals, correlation is equivalent to the convolution operator, which we meet in Fourier analysis. The convolution of two signals in the time domain is equivalent to their product in the frequency domain. Therefore correlation is essentially a filtering operation. Optimum results are obtained when the signal expected is of exactly the same form as the one
being received as above. It can be shown that in this case correlation is equivalent to filtering a signal with spectral content $X(f)$ by a filter with transfer function $X^*(f)$. This type of filter is called the matched filter and is important in signal processing theory. It maximises signal to noise ratio [Skolnik 81].

A drawback of the correlation receiver is that it is very sensitive to a poor model of the reference signal.

### 4.3.2 Pulse Compression Sonar

Even using a correlation receiver, the sensitivity is not usually enough to distinguish between echoes closer than a few centimetres. This resolution may be required to see detail on targets, but it also arises in general navigation tasks, such as central travel through a doorway. Kuc and Kleeman have reported a sensor which generates very short pulses to distinguish targets less than 1cm apart [Kleeman and Kuc 95]. However in most sensors a deviation of several centimetres in the path is needed before the reflections from the two door sides can be distinguished reliably.

Problems arise because the pulses from two echoes overlap. To avoid this the reference pulse must be as narrow as possible. Through Fourier theory we know that a narrow pulse in time corresponds to a wide frequency
range, or bandwidth. Hence we need a wide bandwidth for good resolution.

The easiest way to improve resolution therefore is to increase bandwidth. One way of doing this is to use very short, sharp pulses (the approach taken by Kuc and Kleeman and in medical and non-destructive testing ultrasound). For sonar in air, this is often not a good option as the relatively high attenuation means that the signal to noise ratio is too low. The alternative way to increase bandwidth is by frequency modulating the signal. This technique is observed in the natural world (most bats emit chirps, rather than single frequency calls, and has been developed as standard technology in radar) [Simmons et al 95; Peremans and Hallam 98; Altes 95].

Pulse compression sonar is similar to the standard single frequency sonar, in that time of flight is measured directly. However rather than transmitting a single frequency pulse, the pulse is frequency modulated, for example as a frequency which changes linearly with time, a chirp. The echo is located using a correlation receiver. Peremans [Peremans and Hallam 98] discusses in detail the advantages which may be gained.

The bandwidth of the transducers limits the signal which can be sent, usually to about one octave (a doubling in frequency). For most electrostatic transducers, which have characteristics similar to the Polaroid one, the overall transfer function of the transducers and the path is such that the resulting envelope can be modelled as a Gaussian pulse.

Mathematically a typical pulse transmitted can be described as follows:

$$x_{chirp}(t) = X(t) \cos 2\pi (f_0 t + \frac{\beta t^2}{2})$$ (4.2)

This describes a frequency variation which is linear with time with $df/ dt = \beta$. X(t) is an amplitude modulation to model the finite bandwidth of the transducer.

Gilkerson [Gilkerson 00] describes such a system based on a Polaroid sensor. The signal transmitted is such that $X(t)$ is approximately Gaussian and swept over about 20kHz. A notch filter is included at the resonance frequency to avoid saturation here. The pulses emitted need to be 1-2msec in length; if they are too short there is so little energy that only very short distances can be measured. Figure 4.11 traces the signal reflected from a planar surface as it goes through various stages in the receiver [Gilkerson 00]. On the left the raw echo is shown. In the middle it is shown following
correlation with a pre-stored version. Note the improvement in time localisation. On the right the signal is shown following envelope detection. In these experiments, the reference signal was created through measurement, averaging the echoes received over a number of typical readings.

The improvement in resolution over threshold detection can be seen in figure 4.12 which shows the reflection resulting from two targets. At the top we see echoes from two well separated targets, before and following detection. The centre figure shows a case in which the envelopes overlap and a threshold method would detect only one target, but the targets are well separated by the correlation receiver. The lower figures show how even the correlation receiver fails as the spacing decreases.

4.3.3 Continuous Wave Frequency Modulated Sonar

An alternative type of sonar operates in the frequency domain. Continuous wave frequency modulated sonar (often called continuous time frequency modulated, or CTFM, sonar) was introduced as a navigation aid for the blind by Kay [Kay and Kay 83] in New Zealand in the 1960s. Successive versions of this have been marketed, the most recent being the KASPA sensor described in the last chapter. Similar systems have been used in radar for many years.

CTFM uses a linear pulse chirp which has the same form as the one we have just met:

$$x_{chirp}(t) = X(t) \cos 2\pi \left(f_0 t + \frac{\beta t^2}{2}\right)$$  \hspace{1cm} (4.3)

where $f_0$ is the lowest frequency transmitted and $\beta$ is the constant ramp-
Fig. 4.12. The top figures show the raw signal and the output from the matched filter for two targets, spaced 51 cm apart. The middle figure shows the same for targets spaced 17 cm apart and the lower figure for targets only 7 mm apart.

The instantaneous transmitted frequency is therefore given by

\[ f_T(t) = \frac{d}{dt} \left( f_0 t + \frac{\beta t^2}{2} \right) = f_0 + \beta t \]  

(4.4)
Fig. 4.13  A cycle of the pattern transmitted by the FM sonar, showing the frequency variation with time. The top trace shows the signal transmitted, the lower one the signal received.

The signal received is at a different frequency than the one currently being transmitted (at a lower frequency when the frequency increases over the transmission time). The frequency difference shows how long ago it was transmitted. A typical receiver is shown in figure 4.14.

The mathematical form of the signal transmitted is identical to that from the pulse compression sonar. The difference is that the pulse compression sonar requires a signal to return whilst it is quiet. Hence the time for which it transmits is quite small compared with the time between
adjacent pulses. For the CTFM sonar the signal must return while the
transmitter is active. Therefore the signal takes up most of the repetition
period. The additional power which results leads to improved resolution.
However a dead time is needed to avoid ambiguity between adjacent fre-
cquency sweeps. Typical parameters for such a sensor are shown in table
4.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest transmitted frequency</td>
<td>45kHz</td>
</tr>
<tr>
<td>Highest transmitted frequency</td>
<td>90kHz</td>
</tr>
<tr>
<td>Dead time (for which signal stays at ( f_i ))</td>
<td>24msecs</td>
</tr>
<tr>
<td>Time for the signal to ramp from ( f_i ) to ( f_h )</td>
<td>136msecs</td>
</tr>
<tr>
<td>Sampling interval</td>
<td>40( \mu )sec</td>
</tr>
<tr>
<td>No of points sampled per measurement</td>
<td>2048</td>
</tr>
</tbody>
</table>

Table 4.1 Parameters of the CTFM sonar sensor, developed in the 1970s by David Witt
in Oxford, originally for navigation for the blind. It was used to produce all the sonar
signatures shown in this chapter and the next.

The dead time determines the maximum range from which signals can
return and should be adjusted to match the acoustic power output of the
sensor.

At the receiver, the signal is mixed with the transmitter signal, produc-
ing sum and difference frequencies. The difference frequency is typically in
the audible range. The signal at the sum frequency is removed by filtering.
The frequency content of the signal remaining (at the difference frequency)
contains components with frequencies proportional to the ranges of all re-
reflecting objects.

The difference signal is sampled and a fast Fourier transform (FFT)
performed to extract the range information. For a given sampling frequency
the theoretical range resolution, \( \Delta r \) following the FFT is:

\[
\Delta r = \frac{c}{2\beta NT_s}
\]

(4.5)

where \( c \) is the velocity of sound. Hence from table 4.1 the range resolution
for this sensor is 6mm.

The advantage of the CTFM device is that it does not need a model of
the reflector to achieve this accuracy. Its sensitivity is such that, pointed
at a door, it can distinguish the handle from the surface behind.
4.3.4 Doppler Effects

One area where there has been little work in mobile robotics is Doppler sonar. For radiation moving at velocity $c$ and with frequency $f_0$, the Doppler shift resulting from a target moving with relative velocity $v$ is $(2v/c)f_0$. If we take $f_0$ as 50kHz, $c$ as 340m/s, then we have a frequency shift of about 0.6% or 300Hz for a target moving at 1m/s in the direction of the sensor (a walking speed).

Given that the velocity may not be along the line of sight of the sensor, smaller differences in frequency may have to be detected. Determining, say, a 0.1-1% change for a pulsed sensor, either single frequency or chirped, is impractical in software. A hardware heterodyne stage to find frequency difference is required.

For good frequency resolution, long pulses are needed (the uncertainty principle can be used to show that it is impossible to locate a signal accurately in time and frequency simultaneously). A single frequency continuous wave sensor provides this and is the best way to detect frequency change (as used in radar speed detectors!). However it cannot resolve range at all. A compromise may be found using CTFM sonar. Since this includes a hardware frequency difference stage it can handle the problem of frequency resolution. However because the frequency is changing all the time, the Doppler shift is blurred over a range of frequencies. In addition a double sweep, up and down, is required to distinguish the frequency differences due to range and velocity, and even then there are regions of uncertainty. Pulse chirped sonar can also be used. The shape of the frequency modulation determines the balance between resolution in time and frequency. Skolnik includes a section on using these modulation schemes in the context of radar [Skolnik 81].

4.4 Summary

This chapter has introduced the concept of the sonar signature and shown how it can be used to improve the reliability and efficiency of measurements. A number of technologies to acquire the signature have been described.
Chapter 5

Smooth and Rough Target Modelling:
Examples in Mapping and Texture Classification

Penelope Probert Smith

In this chapter we develop models of the range and orientation signatures for different types of surface. We show how these can be used to locate and discriminate between the target primitives described in section 4.1.3 of chapter 4, given a single rotating sensor. The technique is illustrated in a simple map building exercise in an indoor room. In addition we show empirically how the energy distribution in the scattered energy from a textured surface can be used to classify different surface types. Because of its reliability (which results mainly from the improved signal to noise ratio as a result of the continuous - or nearly continuous - output), results are shown for CTFM sonar. Table 4.1 in the last chapter gives the parameters for the device, which had a beam pattern similar to the Polaroid sensor.

The method relies on models of targets, both rough and smooth, based on their reflective properties. It assumes that we have a single rotating sensor. This is in contrast to the method described in the next chapter in which a small array of sonar transducers is used to determine the feature (a more robust method for smooth targets). Discrimination between the geometric primitives uses amplitude to determine the feature type.

Previous work in robotics has assumed that targets are smooth. Using the rough model provides a more robust method of target recognition as well as offering information on texture.

5.1 Power Received by the Transducer

As discussed in chapter 3 we can write the amplitude of the sonar wave received following reflection from a target at an angle \( \theta \), \( P_\theta \), as follows (equation 3.3):
\[ P_\theta = K \rho G L_{at}^{pressure} \Psi_{t-r}(\theta) \]  

\[ \Psi_{t-r} \] is the joint transmitter-receiver directivity pattern (equation 3.11 of chapter 3).

- \( K \) is a constant over range and orientation for a particular sensor.

- Parameters \( \rho \) and \( G \) are related to the target as described in equation 3.13. \( \rho \) is the backscattering coefficient and related to surface orientation and material. \( G \) is the target geometry coefficient which depends on target shape. It is defined so that it is unity for a planar surface.

- \( L_{at}^{pressure} \) represents the atmospheric and spreading losses for a planar reflector at distance \( r \) (see section 3.6 in chapter 3):

\[ L_{at}^{pressure} = \frac{e^{-2\sigma r}}{2r} \]  

\( \rho \) and \( G \) together are the basis of the target models developed in this chapter.

### 5.2 Smooth Surface Model

#### 5.2.1 Backscattering Coefficient

The backscattering coefficient of a smooth surface can be derived by considering the physics of propagation of a sonar wavefront reflected from a surface [Bozma and Kuc 91]. At an angle \( \theta \) to the sensor, it can be shown that the backscattering coefficient of a smooth surface of reflectivity \( R \) is [Kao 99]:

\[ \rho \approx \begin{cases} R & \theta = 0 \\ 0 & \text{elsewhere} \end{cases} \]  

This corresponds to a pure specular reflection, which is zero except at normal incidence.
5.2.2 The Target Geometry Coefficient

For smooth solids the reflectivity $R$ is normally close to 1. The main factor affecting the amount of sound received (assuming the orientation is such that backscattering occurs) is the shape of the target. This effect is described by the target geometry coefficient, $\mathcal{G}$.

We characterise targets in terms of the standard geometric primitives: line, corner, edge. We assume that the corner includes an angle of 90° and the edge of 270°, both to within about 5°.

For a smooth surface, we can model the sound reflection from a surface using a ray theory approximation. In chapter 6, Peremans defines an effective range $r_{eq}$. Given the true range, $r$, we see from section 6.2.2 that:

- For a plane and a corner:
  \[ \mathcal{G} = 1 \quad (5.4) \]

- For a 90° edge $\mathcal{G}$ is distance dependent. Peremans suggests that $\mathcal{G}$ depends on $1/r$. Other models have suggested a relationship of $\sqrt[\lambda/r}$ [Politis 00] where $\lambda$ is the wavelength of the sound. Both cases predict that $\mathcal{G}$ is significantly less than unity.

In practice we do indeed see a significant difference between the values of $\mathcal{G}$ for these two situations. However in addition, in spite of the theoretical values, there is a difference between the values of $\mathcal{G}$ for a corner and a plane, probably owing to imperfections at the corner. This is confirmed experimentally in table 5.1.

5.2.3 Mapping Experiments

The target geometry coefficient together with the positional data from the range and orientation signatures were used to determine features in an experiment to map an indoor room. The test room is shown in Figure 5.1. The size of the room is approximately $2.5 \times 3m^2$. There is a wardrobe in the south of the room, whose door consists of two sliding pieces made of wood. The exit door at the south-west corner is also made of wood. In the west of the room, there is an electric heater. The metallic surface of the heater is flat and smooth. North of the heater, there is a surface made from a cupboard which actually covers the computer equipment located in the north-west corner.
Table 5.1  Results from estimating a factor proportional to $G$ using equation 5.1. Results are shown for a plane, corner and edge, each made up in the laboratory from a sheet of metal, and so assumed to have the same $\rho$. $G$ is underestimated for the surface at 0.4m due to sonar signal saturation.

<table>
<thead>
<tr>
<th>Range (r)</th>
<th>Plane</th>
<th>Corner</th>
<th>Edge</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>460</td>
<td>111</td>
<td>24</td>
</tr>
<tr>
<td>0.6</td>
<td>580</td>
<td>108</td>
<td>29</td>
</tr>
<tr>
<td>0.8</td>
<td>588</td>
<td>110</td>
<td>20</td>
</tr>
<tr>
<td>1.0</td>
<td>686</td>
<td>105</td>
<td>-</td>
</tr>
<tr>
<td>1.2</td>
<td>684</td>
<td>115</td>
<td>-</td>
</tr>
<tr>
<td>1.4</td>
<td>507</td>
<td>140</td>
<td>-</td>
</tr>
<tr>
<td>1.6</td>
<td>565</td>
<td>118</td>
<td>-</td>
</tr>
<tr>
<td>1.8</td>
<td>588</td>
<td>143</td>
<td>-</td>
</tr>
<tr>
<td>2.0</td>
<td>587</td>
<td>142</td>
<td>-</td>
</tr>
</tbody>
</table>

Fig. 5.1  The test room

A sensor was placed at each position on a 4x4 grid. *

5.2.3.1 Finding the Position of Each Feature

Most applications in planning require the range and orientation of targets to be measured. Smooth targets generally have a well defined range signature with a clear peak at the range of normal incidence, and this can be extracted using peak detection techniques. Frequently several reflections

*The grid positions were chosen heuristically to give reasonably even coverage over the free area of the room; they were not chosen to optimise the map.
may be present in the range signatures and a decision has to be made on which are valid. Kao uses an adaptive threshold based on the signal to noise ratio [Kao and Probert 00] for this.

Orientation is extracted using the orientation signature, which is created for each reflector as the sensor rotates (see chapter 4, section 4.2.2). The difficulty in creating the orientation signature if the environment is at all complex is data association - which peak in one range signature corresponds to a peak in the next. Data association can be a big problem in vision, in which there may be sudden changes in intensity between successive images (for example owing to changes in ambient light), since only an intensity image is available. However having range data available makes association straightforward in most cases.

The association algorithm assumes that the range of each reflector (taking into account any sensor movement) remains approximately constant. It fails if a reflector is too weak to be detected consistently. Amplitude can be used as a secondary indicator of association.

The orientation of each reflector is determined by peak detection on its orientation signature. When the angular sampling resolution is coarse the results are too approximate to be of use. Instead the orientation is determined through fitting the region of the orientation signature near the peak to a parabolic curve and finding the maximum. In the experiments this method provided a bearing estimate within half of the scan resolution.

5.2.3.2 Finding Geometric Type

The geometric type is found through an estimate of $G$ based on the peak amplitude. Targets are specified as one of $\{\text{line, point}\}$ where point includes corners and edges.

5.2.3.3 Data Integration

To improve positional accuracy, an extended Kalman filter was used to integrate readings between successive positions. The Kalman filter combines a prediction of the next measurement with the actual measurement taking into account the uncertainty in each to determine the most likely target position (chapter 2). The algorithm must be supplied in advance with an estimate of the variance in measurement. This is pre-determined experimentally.
Target types were assigned when the value of $G$ fell within predetermined bounds on the same target in more than one position. Figure 5.2 shows the first stages of building up a map. "w" shows the line targets and "c" shows the point targets. Once targets are extracted, simple heuristics can be used (with care) to build up a map (e.g. a line target is bounded by point targets).

After the sensor was placed at all points on the $4 \times 4$ grid the final map was obtained as shown in Figure 5.3.

Around each point target is drawn an ellipse, which represents the bound of all points lying within eight standard deviations of the estimated target position. The position of this ellipse results from the Kalman filter, along with the estimate of the most likely target position. Because sonar is more accurate in finding range than orientation, the ellipses are elongated in the direction perpendicular to the line of sight from the sensor to the reflector.

A number of ghost targets can be seen. These are caused by two phenomena: from harmonics (resulting from non-linearity in the receiver) and
Fig. 5.3 The targets identified after the $4 \times 4$ grid is complete. Adjacent line targets are extended to form walls using simple heuristics [Kao and Probert 00]. The grid of robot positions is shown on a white background. The walls are shown as "w" and the point targets are shown as "c"s. The uncertainty ellipsoids for the points, which represent uncertainty at eight standard deviations, are also shown. The map may also include some ghost targets which were caused by multiple reflections of the sonar wave. A line target is wrongly identified as a point (marked with asterisk) on the right hand wall. Targets marked but with no label have not been assigned a type because the data is too uncertain (probably because of low amplitude) from multiple reflections in the environment. Heuristics based on amplitude and range can be used to eliminate these [Kao 99]. In addition there is an incorrectly identified target (point rather than line segment on the right hand wall).

On many commercial robots, the angular interval between transducers is several degrees. The robustness of the method to changes in angular interval was examined. For intervals between $1.8^\circ$ and about $10.8^\circ$ only one genuine target was lost (a corner on the north brick wall which corresponded to an uneven area) although the ghost targets, as we see in table 5.2, disappeared.

<table>
<thead>
<tr>
<th>scan resolution</th>
<th>1.8</th>
<th>3.6</th>
<th>5.4</th>
<th>7.2</th>
<th>9.0</th>
<th>10.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>targets</td>
<td>1.00</td>
<td>1.00</td>
<td>0.91</td>
<td>0.83</td>
<td>0.57</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Table 5.2 The number of targets recognised for different scan resolutions, expressed as a fraction of those recognised when the scan increment is $1.8^\circ$.

However the method is not robust if the smooth surface is replaced by a rough one. A new model is required. A suitable model is discussed in the next section.
5.3 Rough Surface Planar Models

Rough targets are altogether more complicated. A rough target is defined as one for which the variation in height is significant compared with wavelength. In this section we consider only planar surfaces.

As described in the last chapter, a reflection from a rough or textured surface may have two components, corresponding to both smooth and rough components of reflection. We see this in figure 5.4. The first part of the energy returned results from spurious effects such as direct feed through from the transmitter, electrical interference and ringing. The next part is the smooth component reflected from the surface. This peaks at the distance in the direction of normal incidence, which was 0.75m. The rough component occupies the next region. In theory it peaks at the angle along the bore-sight (centre line of the beam), about 1.06m, but in practice the noisy nature of the reflection (probably caused by differences in orientation of elements of the surface - see section 4.2.3 in chapter 4) - makes this hard to assess. Finally the signature trails off in noise.

![Fig. 5.4 A typical experimental range signature taken from a floor with an office carpet. The components of smooth and rough reflection are marked. The solid line shows the simulated reflection [Politis and Probert Smith 01]](image-url)
5.3.1 Backscattering Coefficient of Rough Surface

For a smooth surface power is received from only one point, and the reflections can be analysed geometrically using a ray model. However for a rough surface power is returned from every point within the sonar beam.

The signal incident on an element at position $(r, \theta)$ on the surface depends on its distance $r$ and the directivity of the transmitter $\Psi_t(\theta)$. The signal backscattered can be determined using boundary conditions on the forward and reverse waves and their derivatives. Details are given in [Bozma and Kuc 92].

A rough planar surface can be modelled using a Gaussian distribution over surface height [Bozma and Kuc 92]. We let the two dimensional characteristic function of a surface over $(x_1, x_2)$ be:

$$\chi(x_1, x_2) = e^{-s^2/2(x_1^2+Cx_1x_2+x_2^2)}$$  \hspace{0.5cm} (5.5)

where $s$ is the standard deviation in surface height and $C$ the correlation coefficient between the two dimensions. Assuming $C$ to depend only on a distance $\tau$, we write:

$$C = e^{-\frac{s^2}{\tau^2}}$$  \hspace{0.5cm} (5.6)

$T$ describes the correlation distance along the surface. If $T$ is small we would expect different parts of the surface to be similar on a small scale; if it is large we expect the surface to exhibit little regularity in pattern. Note that the scale is always considered relative to wavelength.

Because it is a statistical measure, the backscattering coefficient for a small element at an angle $\theta$ is described only in terms of its expectation $\langle|\rho|\rangle$. In [Kao and Probert 00] Kao and Probert show that:

$$\langle|\rho|\rangle = K \frac{TR}{2k \cos^3 \theta s} e^{-\frac{\sin^2 \theta T^2}{2s^2}}$$  \hspace{0.5cm} (5.7)

where $K$ includes terms independent of angle and roughness, $k$ is the wave number ($k = 2\pi/\lambda$ for wavelength $\lambda$), $\theta$ is the angle of incidence and $R$ the surface reflection coefficient as before. Note the dependence on the group $T/s$ which we identify as a surface texture parameter.
5.3.1.1 Finding Position of Rough Surfaces

For rough surfaces, it is clear from experimental data that the peak is not well defined in the range or orientation signatures. Therefore simple peak detection algorithms are unreliable. Instead a curve fitting method is used on the range signature.

Following Kao [Kao and Probert 00] we substitute from equation 5.7 into equation 5.1 to find the reflection for a point on a rough surface at angle $\theta$ to a sensor orientation $\beta_o$:

$$
\langle P_\theta \rangle = A_0 G L_{at}^{\text{pressure}} K \frac{TR}{2k \cos^3 \theta s} e^{-\frac{T^2}{2s^2} \tan^2 \theta} e^{-\frac{4(\theta - \beta_o)^2}{\beta_o^2}}
$$

(5.8)

where we use expression 3.11 in chapter 3 for the $\Psi_{t-r}(\theta)$. $L_{at}^{\text{pressure}}$ is approximately constant over the orientation signature of a single surface. For convenience we define $a$ such that:

$$
\langle P_\theta \rangle = \frac{a}{\cos^3 \theta} e^{-\frac{T^2}{2s^2} \tan^2 \theta} e^{-\frac{4(\theta - \beta_o)^2}{\beta_o^2}}
$$

(5.9)

Note that $a$ is independent of angle.

The peak amplitude in the range signature can be found by finding the derivative:

$$
\frac{\partial}{\partial \theta} \langle P_\theta \rangle = -\frac{a}{\cos^3 \theta} e^{-\frac{T^2}{2s^2} \tan^2 \theta + \frac{4(\theta - \beta_o)^2}{\beta_o^2}} \left[ \tan \theta + \frac{T^2}{2s^2} \tan \theta \sec^2 \theta + \frac{8(\theta - \beta_o)}{\beta_o^2} \right]
$$

Setting this to zero to find the peak leads to the equation:

$$
\frac{T^2}{s^2} \tan \theta \sec^2 \theta + \frac{8(\theta - \beta_o)}{\beta_o^2} + \tan \theta = 0
$$

(5.10)

Using the first order Taylor expansion: $\sec \theta \approx 1$ and $\tan \theta \approx \theta$ we find:

$$
\theta = \frac{4\beta_o}{\left(\frac{T^2}{2s^2} \beta_o^2 + 4\right)}
$$

(5.11)
The approximation in Equation 5.11 holds if $\theta_o^2 \ll 1$. As the bearing angle alters, the range $r_p$ at which the peak occurs is therefore:

$$r_p = r_o \sec \theta \approx r_o (1 + \theta^2 / 2)$$  \hspace{1cm} (5.12)

$$r_p = r_o \left(1 + \frac{16\beta_o^2}{2 \left[ \frac{3}{2s^2} \theta_o^2 + 4 \right]^2} \right)$$  \hspace{1cm} (5.13)

where $r_o$ is the distance in the direction of normal incidence.

It can be seen from equation 5.13 that the range of the peak is approximately a second order polynomial function with bearing angle, $\beta_o$. It is clear from equation 5.9 that the peak amplitude, $A_p$, varies exponentially with bearing angle. Using different values of bearing angle $\beta_o$ a set of values \{A_p, r_p, $\beta_o$\} can be set up. Then either of the sets \{A_p, $\beta_o$\} or \{r_p, $\beta_o$\} can be used to determine the parameters $T/s$, $r_o$ and the surface bearing.

Because fitting a low order polynomial is significantly more robust than fitting an exponential, the set \{r_p, $\beta_o$\} is preferred. Results are shown in table 5.3

<table>
<thead>
<tr>
<th>$r_o, m$</th>
<th>real $r_o, m$</th>
<th>bearing(°)</th>
<th>$T/s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.60</td>
<td>1.60</td>
<td>1.82</td>
<td>2.97</td>
</tr>
<tr>
<td>2.18</td>
<td>2.20</td>
<td>-0.07</td>
<td>3.44</td>
</tr>
<tr>
<td>2.77</td>
<td>2.80</td>
<td>0.47</td>
<td>5.16</td>
</tr>
<tr>
<td>3.36</td>
<td>3.40</td>
<td>-1.53</td>
<td>7.62</td>
</tr>
</tbody>
</table>

Table 5.3  Rough surface estimation

The first column is the estimated normal range to the rough surface and the second column is the actual hand-measured value. The third column is the estimated surface bearing; the actual bearing was zero. The last column is the estimated surface texture parameter.

Table 5.3 tabulates the estimation results. The error in the range estimate is a few centimetres and the error in bearing estimate is almost within 1.8°, the scan increment.

\(^\dagger\)For the sensor used in the experiment, $\theta_o = 0.17$ radians
We see that the last column $T_s$ changes with distance although it should be a property only of the surface type. This is because equation 5.13 has been derived for an infinite rough surface, whereas in the experiment at the longer distance some of the reflected power was from a smoother part of the surface too. The later figures therefore tend to overestimate (since $T_s$ is large for smooth surfaces).

5.4 Mapping Heterogeneous Environments

This section describes an extension of the mapping algorithm described in section 5.2.3.1 to include the rough surface model. The same room was used, with the east wall covered with a rough surface. To achieve this, the east wall was covered with aluminium foil, folded into small squares and then smoothed out, which experiments show makes a consistently rough surface.

Figure 5.5 shows the results when just the smooth model is used, as before. A number of false edges were identified on the rough wall. To overcome this the rough planar model was used as well.

First a decision was made as to whether the target was smooth or rough, using the texture parameter $T/s$, extracted as we have just described, to distinguish between them. If $T/s$ was above a threshold the surface was deemed to be smooth. Rough and smooth targets were then processed using the appropriate method and position and (for smooth targets) $G$ estimated.

This resulted in the map shown in figure 5.6. We see that this time the east wall is identified correctly: the false targets are eliminated.

5.5 Texture: Classifying Surfaces

The surface texture can also be used to provide a different feature for mapping. Rather than classifying a surface as just rough or just smooth, we can distinguish between surfaces.

A number of different features have been examined for classification, including those based on spatial frequency which are commonly used in vision and underwater sonar. However having available both the specular

$\dagger$The value depends on the surface examined; in this case it was taken as 4
and diffuse components of reflection has led to a particularly simple set of features on which recognition can be based.

5.5.1 Reflections from Real Surfaces

If we refer back to figure 5.4 we see a typical range signature with two components of reflection. Figure 5.7 shows the range signatures from a number of real surfaces, all taken from inside the Jenkin engineering building in Oxford. The first (surface$_1$) corresponds to a smooth flat desk. The second (surface$_2$) is an old tiled area, slightly rough and without a prominent pattern. The third (surface$_3$) is a floor covered by a plastic carpet introducing a certain degree of roughness. The fourth (surface$_4$) corresponds to a proper carpet, not very thick, while the fifth (surface$_5$) is a thick carpet. Finally the sixth surface (surface$_6$) is an asphalt pavement.

We see significant variation of the backscattered power with surface type. We can use the shape of the signature to distinguish between different surfaces.
First a suitable angle of view needs to be determined. If the angle is too small then most power is reflected along the normal direction whereas if it is too large very little power is returned. An angle of 45° was chosen as it provided good contrast experimentally. 

Note that although the differences between the classes is clear in the theoretical curves, as derived in [Politis and Probert Smith 01], this difference can be obscured by the peaks and troughs in the real data. These arise because the surface is in fact made up of a large number of elementary reflectors at different orientations; better models take these into account [Zografos and Probert 00]. In fact some of the signatures almost appear to show a periodic pattern although these peaks in fact are random in position. Fourier analysis can be used to distinguish these, or a method based

§For a sensor of substantially different beam-width, a different angle would be more suitable. In general it is important to ensure that reasonable power is incident on the surface along the direction of the normal.
on the spacing of the peaks [Politis and Probert Smith 01].

Fig. 5.7 Experimental (dashed lines) range-amplitude profile from a number of surfaces under an angle of 45°. Solid lines show theoretical predictions. (a) corresponds to \( \textit{surface}_1 \), (b) to \( \textit{surface}_2 \) etc.

### 5.5.2 Pathways Classification

The shape of the range signature depends in a complicated way on physical quantities associated with the surface. It is not possible to pick out, say, the mean deviation in height, from such a signature. However what is possible is to distinguish between a number of candidate surfaces, using methods which are based on training a classifier. Texture classification is a useful part of a localisation system in which the candidate surfaces can be observed beforehand.

A number of classifiers might be used to distinguish between surfaces. Experiments have taken place with neural networks such as multilayer perceptrons and radial basis functions, and with statistical classifiers. For the small training set used here, a statistical classifier, the K-nearest neighbour (KNN) classifier [Bishop 95], gave the best results.

The KNN classifier was developed for the \( \textit{surfaces}_{1-6} \). Sixteen mea-
measurements were taken of each surface, looking downwards with a bore-sight
elevation angle of $45^\circ \pm 2^\circ$. The readings were taken at different points
on the surface (in position and azimuth) so the readings are typical of the
surface properties rather than of a particular view of the surface.

5.5.3 Finding Suitable Features

The choice of classifier often matters much less than the features on which it
will be trained. Some workers have attempted to train classifiers on the full
range signature [Dror et al 95; Harper and McKerrow 95b] but an enormous
amount of training data is required to get any reasonable generalisation. It
is normally much more successful if suitable features are identified first.

The distribution of energy with range proved successful as the prime
indicator of surface type. Let $E_{\text{smooth}}$ and $E_{\text{rough}}$ correspond to the energies
of the smooth and rough components of the range signature respectively:

$$
E_{\text{smooth}} = \int_{l_0}^{l_1} S_a^2(r)dr, \quad E_{\text{rough}} = \int_{l_1}^{l_2} S_a^2(r)dr.
$$

(5.14)

where $S_a(r)$ is the amplitude at a distance $r$. $l_0$ and $l_2$ are chosen to match
the beam width and range to the surface. The partitioning between the
smooth and rough surfaces is determined by the distance $l_1$, which was
chosen by examining the images. It will vary with a different radiation
patterns, different bore-sight angle or range. The values used in these
experiments were: $l_0 = 0.75m$, $l_1 = 0.85m$, $l_2 = 1.5m$. Figure 5.8 plots
$E_{\text{smooth}}$ and $E_{\text{rough}}$ for the data gathered.

The K-nearest neighbour classifier has two phases. In the training phase
it attempts to find cluster centres and boundaries for each class based upon
these six distributions. In the generalisation phase it is presented with a
new unknown point and assigns it to one of these clusters. The parameter $K$
determines how many data points will be within each cluster and essentially
acts as a smoothing parameter. If $K$ is too large the classifier is poor at
discrimination, too small and it will not learn.

The success of the classifier depends on how well the classes are sepa­
rated in the feature space. In this case separation is excellent as we see in
Figure 5.8. The success rate of the classifier was above 99% for most values
of $K$. Other experiments with outdoor paths (gravel, grass, planted areas,
tarmac) have shown that other features may be used as well. A multi-layer
perceptron classifier is being used to handle large data sets. A particular advantage of this method is that it is robust to uncertainty in orientation of a few degrees.

5.5.4 Remarks

Texture recognition is a useful tool in localisation, especially to complement a geometric feature based method. Downwards and sideways facing sensors could therefore provide valuable information to assist in localisation, especially in environments in which there are not many structural features (for example walking along pavements, in open spaces).

5.6 Summary

In this chapter we have used simple models of smooth and rough surfaces, to build maps in regions with both smooth and rough features. In addition we have described the use of sonar for recognising texture through exploiting both specular and diffuse components of the reflected power.
Acknowledgment

This work is described more fully in [Kao and Probert 00; Politis and Probert Smith 01] (©1999 Sage Publications, Inc.).
Chapter 6

Sonar Systems: A Biological Perspective

Herbert Peremans

In this chapter we will review biologically inspired approaches to ultrasonic sensing. These are man-made systems which implement functional principles of their biological counterparts, i.e. bat echolocation systems. It is expected that by using such an approach the performance of artificial ultrasonic sensing systems can be improved considerably.

6.1 Introduction

One of the most important features of a navigation sensor, i.e. a sensor that provides a mobile robot or a robot arm the information it requires to move around in the real world, is the rate at which it produces information about the environment. Ideally, such a sensor would generate very accurate information about a very large area at a very fast measurement rate. Unfortunately, conventional sonar range sensors, e.g. the Polaroid range sensor [Biber et al 80], suffer from a very low rate of information as can be understood by looking at their operation.

A conventional sonar range sensor operates by transmitting a pulse and subsequently measuring the time of flight of the first detected echo. Usually the same transducer is used both as a transmitter and a receiver. Knowing the speed of sound allows this time of flight to be transformed into a distance measure, i.e. the distance from the transmitter to the reflecting object and back to the receiver.

Such a measurement scheme allows a fairly accurate distance measurement but determining the bearing of the reflecting object is very inaccurate. Indeed, as can be seen from Fig. 6.1, the angular accuracy is effectively determined by the beam-width of the transducer, e.g. the beam-width of the Polaroid transducer is given by $\varphi \simeq 25^\circ$.

Attempts at increasing the amount of information per measurement
or equivalently decreasing the uncertainty left after the measurement by restricting the beam width of the ultrasonic transducer, as is sometimes suggested [Crowley 85; Brown 85], are missing the point. This can be understood by looking more closely at the results of a measurement (Fig. 6.1): it states that an object is present in the object zone, which is determined by the beam width $\varphi$ and the measured distance $r$ from the sensor. But also and equally important, it states that no object is present in the region between the sensor and the object, the empty zone. Restricting the beam width trades information about empty space for information about occupied space, necessitating a larger number of measurements to sweep a specified region.

Also, the conventional sonar sensor processes the signal received up to the first echo only thereby hiding all reflecting objects behind the one closest to the sensor. Finally, the output of the sensor is limited to time of flight information making it impossible to extract reflector shape information. From these considerations we conclude that the conventional sonar sensor does indeed extract only a very limited amount of information from each measurement.

The small amount of information extracted from a single measurement is only half the explanation for the low information rate of a conventional sonar sensor, equally problematic is its low measurement rate. The mea-
sourement rate of every sonar sensor, as it has to wait for the transmitted sound pulse to arrive at the receiver, is fundamentally limited by the speed of sound (\( \approx 340 \) m/s). Note that this measurement latency makes it very difficult to trade number of measurements for amount of information extracted from a single measurement. Indeed, we believe that the often stated equivalence regarding the information content of a large number of simple sensor measurements and a small number of sophisticated sensor measurements does not apply to sonar sensors because of this slow measurement rate.

The low information rate would seem to indicate that sonar sensors are not very well suited to be used as navigation sensors. Nevertheless, considerable evidence has been collected [Griffin 58] showing that bats rely almost exclusively on their airborne sonar system to navigate and hunt for prey. We believe that the discrepancy between the performance of natural echolocation systems and artificial ones can be explained by three principal features of natural sonar systems that are missing from the conventional range sensor. First, advanced signal processing is applied to the entire echo signal received (the sonar signature of chapter 4, allowing the extraction of much more information than the time of flight of the first echo. Next, multiple transmitters/receivers schemes are used to create correlations between the echo signals received. These correlations contain additional information about the environment. Finally, the sensor system has the capability to actively, i.e. by adapting some of its configuration parameters, explore the environment allowing a more efficient extraction of information.

The systems described in this chapter attempt to remedy the low information rate of the conventional sonar sensor by proposing schemes that significantly increase the amount of information extracted from a single measurement. They will all do this by realising to varying degrees some or all of the "biological sonar" features indicated above.

6.2 Echo Formation

Before discussing the extraction of information from the echo signal received by the different sensor systems we first summarise the echo formation process itself, i.e. the process by which environmental features give rise to particular features of the echo signal.
6.2.1 Transformations

A measurement of an ultrasonic sensor consists of the following operations, depicted in Fig. 6.2(a):

1. An electrical signal is applied to the transducer which causes the surface of the transducer to move.
2. The movement of the transducer surface generates an acoustical wave.
3. The acoustical wave travels through the air.
4. Upon hitting an object part of the acoustical wave is reflected and returned towards the receiver.
5. Upon reception of the reflected echo the surface of the transducer is made to move accordingly.
6. This movement is again transformed in an electrical signal.

Note that in many sensor systems, receiver and transmitter are the same physical transducer.
Echo Formation

Assuming that each of the steps above can be faithfully modelled by a linear transformation, Fig. 6.2(b), we can write the signal received:

\[ r(t) = s(t) \ast h_{tr}(t) \ast h_{rad}(t) \ast h_{air}(t) \ast h_{refl}(t) \ast h_{rad}(t) \ast h_{rec}(t), \]

with \( s(t) \) the signal used to excite the transmitter, \( \ast \) denoting convolution, and the \( h(t) \)'s denoting the impulse responses corresponding with the different filter operations as detailed below. The validity of this filter model is confirmed by the good agreement between its predictions and the experimental data [Kuc and Siegel 87; Kleeman and Kuc 95; Peremans 97].

This expression is equivalent to a transform into the time domain of equation 3.3 in chapter 3, but includes explicitly the transducer impulse responses \( h_{tr}(t) \) and \( h_{rec}(t) \).

We consider each of these terms. In the expressions which follow \( f_0 \) is the resonant frequency of the transducer and \( B \) its bandwidth; with \( \theta \) the off-axis angle of the observation point [Bozma and Kuc 91], \( v_s \) the speed of sound in air, \( a \) the radius of the circular transducer and \( J_1() \) the Bessel function of the first order [Morse and Ingard 68; Peremans 97]. \( \alpha_{dB}(f) \) is the absorption coefficient of air in decibels/meter at frequency \( f \), \( r_1 \) and \( r_2 \) the distances between transmitter and reflector and between reflector and receiver respectively and \( r_{eq} \) determines the effective spherical spreading of the wave-fronts [Kleeman and Kuc 95; Kuc 94].

1. \( h_{tr} \) and \( h_{rec} \) describe the low pass filtering effect of the transmitter and receiver. This effect is mentioned in chapter 4 which discussed how the transmitter and receiver imposed a Gaussian amplitude modulation on the signal. Transferring this into the frequency domain we write:

\[
H_{tr}(f)H_{rec}(f) = e^{-\frac{(f-f_0)^2}{2\sigma^2}},
\]

(6.1)

2. \( h_{rad} \) describes the radiation pattern (chapter 3, section 3.3), which in general depends on both angle and frequency. In the frequency domain,

\[
H_{rad}(f, \theta) = \frac{J_1(ka \sin \theta)}{ka \sin \theta},
\]

(6.2)

where \( k = 2\pi f/v_s \). \( H_{rad} \) is equivalent to \( \Psi_t = \Psi_r \) in equation 3.3
(3) Finally we describe the attenuation from loss and spreading.

\[ H_{\text{air}}(f) = \frac{10^{-\alpha_{dB}(f)(r_1+r_2)}}{r_{eq}}e^{-j2\pi f \frac{r_1+r_2}{v_s}}, \tag{6.3} \]

\( H_{\text{air}}(f) \) includes terms in loss and spreading which were included in equation 3.3.

Below, we discuss in more detail the only remaining transformation \( h_{\text{refl}}(t) \), with Fourier transform pair \( H_{\text{refl}}(f) \). \( H_{\text{refl}}(f) \) determines the form of \( r_{eq} \) in \( H_{\text{air}}(f) \).

6.2.2 Reflection

We adhere to Kuc's acoustic imaging model [Kuc and Siegel 87], based on geometrical wave propagation in a specular environment. Furthermore, all except the last sensor system described in this chapter assume geometrical wave propagation in a horizontal plane, equivalent to the common assumption of a 2D-world model. As a consequence, the position of a reflecting object is defined by the location of its intersection with the horizontal plane through the sensor. Note that if the objects are not cylindrical, this assumption might result in erroneous readings. We also adhere to Kuc's categorisation of reflectors into three classes (Fig. 6.3), because of the close agreement between the results predicted by this model and the real measurements in indoor environments.

6.2.2.1 Reflections from a Planar Reflector

We take a planar reflector to act like a perfectly smooth and rigid reflector [Morse and Ingard 68], i.e. it behaves like a mirror to the acoustic waves.

In that case, we can replace the situation where the sound waves arrive at the receiver after reflection off the planar reflector by one where the latter is taken away and a virtual transmitter is placed at the mirror image of the real one as shown in Fig. 6.3(b). From this figure we conclude that the wave-front arriving at either receiver after reflection can not be distinguished from one originating at the virtual transmitter. Hence the parameter describing the spherical spreading of the wave-front in equation (6.3) should be set to \( r_{eq} = r_1 \) or \( r_{eq} = r_2 \) respectively.

Note that this specular reflection model is no longer valid in the presence of rough surfaces and has to be replaced by a more complicated one.
[Bozma and Kuc 91]. However, due to the fairly long wavelength of sound, 7 mm at 50 kHz, most reflecting surfaces in indoor environments result in predominantly specular reflections.

6.2.2.2 Reflections from a Corner

Two perfectly smooth and rigid planar and orthogonal reflectors that intersect are called a corner if the transducer is on the inside of the intersection angle (Fig. 6.3). In that case, [Pierce 94] the reflector behaves like a 2D corner cube reflector, i.e. sound waves are reflected back in the same direction they arrive from. As with the planar reflector we can replace the situation where the sound waves arrive at the receiver after reflection off the corner by one where the latter is taken away and a virtual transmitter is placed at the mirror image of the real one as shown in Fig. 6.3(c).

From this figure we conclude that the spherical wave-front arriving at the receiver is again identical to one coming from the virtual transmitter at range \( r_{eq} = r_1 \) or \( r_{eq} = r_2 \) respectively (equation (6.3)). Hence, the corner and a plane through the centre of the corner and orthogonal to the line through the transmitter and the centre of the corner result in the same position for the virtual transmitter. However, because a corner gives rise to two consecutive reflections the orientation of this virtual transmitter is
different than would be the case for a single reflection on a planar reflector as can be seen by comparing Figs. 6.3(b) and (c).

As stated above, our reasoning is based on geometric wave propagation. The validity of this approximation is guaranteed as long as all geometric features present in the environment are large compared to the wavelength. However, this is no longer true in the presence of a corner. Solving the wave equation for a corner [Pierce 94] one finds in addition to the reflected wave, described above, another diffracted component originating at the intersection between the two planes. But, because this diffracted component is so much smaller than the reflected one, it can be safely neglected for our purposes.

6.2.2.3 Reflections from an Edge

For the third class of reflector, i.e. the edge, the diffracted component will be the only signal present since, as can be seen in Fig. 6.3(a), the reflected component, i.e. the component based on geometric wave propagation, will be directed away from the receiver. In that case it can be shown [Kuc and Siegel 87] that the wave-fronts arriving at the receivers after diffracting off the edge behave as if the sound waves arriving at the edge are retransmitted by a point source located at the edge. Hence with the notation used in equation (6.3) the spherical spreading is described by $r_{eq} = r_1 r_2$. The proportion of energy “retransmitted” by the edge is a complex function of the exact shape of the edge [Pierce 94] but will not be considered here.

This completes the system theoretic model for the echo formation process as shown in Fig. 6.2.

6.3 Monaural Sensing

We will call every ultrasonic sensor system consisting of one transmitter and one receiver, whether the same transducer is used for both the transmitter and the receiver or not, a monaural sensor. The multi-aural sensor systems described below will consist of multiple such monaural sensors. Processing a measurement of a multi-aural sensor system consists of two parts: first, for each monaural sensor the relevant information is extracted from the signal received; next, the information from the multiple monaural systems is combined into the desired environment feature estimates.
6.3.1 Inverting the Echo Formation Process

From the above description of the echo formation process we conclude that different features of the echo signal, i.e. echo delay, echo strength and echo spectrum, can be related to different features of the reflecting object. The task of the sonar sensors described below will be to invert the transformations applied during the echo formation process and deduce useful information about the reflector from these features of the echo signals received.

We look upon the signal processing performed on the data received by a monaural sensor as a pre-processing operation. In particular, to support both reflector localisation and reflector recognition capabilities the intermediate data representation resulting from this pre-processing should simplify the extraction of: arrival time information of multiple echoes, amplitude information of multiple echoes, spectral content of multiple echoes. The biologically inspired approach described below has the advantage of giving rise to a data representation that makes extraction of all three types of information possible.

6.3.2 Extraction of Information: Cochlear Processing

The signal processing operations performed upon the echoes received, denoted as ‘cochlear processing’, are based upon a simplified model of the processing performed by the mammalian cochlea [Slaney 93], i.e. a joint time-frequency analysis of the incoming signal. We model this analysis by a bandpass filterbank, i.e. gammatone filters, with subsequent envelope extraction in each channel, Fig. 6.4(a).

The layout of the auditory filterbank follows the general mammalian pattern of keeping filter quality constant \( Q = 20 \) as centre frequency varies \((35\text{kHz-160kHz})\). Keeping filter quality constant implies that bandwidth \( (f_{bw}) \) is a linear function of centre frequency \( (f_c) \): \( f_{bw} = f_c/Q \). Additionally, overlap between neighbouring filters is kept constant [Slaney 93], which results in a wider spacing of filters as bandwidth increases. All these properties are evident in the frequency responses shown in Fig. 6.4(b).

The outputs of the filterbank are subsequently processed with a half-wave rectifier and a low-pass filter, a simple yet fairly accurate model of the neural transduction mechanism. These last two steps are equivalent to an amplitude demodulation scheme that approximately recovers the envelopes.
of the outputs of the bandpass filters in the filterbank. The logarithms of
the outputs of these lowpass filters are taken to be the final, spectrogram­
like, output of the system. In section 6.4.4 we describe a sensor system
that makes explicit use of this spectrogram representation generated by
the cochlear model.

Although the sensor systems described below will make use of classic
correlation techniques to accurately estimate the arrival times of the echoes
it is shown by Altes [Altes 80] that a spectrogram representation of the
signal received could also be used. The accuracy of the echo arrival time and
echo energy estimates can be shown to be equivalent to that of classic semi­
coherent and coherent receivers. A biologically motivated implementation
of such a fully coherent receiver making use of this spectrogram-like data
representation is described in [Saillant et al 93; Peremans and Hallam 98].

The increased flexibility of this scheme compared with a standard spec­
trogram is an additional benefit. The experimenter now has complete con­
trol over the parameters, i.e. central frequencies and Q-factors, defining
the behaviour of the filters in the filterbank. Hence, the filterbank can be
specifically designed to facilitate the extraction of particular features of the
echoes [Peremans et al 00; Walker et al 98].

6.4 Multi-Aural Sensing

Monaural sensors that process the entire signal to extract echo amplitudes,
echo arrival times, and possibly echo spectrograms already constitute a
significant improvement over the standard time-of-flight (TOF) ultrasonic range sensor [Biber et al 80] with respect to amount of information extracted from a single measurement. Nevertheless, they still suffer from some of the same limitations: angular localisation capabilities limited by the transducer beam-width, very limited object recognition capabilities. We will now show that if we combine monaural sensors, i.e. construct multi-aural sensors consisting of multiple transmitters and receivers, we can improve upon these monaural sensors considerably.

All but the last multiple receiver system described below will use arrival time information to locate the reflecting object(s), both distance and bearing, by triangulation. To recognise the reflector type some sensor systems will use additional arrival time information whereas others will use amplitude information. Nevertheless, it should be realised that for all these sensor systems localisation and object recognition capabilities are linked together. Since, depending on the reflector’s type, the measurement results, i.e. the measured arrival times, should be interpreted differently to determine the position of the reflector. The two possible triangulation scenarios for the three classes of elementary reflector types are represented in Fig. 6.3. When looking at an edge all three reflections originate from one point, whereas when looking at a plane the reflections come from three different points. As stated before, the latter situation is equivalent to one where all three reflections originate at a virtual transmitter placed at the mirror image of the real one (Fig. 6.3(b)). The third type of reflector, i.e. a corner, gives rise to exactly the same pattern of arrival times as a planar reflector (Fig. 6.3(c)). Hence, when the reflector is a corner the triangulation scenario is identical to the one corresponding with a plane.

Depending on the number of transmitters, the number of receivers and whether the output of the receivers is limited to arrival time information only or provides amplitude information as well different multi-aural sensor systems have been proposed.

6.4.1 **Echo Amplitude and Echo Arrival Time: Two transmitters, Two receivers**

6.4.1.1 **Sensor Setup**

The sensor consists of two transmitter/receiver pairs as shown in Fig. 6.5. The measurement cycle of this sensor is made up of the following steps
First, transmitter 1 emits a pulse. Next, at each of the two receivers, the arrival time as well as the amplitude of the echo are extracted from the raw sensor signals. We denote the corresponding travelled distances by \( m_{11}', m_{21}' \) and the amplitudes by \( A_{11}', A_{21}' \). Next, transmitter 2 emits a pulse and again at each receiver the arrival time and amplitude of the echo are extracted. We denote the travelled distances by \( m_{12}', m_{22}' \) and the amplitudes by \( A_{12}', A_{22}' \).

### 6.4.1.2 Localisation of Planes and Corners

Based on the measurement situation, Fig. 6.5(a), it can be shown [Barshan and Kuc 90] that the position \((r, \theta)\) of a plane is given by

\[
\begin{align*}
  r &= \frac{m_{11}' + m_{22}'}{4}, \\
  \theta &= \arcsin\left(\frac{m_{22}' - m_{11}'}{2d}\right)
\end{align*}
\]

where \(d\) is the distance between the centres of the transducers. If the reflector is a corner, Fig. 6.5(b), this turns into:
\[
\frac{r}{4} = \frac{m_{12'} + m_{21'}}{4}
\]

\[
\theta = \begin{cases} 
\arcsin\left(\frac{(m_{22'} - s) - m_{12'}}{2d}\right) & \text{if } m_{22'} - s > m_{12'}, \\
\arcsin\left(\frac{(m_{11'} - s) - m_{12'}}{2d}\right) & \text{if } m_{11'} - s > m_{12'},
\end{cases}
\]

with

\[
s = \frac{\sqrt{m_{12'}^2 + d^2} - m_{12'}}{2}.
\]

Note that the position of a corner determines the point of intersection of the two planes but it does not determine the orientation of the corner.

To use these equations one needs to know whether the reflector is a plane or a corner, as this will determine whether equation 6.4 or equation 6.5 is required to transform the measurement vector. In the next section we describe a statistical test to distinguish between these two reflector types.

6.4.1.3 Recognition of Planes and Corners

In section 6.2 we noted how the directivity of the transducer results in angle dependent filtering both during transmission and reception of a pulse. For a very short pulse, i.e. a wide bandwidth pulse, the resulting directivity pattern no longer contains any nulls or side-lobes so typical for the directivity pattern of a long duration/single frequency pulse. This can be understood by observing that the positions of the nulls and side-lobes are frequency dependent and hence a wideband pulse gives rise to a continuum of nulls and peaks resulting in a net smooth beam.

From experiments [Barshan and Kuc 90], it is found that instead the directivity pattern of the transmitting (receiving) transducer can be well approximated by a Gaussian beam pattern. Hence, the pressure amplitude can be written

\[
A(\theta) = A_{\max}e^{-\frac{2\theta^2}{\theta_0^2}},
\]

with \(\theta_0 \approx 10^\circ\) for the Polaroid transducer. When both the directivity of the transmitter and that of the receiver is taken into account the angular
dependency of the amplitude of the signal received is given by

\[ A(\theta_1, \theta_2) = A_{\text{max}} e^{-\frac{2\theta_1^2}{\theta_0^2}} e^{-\frac{2\theta_2^2}{\theta_0^2}}, \]

with \( \theta_1 \) and \( \theta_2 \) the inclination angles of the transmitter and receiver respectively (Fig. 6.6) and \( A_{\text{max}} \) the amplitude measured if transmitter and receiver are aligned.

Applying this result to the situations shown in Fig. 6.5 we get for the plane

\[
\begin{align*}
A_{11'}(\theta) &= A_{22'}(\theta) = A_{\text{max}} e^{-\frac{4\theta^2}{\theta_0^2}}, \\
A_{12'}(\theta) &= A_{21'}(\theta) = (A_{\text{max}} e^{-\frac{4\theta^2}{\theta_0^2}}) e^{-\frac{4\theta^2}{\theta_0^2}},
\end{align*}
\]

with

\[
\theta_s = \arctan \frac{d \cos(\theta)}{2r} \simeq \arctan \frac{d}{2r},
\]

for \( \theta < \theta_0 \) with \( \theta_0 = 10^\circ \). For the corner we get

\[
\begin{align*}
A_{11'}(\theta) &= A_{\text{max}} e^{-\frac{4(\theta - \theta_a)^2}{\theta_0^2}}, \\
A_{22'}(\theta) &= A_{\text{max}} e^{-\frac{4(\theta + \theta_a)^2}{\theta_0^2}}, \\
A_{12'}(\theta) &= A_{ba'}(\theta) = A_{\text{max}} e^{-\frac{4\theta^2}{\theta_0^2}},
\end{align*}
\]

with

\[
\theta_s = \arctan \frac{d \cos(\theta)}{2r + d \sin(\theta)} \simeq \arctan \frac{d}{2r},
\]

for \( \theta < \theta_0 \).
From Fig. 6.7 it can be seen that to differentiate between corners and planes in the presence of additive, zero-mean, Gaussian noise we can use the following decision algorithm

if \((A_{11}'(\theta) - A_{12}'(\theta) > 6\sigma AND A_{22}'(\theta) - A_{12}'(\theta) > 6\sigma)\) then PLANE

if \((A_{12}'(\theta) - A_{11}'(\theta) > 6\sigma OR A_{12}'(\theta) - A_{22}'(\theta) > 6\sigma)\) then CORNER,

with \(\sigma\) the standard deviation of the amplitude noise. The \(\pm 3\sigma\) boundaries around the expected values of the amplitudes include 99.7\% of the measurements based on the Gaussian noise assumption. Note that these conditions are mutually exclusive so when the amplitudes differ by more than \(6\sigma\) a decision will always be reached. However, as \(\theta\) grows larger (Fig. 6.7) it gets more likely that neither of the two conditions holds. In that case it will no longer be possible to determine the reflector type based on this test.

### 6.4.2 Echo Arrival Time Information: Two Transmitters, Two Receivers

The previous sensor allowed localisation and recognition of planes and corners but not edges. The sensor described now will allow both localisation and recognition of all three primitive reflector types: edges, planes and corners.
6.4.2.1 Sensor Setup

The sensor consists of a transmitter/receiver and a second receiver positioned close together plus a second transmitter positioned further away as shown in Fig. 6.8. One transducer is used both as a transmitter and a receiver; the other two are used either as a receiver only or as a transmitter only. As was the case with the previous sensor system each transmitter-receiver pair forms an accurate, monaural, range sensor.

The measurement cycle of this sensor is made up of the following steps [Kleeman and Kuc 95]. First, transmitter 1 emits a pulse. Next, at each of the two receivers, the precise arrival time of the echo is extracted from the signal received. This echo arrival time information is extracted by a classic matched filter/peak detection approach (chapter 4). We denote the corresponding travelled distances by $m_1, m_2$. Next, transmitter 2 emits a pulse and the precise arrival time of the echo is again extracted from the signals at each of the two receivers. We denote these travelled distances by $m_3, m_4$.

6.4.2.2 Localisation of Edges and Planes/Corners

From triangulation one can show [Kleeman and Kuc 95] that the position, Fig. 6.9, of the reflector $(r, \theta)$ can be derived from the information extracted from the first transmission. The position is given by

\[
\begin{align*}
    r &= m_1/2, \\
    \theta &= \arcsin\left(\frac{(r_1 - r_2) \cos \phi}{d}\right) + \phi, \\
    \phi &= \frac{1}{2} \arcsin\left(\frac{d \cos \theta}{r_2}\right),
\end{align*}
\]
where

\[
\begin{align*}
(r_1, r_2) &= (m_1, m_2)_{p/c}, \\
(2r_1, r_1 + r_2) &= (m_1, m_2)_e,
\end{align*}
\]

(6.8)

(6.9)

for a plane/corner and an edge, respectively. This position refers to point \( P \) in the figure. As is clear from a comparison of Fig. 6.9 and Fig. 6.3, \( P \) will coincide with the true position of an edge whereas in the case of a plane or a corner \( P \) will denote the position of the virtual image of the transmitter.

As was the case before, one needs to know whether the reflector is an edge or a plane/corner to use these equations, since this will determine whether equation 6.9 or equation 6.8 is required to transform the measurement vector. In the next section we will describe a statistical test to recognise these three reflector types.

6.4.2.3 Recognition of Edges, Planes and Corners

The measurement situation for a plane and a corner reflector is shown in Figs. 6.10(a),(b). If we denote by \( r_1 \) and \( \theta_1 \) the position of the virtual transmitter 1 and by \( r_2 \) and \( \theta_2 \) the position of the virtual transmitter 2 with respect to the sensor system we can see from Fig. 6.10 that both for a plane and a corner

\[
\begin{align*}
r_2 &= \sqrt{r_1^2 - 2r_1b \sin \theta_1 + b^2}, \\
\theta_2 &= \theta_1 + \beta,
\end{align*}
\]
Fig. 6.10 (a) Plane reflector, (b) corner reflector, (c) edge reflector

with $\beta_{\text{plane}} = -\beta_{\text{corner}} = \arctan\left(\frac{b \cos \theta_1}{r_1 - b \sin \theta_1}\right)$. Hence, planes and corners can be distinguished based on the sign of $\beta$. We now show that edges can be recognised this way too.

The measurement situation for an edge reflector is shown in Fig. 6.10(c). From this figure we conclude that for an edge the position of the point of reflection always coincides with the position of the edge itself. Consequently, the bearing of the reflection from an edge is the same for the two transmitters, i.e. $\theta_1 = \theta_2$. Furthermore, if we denote the distance between transmitter 1 and the edge by $r_1$ and the distance travelled by the pulse originating at transmitter 2 and reflected by the edge towards receiver 1 by $r_2$ (Fig. 6.10(c)) it can be seen that

$$r_2 = r_1 + \sqrt{r_1^2 - 2r_1b \sin \theta_1 + b^2},$$

$$\theta_2 = \theta_1 + \beta,$$

with $\beta_{\text{edge}} = 0$.

Based on the relations derived above one can estimate the values of
$r_2$ and $\theta_2$ from the measurements of $r_1$ and $\theta_1$ given an hypothesis about the particular type of reflector present. The actually measured values of $r_2$ and $\theta_2$ can then be compared with these expected values and assuming a particular noise model the statistical significance of the deviations can be determined. Such a statistical test is derived in [Kleeman and Kuc 95] assuming a zero-mean Gaussian noise model. The test associates, based on the chi-square distribution, a confidence level with each reflector type. The reflector type hypothesis for which the confidence level is acceptable is taken to be the result of the classifier. If none or more than one reflector type has an acceptable confidence level the reflector is classified as being of type “unknown”.

6.4.3 Echo Arrival Time Information: One Transmitter, Three Receivers

Like the previous sensor, the sensor described here will make use of arrival time information only to allow both localisation and recognition of all three primitive reflector types: edges, planes and corners. Indeed, it is shown in [Kleeman and Kuc 95] that “two transmitters and two receivers are necessary and sufficient for discriminating planes, corners and edges in two dimensions” whenever the sensor is limited to extracting arrival time information from the echoes only. This condition is fulfilled by the current sensor setup by exploiting the movement of the sensor in between successive measurements.

6.4.3.1 Sensor Setup

The sensor as shown in Fig. 6.11, consists of one transmitter/receiver and two additional receivers: the three transducers are lined up and spaced distance $d$ apart. The central transducer is used both as a transmitter and a receiver; the two peripheral ones are used solely as receivers. As before, each transmitter-receiver pair forms an accurate, monaural, range sensor.
Fig. 6.12 Measuring distance of flight for a (a) plane, (b) corner, (c) edge

The measurement cycle of the sensor is made up of the following steps [Peremans et al 93]. First, the central transducer emits a pulse. Next, at each of the three receivers, the precise arrival time of the echo is extracted from the raw sensor signals. This echo arrival time information is extracted by a classic matched filter/peak detection approach (chapter 4). We denote the corresponding travelled distances by $m_1$, $m_2$ and $m_3$.

### 6.4.3.2 Localisation of Edges and Planes/Corners

Based on similar geometrical considerations as for the previous sensor system (Fig. 6.12) one can show [Peremans et al 93] that the position of the reflector $(r, \theta)$ is given by

\[
\begin{align*}
    r &= m_1/2 \\
    \theta &= \arcsin \left( \frac{r_3^2 - r_2^2}{4dr_1} \right),
\end{align*}
\]

where

\[
\begin{align*}
    (m_1, m_2, m_3)_{p/c} &= (r_1, r_2, r_3) \quad \text{(6.11)} \\
    (m_1, m_2, m_3)_e &= (2r_1, r_1 + r_2, r_1 + r_3), \quad \text{(6.12)}
\end{align*}
\]

for a plane/corner and an edge, respectively. Note that no distinction can be made between a plane and a corner based on the measured distances.
Knowledge of the type of reflector, whether it is an edge or a plane/corner, is again necessary to effectively localise the reflector using these equations, as the reflector type will determine whether equation 6.12 or equation 6.11 is required to transform the measurement vector. In the next section we describe statistical tests to recognise the three reflector types.

6.4.3.3 Recognition of Edges, Planes and Corners

To distinguish between edges, planes and corners two statistical tests are proposed [Peremans and Campenhout 92; Peremans and Campenhout 93]: one that distinguishes between edges and planes/corners and another that distinguishes between planes and corners/edges. By combining the results of these tests which can be evaluated in parallel, all three object types can be recognised.

**Discriminating between edges and planes/corners**

To discriminate between edges and planes/corners the first statistical test is based on the measurement vectors corresponding to planes/corners and edges lying on different surfaces in \((m_1, m_2, m_3)\)-space. These surfaces can be readily determined by realising that the identity

\[ r_2^2 + r_3^2 = 2(r_1^2 + d^2). \]

is valid for both classes of reflectors, Fig. 6.12. Applying the appropriate transformations, i.e. equation 6.12 and equation 6.11, to this constraint and denoting by \(f_{p/c}(\cdot) = 0\) and \(f_e(\cdot) = 0\) the surfaces corresponding to a plane/corner reflector and an edge respectively, one obtains

\[
\begin{align*}
    f_{p/c}(\mathbf{m}) &= 2m_1^2 + 2d^2 - m_2^2 - m_3^2, \\
    f_e(\mathbf{m}) &= m_1(m_2 + m_3) + 2d^2 - m_2^2 - m_3^2.
\end{align*}
\] (6.13) (6.14)

Determining which surface the measurement vector is closest to the test distinguishes between edges and planes/corners in the presence of noise.

**Discriminating between planes and edges/corners**

The test derived above treats corners and planes identically. Hence, another test is needed to discriminate between those two types of reflectors. As can be seen from Fig. 6.12 a corner behaves identically to a plane orthogonal to the line of sight of the sensor. However, when we change the viewpoint
(Fig. 6.13), the new line of sight stays parallel to the old one in the case of a plane, whereas the line of sight will rotate in the case of a corner. This can be understood by noting that the line of sight will always go through the centre of a corner, as can be seen in Fig. 6.12. Hence the centre of the corner as it is a fixed point will be the centre of rotation for the line of sight as the sensor is moved. Therefore, by testing whether the lines of sight meet at a real point, i.e. the centre of the corner, or at infinity we can discriminate between corners and planes. Note that, with respect to this test, edges will behave the same way as corners do making it complementary to the first test described above.

For this test the sensor readings consist of the object's distance and bearing as given by equation 6.10 with respect to the different positions of the sensor. First, the intersection point from a pair of measured directions $(\theta_i, \theta_{i+j})$ is calculated. Next, the actually measured distance $r_{i+j}$ is divided by the distance $r_s$ of this calculated intersection point from the sensor. Comparing these two distances we get (Fig. 6.13)

$$t = \frac{r_{i+j}}{r_s} = \frac{r_{i+j} \sin(\theta_{i+j} - \theta_i)}{\delta x_{i,i+j} \sin(\theta_i)}, \quad (6.15)$$

where $t = 1$ when looking at a corner and $t = 0$ when looking at a plane, i.e. the two parallel lines intersect at infinity. Determining which value the measured $t$ is closest to the test distinguishes between edges/corners and
planes in the presence of noise.

**Sequential probability ratio tests**

The tests described above can both be implemented to make use of the minimum amount of information only: i.e. 1 measurement for the edge-plane/corner test and 2 measurements from different viewpoints for the plane-edge/corner test. However, to get the misclassification errors sufficiently low, more measurements are generally needed [Peremans and Campenhout 92]. On the other hand, one usually strives to know the reflector type after collecting as few measurements as possible. For these reasons, a sequential probability ratio test [Wald 73] is derived in [Peremans and Campenhout 93]. Such a test can be shown to be an optimal compromise in the sense that it minimises the number of measurements necessary to achieve specified error probabilities for a “two class” test. This approach guarantees that we will not attempt to distinguish between two reflector classes until sufficient evidence is gathered so the classification can be done reliably. In addition, it is shown in [Peremans and Campenhout 92] that, assuming the measurements to be independently distributed gaussian variables, the sequential probability ratio test can be computed recursively, resulting in a very efficient test computationally.

### 6.4.3.4 Localisation of Curved Reflectors

Recall from above that the position of a reflector is given by the location of its intersection with the measurement plane. For curved reflectors we will assume this intersection to be a plane curve [O’Neill 66]. This type of reflector is a generalisation of the piecewise linear reflectors referred to above, i.e. edges, planes and corners. The shape of such a plane curve in the local neighbourhood of a point \( p \) is determined by its curvature \( \kappa \) in \( p \) which can be geometrically interpreted as being the inverse of the radius of the osculating circle, i.e. the best circular approximation of the curve at \( p \), as shown in Fig. 6.14.

To determine the position of the curved reflector that gave rise to the measured triple one needs to find the reflector boundary that is tangent to the three constraint curves defined by the three measured distances \( m_1, m_2, m_3 \). Hence, both for a circular reflector and a curved reflector, locally approximated by its osculating circle, the measurement situation is as shown in Fig. 6.15(a). The locus of points for which \( m_1 \) is constant is
a circle as transmitter and receiver have the same position. The loci of points for which the distances $m_2$ or $m_3$ are constant are ellipses. The foci of these ellipses coincide with the positions of the transmitter and the corresponding peripheral receiver. Introducing one further simplification, i.e. approximating the ellipses by circles with their centres halfway between the two foci (Fig. 6.15(b)) leads to the problem of finding a circle tangential
to three given circles. The radii of these circles can be expressed in terms of the measured distances [Peremans et al. 93] as

\[ r_1 = m_1/2, \quad r_2 = \left(\sqrt{m_2^2 - d^2}\right)/2, \quad r_3 = \left(\sqrt{m_3^2 - d^2}\right)/2. \]

From Fig. 6.15(b) we get

\[ r_c = \frac{2r_1^2 - r_2^2 - r_3^2 + d^2/2}{2(r_2 + r_3 - 2r_1)} \]  \hspace{1cm} (6.16)

and

\[ r = r_1, \quad \theta = \arcsin\left(\frac{r_3^2 - r_2^2 + 2r_c(r_3 - r_2)}{2d(r_1 + r_c)}\right). \]  \hspace{1cm} (6.17)

For a circular reflector these equations define both the reflector’s position \((r, \theta)\) as well as its radius \(r_c\). For a curved reflector \((r, \theta)\) and \(r_c\) denote the position and radius of an osculating circle as shown in Fig. 6.14.

Using the curvature information one can also discriminate between different types of reflectors. Such a curvature-based scheme can be seen to include as a special case the previous classification scheme by noting that edges can be mapped on curved reflectors with \(r_c = 0\) and planes onto curved reflectors with \(r_c = \infty\). Since corners and planes give rise to identical measured distances, curvature information will not distinguish between them either. As before, multiple measurements from different viewpoints will be required. This scheme can be shown [Peremans et al. 93] to generalise, within the context of geometrical wave propagation, the notion of a discrete set of different reflector types to a continuum of reflector types. However, it should be noted that the results obtained so far are only valid in noiseless conditions. The presence of noise will place a limit on the accuracy of the results. In [Peremans et al. 93] it is shown that the radius of curvature in particular is very sensitive to noise. This effectively limits the practical use of the radius of curvature as a decision parameter to distinguish between only two or three different types of reflectors.

### 6.4.4 One Transmitter, Two Receivers: 3 Dimensional World Model

For mobile robots operating in man-made environments the 2D world assumption is usually sufficient. However, the use of ultrasonic sensors fitted to the gripper, i.e. sensor-in-hand systems, of a robot-arm in order to
steer the arm adaptively requires a full 3D world model. A number of researchers have proposed ultrasonic sensor systems for localising objects in 3D, [Brown 85; Kuc 93; Hong and Kleeman 95; Akbarally and Kleeman 95; Delepaut 86; Barshan and Sekmen 99]. All of these systems are straightforward extensions of the ones described above, in particular, they all make use of differences in arrival times to calculate the position of the reflector in 3D. Consequently, they all tend to be rather big and use at least three receivers. On the contrary, the sensor system [Peremans et al 98] described below uses inter-aural differences in spectral information, allowing a very small head size. Furthermore, only two, dynamically re-orientated, receivers are needed to perform 3D target localisation. This sensor is referred to as a biomimetic sensor [Peremans et al 00] because both the morphology of the sonar head as well as the measurement strategies it employs are modelled on aspects of the bat’s echolocation system [Walker et al 98; Walker et al 98].

6.4.4.1 Sensor Setup

The biomimetic sensor, as shown in Fig. 6.16 (left), consists of a 6 DOF sonar head that allows panning and tilting of the neck, and independent panning and tilting of each of the two ears. The transmitter is mounted at the centre of the head, moving along with the pan and tilt movements of the neck. The transmit pulse is a broadband call consisting of a frequency swept, from 30kHz to 100 kHz, sinusoidal signal. The echo signals received are processed with the cochlear model described in section 6.3.2. Next, the
sensor system extracts for each frequency channel of the cochlear model the maximum amplitude. These sets of amplitudes, one set for each receiver, are the measurement result returned by the sensor.

6.4.4.2  Localisation of a Point-Like Reflector in 3D

In the case of a small head size it is generally believed that the use of inter-aural intensity differences (IIDs) instead of inter-aural time differences (ITDs) results in the most reliable position estimate. In a binaural system the IID value at a fixed frequency \( f \) corresponding to a particular echo can be determined from the directionally dependent sensitivity of the receivers only. Other factors influencing intensity like transmitter directivity, spreading losses, absorption in air, and reflection losses can be assumed to be identical for both receivers. Modelling this directivity by that of a circular piston, see equation 6.2, we get for the left and the right receiver respectively

\[
D_{\text{piston}}(\theta_i) = 20 \log\left(\frac{|J_1(ka \sin(\theta_i))|}{|ka \sin(\theta_i)|}\right) \text{ with } i \in [\text{left}, \text{right}]
\]

where \( \theta_{\text{left}} \) and \( \theta_{\text{right}} \) are the angles between the maximal sensitivity axes of the receivers and the respective reflecting object's lines-of-sight. Figure 6.17 clarifies the definition of these two angles; it also shows their relationship.
with the azimuth and elevation angles used below to describe the location of a reflecting object (fixed distance \( r = 0.5 \text{m} \)) with respect to the sonar head. Finally, the IID at a particular frequency \( f \) can then be written as

\[
\text{IID}(\text{azimuth, elevation}) = D_{\text{piston}}(\theta_{\text{left}}) - D_{\text{piston}}(\theta_{\text{right}}).
\]

A number of systems have been proposed [Kay 80; Kuc 94] that use a measured IID value to derive the azimuth angle for a target in the horizontal plane, i.e. elevation=0°. Fig. 6.18 shows how in that case the IID values relate to azimuth. Note that once the azimuth angle grows larger (smaller) than the angle for which the IID reaches its maximum (minimum), ambiguity is introduced. To extend this approach beyond the horizontal plane we study the theoretical IID maps for the sonar head when both target azimuth and elevation are allowed to vary. Figs. 6.19(a),(b) show the constant IID lines, i.e. iso-IID contours, of two such maps for two different frequencies. From this figure it is clear that, although each iso-frequency IID map is ambiguous (i.e. contains iso-IID contours that run through numerous bearings), the ambiguous regions in different frequency maps corresponding with particular IID values usually do not coincide. Thus, the combination of a set of IID maps - each corresponding to a different echolocation call
Multi-Aural Sensing

Fig. 6.19 Contour lines, every 5dB, of the IIID maps for $\Phi_R = \Phi_L = 0^\circ$ and (top left) 45 kHz, (top right) 65 kHz; 50 kHz and (bottom left) $\Phi_R = +5^\circ, \Phi_L = -5^\circ$, (bottom right) $\Phi_R = -5^\circ, \Phi_L = +5^\circ$

frequency - can be used to reduce target positional ambiguity.

Note that this strategy will not eliminate the ambiguity along the vertical midline as a 0 dB iso-IIID contour runs along this midline in all maps. To overcome this remaining ambiguity a two stage measurement procedure is introduced. First, both receivers are given an elevation offset, $+5^\circ$ and $-5^\circ$ respectively, and a measurement is performed. Next, the ears are flapped, i.e. the elevation offsets are changed to $-5^\circ$ and $+5^\circ$ respectively and a new measurement is performed. It is clear from Figs. 6.19(c),(d) that flapping the ears ensures minimal overlap between the ambiguity regions for the two ear configurations. This measurement procedure makes this sensor the first one to demonstrate that active sensing, i.e. actively adapting the sensor parameters, can bring additional benefits to sonar sensing.

To combine multiple IIID maps a number of evidence accumulation schemes can be used. Below, we show the results for a probabilistic evidence accumulation scheme based on probability maps [Peremans et al
A probability map is defined as a two-dimensional array, where the value of each cell represents the probability, denoted by \( P(cell_i | A_m) \), that a target at the corresponding position \((\text{azimuth}, \text{elevation})_{cell_i}\) caused a particular measured IID value \(A_m\). Employing Bayes rule, the posterior probability of a particular cell after combining \(n\) statistically independent measured IID values is given by:

\[
P(cell_i | A_m^n, A_m^{n-1}, \ldots, A_m^1) = \frac{P(A_m^n | cell_i) P(cell_i | A_m^{n-1}, \ldots, A_m^1)}{P(A_m^n | A_m^{n-1}, \ldots, A_m^1)}.
\]

The term in the denominator is a normalisation term, calculated by normalising the posterior probability map. The first term in the numerator represents the measurement model, i.e. additive, zero mean, Gaussian noise. The second term in the numerator (i.e., the prior probability map) represents the evidence accumulated through the previous measurements. At start up, this prior probability map is set equal to a uniform distribution.

Fig. 6.20 shows the posterior probability maps resulting from the combination of the IID maps, one for each frequency channel, for each ear configuration as well as the subsequent combination of those two maps. The true position of the target, i.e. a small sphere (\(r=1\)cm) dangling from a wire in front of the sonar head, is given by \(r = 50\)cm, \(\text{azimuth} = 2^\circ\) and \(\text{elevation} = -6^\circ\). Note that the maximum probability point from the final probability map, i.e. the white cell, stands out quite clearly from the background. From the final probability map we derive a single object position, i.e. \((\text{azimuth}, \text{elevation})\), by collapsing it into a single position estimate. For this experiment the maximum probability position estimate...
is azimuth = 2° and elevation = −6° and the mean position estimate is azimuth = 1.94° and elevation = −6.07°.

6.5 Summary

This chapter has developed the concept of biologically inspired sonar systems. Through developing the concept of echo formation as a series of transfer functions, it has isolated the effect of the reflector and developed methods by which sonar sensors, from a single transducer to small arrays, can discriminate reliably between different geometric primitives. It closes with a discussion of a sensor head which mimics the natural world, with a single transmitter and two flapping “ears” and shows that inter-aural intensity rather than range differences may be a more faithful model of binaural processing in such a head.
This page is intentionally left blank
Chapter 7

Map Building from Range Data Using Mathematical Morphology

Billur Barshan and Deniz Baškent

7.1 Introduction

This chapter presents a novel method for surface profile extraction using multiple sensors. The approach presented is based on morphological processing techniques to fuse the range data from multiple sensor returns in a manner that directly reveals the surface profile. The method has the intrinsic ability of suppressing spurious readings due to noise, crosstalk, and higher-order reflections, as well as processing multiple reflections informatively. The approach taken is extremely flexible and robust, in addition to being simple and straightforward. It can deal with arbitrary numbers and configurations of sensors as well as synthetic arrays. The algorithm is verified both by simulations and experiments in the laboratory by processing real sonar data obtained from a mobile robot. The results are compared to those obtained from a more accurate structured-light system, which is however more complex and expensive.

Perception of its surroundings is a distinguishing feature of an intelligent mobile robot. An inexpensive, yet efficient and reliable approach to perception is to employ multiple simple sensors coupled with appropriate data processing.

Sonar sensors are robust and inexpensive devices, capable of providing accurate range data, and have been widely used in robotics applications. However, as we have seen already in this book, the angular resolution of sonar sensors is low, resulting in an uncertainty in the location of the object encountered. Furthermore, data from these sensors are often considered difficult to interpret due to specular reflections from most objects, especially when multiple or higher-order reflections are involved.

In robotics, most of the research on sonar has focused on surfaces with fixed or piecewise-constant curvature, mostly composed of target primi-
tives such as planes, corners, edges, and cylinders [Ayrulu and Barshan 98; Barshan and Kuc 90; Hong and Kleeman 97; Kleeman and Kuc 95; Kuc and Siegel 87; Leonard and Durrant-Whyte 91; Peremans 93]. In an earlier study which considers surfaces with varying curvature, a purely analytical approach is taken using ultrasonic echo trajectories and differential geometry to extract the surface features [Brown 86].

Sonar sensors have also been extensively used for map building and obstacle avoidance in robotics. The different geometric approaches in map building basically fall into two primary categories: feature-based and grid-based. Feature-based approaches are based on extracting the geometry of the environment from sensor data as the first step in data interpretation (e.g. edge detection, straight-line or curve fitting to obstacle boundaries) [Cox 91; Crowley 85; Crowley 89; Drumheller 97; Grimson and Lorano-Perez 84]. Many researchers have reported the extraction of line segments from sonar data as being difficult and brittle and have proposed alternative feature-based approaches [Kuc and Siegel 87; Leonard and Durrant-Whyte 91; Leonard et al 92]. Approaches based on physical model-based reasoning, including classification of environmental features into target primitives discussed earlier [Ayrulu and Barshan 98; Barshan and Kuc 90; Bozma 92; Hong and Kleeman 97; Kleeman and Kuc 95] can also be considered feature-based methods.

An alternative representation is the certainty or occupancy grid (see chapter 4 section 4.1.1.1). Although grid-based methods have their limitations in terms of memory and resolution, they are advantageous in that they do not commit to making difficult geometric decisions early in the data interpretation process. On the other hand, since different target types are not explicitly represented, it is not as easy to predict what sonar data will be observed from a given position and to give account of individual sonar readings. Typically, when sufficient sensor data have been collected in the grid cells, the data are matched or correlated with a global model. This process can be computationally intensive and time consuming depending on the cell resolution of the grid.

In [Wallner and Dillman 95], a hybrid method is presented for updating the local model of the perceivable environment of a mobile robot. Unlike earlier grid-based approaches, local grids can be used in dynamic environments. The method combines the advantages of feature- and grid-based environment modeling techniques. More detail on the different approaches
to map building can be found in [Borenstein et al. 96].

The approaches described above are often limited to elementary target types or simple sensor configurations. On the other hand, the method presented in this chapter is aimed at the determination of arbitrary surface profiles which are typically encountered in mines, rough terrain, or underwater. The approach is completely novel in that morphological processing techniques are applied to sonar data to reconstruct the profile of an arbitrarily curved surface. It can also be looked at as a novel method for solving a class of nonlinear reconstruction (inverse) problems. It is important to underline that morphological processing is employed here to process the sonar map of the surface being constructed in the robot's memory. The method has sufficient generality to find application in other ranging systems such as radar, optical sensing and metrology, remote sensing, and geophysical exploration.

From a map-building perspective, this method can also be considered as a hybrid of feature- and grid-based methods: Initially, the environment is discretised into rectangular grids. After accumulation of sufficient amount of data, a curve-fitting procedure is employed to extract the geometry of the surface under investigation. The present approach is advantageous over probability-based grid methods since it allows the use of a much finer physical grid. This is because the approach does not rely on accumulation of multiple measurements in each cell. From a different perspective, although it is possible to interpret this method as a spatial voting scheme where cells are locally supported by their neighbours, we find it more appropriate to look at it from a nonlinear signal reconstruction perspective where morphological processing is used to extract reinforced features in the arc map.

The method is extremely flexible in that it can equally easily handle arbitrary sensor configurations and orientations as well as synthetic arrays obtained by moving a small number of sensors. As already mentioned above, a commonly noted disadvantage of sonar sensors is the difficulty associated with handling spurious readings, crosstalk, higher-order and multiple reflections. The method proposed is capable of effectively suppressing spurious readings, crosstalk, and most higher-order reflections. Furthermore, it has the intrinsic ability to make use of echo returns beyond the first one (i.e. multiple reflections) so that echoes returning from surface features further away than the nearest can also be processed informatively.
7.2 Basics of Sonar Sensing

The ultrasonic sensors used in this work measure time-of-flight (TOF), which is the round-trip travel time of the pulse from the sonar to the object and back to the sonar. Using the speed of ultrasonic waves in air \( (c = 343.3 \text{ m/s at room temperature}) \), the range \( r \) can be easily calculated from \( r = \frac{ct_o}{2} \) where \( t_o \) denotes the TOF. Many ultrasonic transducers operate in this pulse-echo mode [Hauptmann 93]. The transducers can function both as receiver and transmitter.

The objects are assumed to reflect the ultrasonic waves specularly. This is a reasonable assumption, since most systems operate below a resonant frequency of 200 kHz so that the propagating waves have wavelengths well above several millimeters. Details on the objects which are smaller than the wavelength cannot be resolved [Brown 86]. The sonars used in our experimental setup are Polaroid 6500 series transducers [Polaroid 97] operating at a resonance frequency \( f_o = 49.4 \text{ kHz} \), which corresponds to a wavelength of \( \lambda = \frac{c}{f_o} = 6.9 \text{ mm at room temperature} \).

The major limitation of sonar sensors comes from their large beamwidth, as we have seen in previous chapters. Although these devices return accurate range data, they cannot provide direct information on the angular position of the object from which the reflection was obtained. Thus, all that is known is that the reflection point lies on a circular arc whose radius is determined by \( r = \frac{ct_o}{2} \) as illustrated in Fig. 7.1(a). More generally, when one sensor transmits and another receives, it is known that the reflection point lies on the arc of an ellipse whose focal points are the transmitting and receiving transducers (Fig. 7.1(b)). Note that the reflecting surface is tangent to these arcs at the actual point(s) of reflection. The angular extent of these arcs is determined by the sensitivity regions of the transducers.

If multiple echoes are detected for the same transmitting/receiving pair, circular or elliptical arcs are constructed to correspond to each echo. (However, not all systems commonly in use are able to detect echoes beyond the first one.)

Most commonly, the large beamwidth of the transducer is accepted as a device limitation which determines the angular resolving power of the system. In this naive approach, a range reading of \( r \) from a transmitting/receiving transducer is taken to imply that an object lies along the line-of-sight of the transducer at the measured range. Consequently, the angular resolution of the surface profile measurement is limited by the rather
large beamwidth, which is a major disadvantage. Our approach, as will be seen, turns this disadvantage into an advantage. Instead of restricting oneself to an angular resolution equal to the beamwidth by representing the reflection point as a coarse sample along the line-of-sight, circular or elliptical arcs representing the uncertainty of the object location are drawn. By combining the information inherent in a large number of such arcs, angular resolution far exceeding the beamwidth of the transducer is obtained.

7.3 Processing of the Sonar Data

Structured configurations such as linear and circular arrays as well as irregular sensor configurations have been considered (The mobile robot used in the experiments has a circular array of sonar sensors.) Figure 7.2(a) shows a surface whose profile is to be determined. Part (b) of the same figure shows the circular and elliptical arcs obtained from circular arrays of sensors, which both rotate and translate to increase the number of arcs generated from the available number of sensors. Further sonar maps obtained using a circular configuration are presented later in Sec. 7.3.3.2.

Notice that although each arc represents considerable uncertainty as to the angular position of the reflection point, one can almost extract the actual curve shown in Fig. 7.2(a) by visually examining the arc map in...
Fig. 7.2 (a) The original surface. (b) The circular sensor array mounted on a mobile robot moves to 20 different locations and collects data by rotating around its center from 45° to 135° in 15° steps. The angles are measured with respect to the positive x axis in the counterclockwise direction. The circular array has been shown at the 20 locations at its 90° position. (c) Result of $n = 3$ thinning: $e = 0.052$, $f_c = 0.341$, $t_{CPU} = 1.11$ s. (d) The original surface (dashed line) and the polynomial fit of order $m = 9$ (solid line), with $E_1 = 3.18$ cm and $E_2 = 0.036$. In this case both coincide.

Fig. 7.2(b). Each arc drawn is expected to be tangential to the surface at least at one point. At these actual reflection point(s), several arcs will intersect with small angles at nearby points on the surface. The many small segments of the arcs superimposed in this manner create the darker features in Fig. 7.2(b), which tend to cover and reveal the actual surface. The remaining parts of the arcs, not actually corresponding to any reflections and simply representing the angular uncertainty of the sensors, remain more sparse and isolated.
Morphological processing will be employed to achieve what is natural for the human visual perception system: the extraction of Fig. 7.2(a) from 7.2(b). Morphological erosion and dilation operators will be used to weed out the sparse arc segments, leaving us with the mutually reinforcing segments which will reveal the original surface.

7.3.1 Morphological Processing

The main application of morphological operations is in modifying regions and shapes of images [Low 96]. Therefore, they are widely used in image processing to accomplish tasks such as edge detection, enhancement, smoothing, and noise removal [Dougherty 92; Myler and Weeks 93]. In this study, morphological processing is used to eliminate the sparse and isolated spikes and segments in the sonar arc map, directly revealing the surface profile.

Erosion and dilation are the two fundamental morphological operations used to thin or fatten an image respectively. These operations are defined according to a structuring element or template. The algorithm for erosion is as follows: The template is shifted over the pixels of the sonar map which take the value 1 one at a time and the template’s pixels are compared with those pixels which overlap with the template [Pitas 93]. If they are all identical, the central pixel with value 1 will preserve its value; otherwise it is set to zero.

The dilation algorithm is very similar to that for erosion, but it is used to enlarge the image according to the template. This time, all eight neighbours of those image pixels which originally equal 1 are set equal to 1.

In this study, the structuring element for dilation and erosion is chosen to be a 3 x 3 square template. Since the template is symmetric, the image will be fattened (dilation) or thinned (erosion) in all directions by one pixel.

The direct use of erosion may eliminate too many points and result in the loss of information characterising the surface. For such cases, the compound operations of opening and closing are considered. Opening consists of erosion followed by dilation and vice versa for closing. Opening helps reduce small extrusions, whereas closing enables one to fill the small holes in the image [Myler and Weeks 93]. Closing is applied prior to thinning, described below, in cases where the points are not closely connected to each other so that the direct use of thinning may result in the loss of too many points. Filling the gaps using closing first may prevent this from happening.
Thinning is a generalisation of erosion with a parameter \( n \) varying in the range \( 1 < n < 8 \). In this case, it is sufficient for any \( n \) neighbours of the central image pixel to equal 1 in order for that pixel to preserve its value of 1. The flexibility that comes with this parameter enables one to make more efficient use of the information contained in the arc map.

In pruning, which is a special case of thinning, at least one \( (n = 1) \) of the neighbouring pixels must have the value 1 in order for the central pixel to remain equal to 1 after the operation. This operation is used to eliminate isolated points [Dougherty 92]. Thus, pruning and erosion are the two extremes of thinning with \( n = 1 \) and \( n = 8 \) respectively.

Since there are many alternatives for morphological processing of sonar data, an error measure is introduced as a success criterion:

\[
e = \frac{\sqrt{\frac{1}{N_k} \sum_{i=1}^{N_k} (m_i - y_i)^2}}{\sigma_y}
\]  

(7.1)

Here, \( i \) is the discrete index along the \( x \) direction and \( y_i \) is the discretised function representing the actual surface profile with variance \( \sigma_y^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \frac{1}{N} \sum_i y_i)^2 \). \( N \) is the total number of columns whereas \( N_k \) represents those columns left with at least one point as a result of some morphological operation. \( m_i \) is the vertical position of the median (center-most) point along the \( i \)th column of the map matrix (e.g. Fig. 7.2(c)). If there are no points in a particular column, that column is excluded from the summation. If the number of columns thus excluded is large; that is, if the morphological operations have eliminated too many points, the remaining points will not be sufficient to extract the contour reliably, even if \( e \) is small. We will denote by \( f_c = N_k/N \) the fraction of columns left with at least one point at the end of a morphological operation. This factor must also be taken into account when deciding on which method provides a better result.

Additionally, the CPU times of the algorithms (\( t_{CPU} \)) are measured. These represent the total time the computer takes to realise the morphological operations starting with the raw TOF data. Morphological operations are implemented in the C programming language and the programs are run on a 200 MHz Pentium Pro PC.

The result of applying \( n = 3 \) thinning to the sonar arc map shown in Fig. 7.2(b) is presented in part (c) of the same figure and the results of various morphological operators applied to the same map are summarised.
### Table 7.1: Results of various morphological operations and curve fitting of order $m = 9$.

<table>
<thead>
<tr>
<th>morphological operation</th>
<th>$e$</th>
<th>$f_c$</th>
<th>$E_1$ (cm)</th>
<th>$E_2$</th>
<th>$t_{CPU}$ (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>thinning ($n = 1$: pruning)</td>
<td>0.168</td>
<td>0.924</td>
<td>7.83</td>
<td>0.089</td>
<td>1.15</td>
</tr>
<tr>
<td>thinning ($n = 2$)</td>
<td>0.074</td>
<td>0.637</td>
<td>4.42</td>
<td>0.050</td>
<td>1.12</td>
</tr>
<tr>
<td>thinning ($n = 3$)</td>
<td>0.052</td>
<td>0.341</td>
<td>3.18</td>
<td>0.036</td>
<td>1.11</td>
</tr>
<tr>
<td>thinning ($n = 4$)</td>
<td>0.045</td>
<td>0.160</td>
<td>5.37</td>
<td>0.061</td>
<td>1.10</td>
</tr>
<tr>
<td>thinning ($n = 5$)</td>
<td>0.012</td>
<td>0.063</td>
<td>16.21</td>
<td>0.183</td>
<td>1.09</td>
</tr>
<tr>
<td>thinning ($n = 6$)</td>
<td>0.014</td>
<td>0.021</td>
<td>464.30</td>
<td>5.246</td>
<td>1.10</td>
</tr>
<tr>
<td>thinning ($n = 7$)</td>
<td>0.007</td>
<td>0.003</td>
<td>1915.45</td>
<td>21.641</td>
<td>1.10</td>
</tr>
<tr>
<td>thinning ($n = 8$: erosion)</td>
<td>—</td>
<td>0.000</td>
<td>—</td>
<td>—</td>
<td>1.10</td>
</tr>
<tr>
<td>closing &amp; thinning ($n = 3$)</td>
<td>0.121</td>
<td>0.760</td>
<td>6.72</td>
<td>0.076</td>
<td>7.10</td>
</tr>
<tr>
<td>closing &amp; thinning ($n = 4$)</td>
<td>0.122</td>
<td>0.641</td>
<td>6.88</td>
<td>0.078</td>
<td>7.09</td>
</tr>
</tbody>
</table>

Since $f_c = 0$ for $n = 8$ thinning, reflecting the fact that all points are eliminated, the values of $e$, $E_1$, and $E_2$ are undefined for this case.

in Table 7.1. Error measures $E_1$ and $E_2$, given in the same table, will be defined and discussed later in Sec. 7.3.2. Since simple erosion results in very small values of $f_c$, we have considered thinning with parameter $n$. The error $e$ tends to decrease with increasing $n$. However, larger values of $n$ tend to result in smaller values of $f_c$ so that a compromise is necessary. For the time being, we note that the thinning parameter $n$ allows one to trade off between $e$ and $f_c$.

#### 7.3.2 Curve Fitting

As a last step, curve fitting is applied in order to achieve a compact representation of the surface profile in the robot's memory. Since our aim is to fit the best curve to the points, not necessarily passing through all of them, least-squares optimisation (LSO) is preferred to interpolation. LSO finds the coefficients of the best-fitting polynomial $p(x)$ of order $m$ (which is predetermined) by minimising

$$E_p^2 = \sum_{i=1}^{N} \sum_{j=1}^{M_i} [p(x_i) - f_{ij}]^2$$

(7.2)

where $E_p^2$ is the sum of the squared deviations of the polynomial values $p(x_i)$ from the data points $f_{ij}$. $x_i$ is the horizontal coordinate corresponding to the $i$th column of the map matrix and $f_{ij}$ is the vertical coordinate.
of the $j$th point along the $i$th column. The index $j$ runs through the $M_i$ points along column $i$, and $N$ is the number of columns. If $M_i = 0$ for a certain column, the inner summation is not evaluated and taken as zero for that column. The coefficients of the polynomial are determined by solving the linear equations obtained by setting the partial derivatives equal to zero [Lancaster 86]. Once an acceptable polynomial approximation is found, the surface can then be represented compactly by storing only the coefficients of the polynomial. Although polynomial fitting has been found to be satisfactory in all of the cases considered, other curve representation approaches such as the use of splines might be considered as alternatives to polynomial fitting.

To assess the overall performance of the method, two final error measures are introduced, both comparing the final polynomial fit with the actual surface:

$$E_1 = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [p(x_i) - y_i]^2}$$

$$E_2 = \frac{E_1}{\sigma_y}$$

The first is a root-mean-square absolute error measure, with dimensions of length, which should be interpreted either with respect to the pixel size, or with respect to the wavelength $\lambda$, which serves as a natural reference for the intrinsic resolving power of the system. The second is a dimensionless relative error measure which can be interpreted as the error relative to the variation of the actual surface.

The curve fitted to the surface map after the thinning shown in Fig. 7.2(c) is presented in part (d) of the same figure. Table 7.1 shows that increasing $n$ improves $e$ but worsens $f_c$ and that $E_1$ and $E_2$ achieve a minimum at some value of $n$ (which in this case happens to occur at $n = 3$ for both $E_1$ and $E_2$). In the simulations, where the actual surface is known, it is possible to choose the optimal value of $n$, minimising $E_1$ or $E_2$. In real practice, this is not possible so that one must use a value of $n$ judged appropriate for the class of surfaces under investigation.

In the simulations, higher-order reflections* are ignored since they are

*A higher-order reflection refers to an echo detected after bouncing off from object surfaces more than once.
difficult to model, although they almost always exist in practice: The key idea of the method is that a large number of data points coincide with the actual surface (at least at the tangent points of the arcs) and the data points off the actual curve are more sparse. Those spurious arcs caused by higher-order reflections and crosstalk also remain sparse and lack reinforcement most of the time. The thinning algorithms eliminate these spurious arcs together with the sparse arc segments resulting from the angular uncertainty of the sensors.

7.3.3 Simulation Results

The approach taken in this paper is very general in that arbitrary configurations and orientations of sensors can be handled equally effectively. We begin by considering the special cases of linear and circular arrays, which might be typical of deliberately designed array structures. Following these, we also consider arrays where the sensors are situated and oriented randomly. This not only exemplifies the general case where the array structure is irregular, but might also have several applications: For instance, it may correspond to the case where arcs are accumulated by any number of randomly moving and rotating sensors, perhaps mounted on mobile robots as in swarm robot applications. Another potential area of application of the random configuration is in array signal processing where the individual sensor positions of a regular array have been perturbed by the wind or waves.

7.3.3.1 Linear Arrays

First, linearly configured sensor arrays were considered. Given the beamwidth of the sensors (25°), it was observed that the number of arcs collected is relatively small. The situation gets even worse as the array is brought closer to the surface since interaction between an even smaller number of sensor pairs becomes possible. This is a consequence of the fact that the angular beamwidth subtends a smaller arc on the surface. It is interesting to note that whereas narrow beamwidths are esteemed for their high resolving power in conventional usage of such sonar sensors, here it would have been desirable to have sensors with larger beamwidths. This would enable a greater number of sensor pairs to produce a greater number of elliptical arcs, better and faster revealing the surface. However, since we
restrict ourselves to the parameters of the most widely available transducer (i.e. the Polaroid), rotating the sensors is considered instead to make up for their limited beamwidth, as shown in Figs. 7.3(a) and (b).

A further consideration is that in practice the number of sensors may be limited. One way to overcome this limitation is to move a smaller array much in the same spirit as synthetic aperture radar (SAR) techniques [Skolnik 81]. However, this is not completely equivalent to the full array since those elliptical arcs corresponding to pairs of sensors not contained within the actually existing array are missing. A further extension is the combination of such movement and rotation as was the case in our first example in Fig. 7.2.

We now return to Figs. 7.3(a) and (b) where the sensors are not translated but are only rotated. The results of morphological processing are presented in parts (c) and (d), and the curves fitted to them are shown in parts (e) and (f) of the same figure. In these and later simulations, the values of $n$ and $m$ chosen are those which give the smallest error $E_1$, unless otherwise indicated.

7.3.3.2 Circular Arrays

The circular configuration corresponds to the arrangement of sonar sensors on the Nomad 200 mobile robot in our laboratory. Only the five sensors facing the surface are activated since the others cannot see the surface. Again, the surface shown in Fig. 7.2(a) is considered. In addition to the array locations in Fig. 7.2(b), two further examples are presented in Figs. 7.4(a) and (b). The result of applying morphological processing and curve fitting to the sonar map in part (b) is presented in parts (c) and (d) of the same figure.

7.3.3.3 Arbitrarily-Distributed Sensors

In this section, the locations and the line-of-sight orientations of the sensors are generated randomly and do not conform to any special structure. In Fig. 7.5(a), a surface is shown whose profile is to be determined with such a configuration. Parts (b) and (c) show the sonar arcs obtained using different numbers of sensors. In Fig. 7.6(a), the surface features obtained after applying $n = 2$ thinning to the sonar arcs in Fig. 7.5(b) is shown. Similarly, applying $n = 5$ thinning to the arc map in Fig. 7.5(c), the result in Fig. 7.6(b) is obtained. The curves fitted to the surface features extracted
Fig. 7.3 Two linear arrays where the sensors are individually rotated from 40° to 140° in 10° steps: (a) An array of 11 sensors with 50 cm spacing. (b) An array of 21 sensors with 25 cm spacing. (c) Result of closing and $n = 5$ thinning applied to part (a): $e = 0.276$, $f_c = 0.772$, $t_{CPU} = 4.51$ s. (d) Result of $n = 3$ thinning applied to part (b): $e = 0.353$, $f_c = 0.864$, $t_{CPU} = 1.28$ s. (e) Polynomial fit of order $m = 8$ to part (c) (solid line), and the actual surface (dashed line). $E_1 = 3.03$ cm, $E_2 = 0.127$. (f) Polynomial fit of order $m = 7$ to part (d), and the actual surface. $E_1 = 7.85$ cm, $E_2 = 0.329$
Map Building from Range Data Using Mathematical Morphology

Fig. 7.4 (a) The robot is located at 70 locations and the front sonar is oriented at 135° with respect to the positive x axis. No rotation. (b) The robot rotates at 35 locations, from 45° to 135° in 15° steps. (c) Result of $n = 5$ thinning applied to part (b): $e = 0.050, f_c = 0.464$. (d) Polynomial fit of order $m = 11$ (solid line) and the actual surface (dashed line). $E_1 = 3.57$ cm, $E_2 = 0.040$. In this case both coincide

are presented in parts (c) and (d) of the same figure.

Although structured arrays such as linear or circular ones are often preferred in theoretical work for simplicity and ease of analysis, the method presented here is capable of equally easily handling arbitrary arrays. In fact, the large number of simulations we have undertaken indicate that arrays consisting of irregularly located and oriented sensors tend to yield better results. This seems to result from the fact that the many different vantage points and orientations of the sensors tend to complement each other better than in the case of a uniform array. Although the question of optimal complementary sensor placement is a subject for future research,
the results imply that it is preferable to work with irregular or randomised arrays rather than simple-structured arrays such as linear or circular ones.

7.4 Experimental Verification

In this section, the method is verified using the sensor systems on the Nomad 200 mobile robot in our laboratory.

7.4.1 System Description

The Nomad 200 mobile robot, shown in Fig. 7.7, has been used in the experiments. It is an integrated mobile robot including tactile, infrared, sonar and structured-light sensing systems, with dimensions 76.2 cm (height) and 45 cm (diameter). The mechanical system of the robot uses a non-holonomic, three-servo, three-wheel synchronous drive with zero gyro-radius. The control of the base translation, base rotation and turret rotation is performed by three separate motors. The robot can translate only in the forward and backward directions but not sideways without rotating first. Servo control is achieved by a MC68008/ASIC microprocessor system. The maximum translational and rotational speeds of the Nomad 200 are 60 cm/s and 60°/s respectively. Nomad 200 has an on-board computer for sensor and motor control and for host computer communication. The communication is managed with a radio link and a graphics interface (server). The robot can be run from a C-language program either through the server or
We next give a brief description of the two sensor modules used in the experiments:

The Sensus 200 Sonar Ranging System consists of 16 sonars which can yield range information from 15 cm to 10.7 m with ±1 % accuracy. The sensors are Polaroid 6500 series transducers [Polaroid 97] which determine the range by measuring the TOF. The transducer beamwidth is 25°. The carrier frequency of the emitted pulses is 49.4 kHz. The system can be interfaced with any type of microcontroller. The power requirements of the system are 100 mA at 5 V or 12 V [Nomad 97].
Fig. 7.7 Nomad 200 mobile robot. The ring of sonars can be seen close to the top rim of the turret, and the structured-light system is seen pointing rightwards on top.

The Sensus 500 Structured-Light System basically consists of a laser diode (as its light source) and a CCD array camera. The laser beam is passed through a cylindrical lens in order to obtain planar light. The operating range of the system is from 30.5 cm to 3.05 m. Within this range, the intersection of the plane of light with the objects in the environment can be detected by the camera. The range is determined by (laser line striping) triangulation, characterised by decreasing accuracy with increasing range [Everett 85]. The power requirement of the system is 2000 mA at 12 V [Nomad 97].

In the experiments, both sonar and structured-light data are collected from various surfaces constructed in our laboratory. The structured-light system is much more expensive and complex, requiring higher-power and sufficient ambient light for operation. Since it reveals a very accurate surface profile, the surface detected by this system is used as a reference in the experiments.

In order to prevent any crosstalk between consecutive pulses, the sonars should be fired at 62 ms intervals since the maximum range of operation of Polaroid transducers is 10.7 m. In the experiments, the sonars are fired at
40 ms intervals. This prevents much of the crosstalk, and in the few cases where erroneous readings are obtained due to crosstalk, these false readings are readily eliminated by the algorithm. This is another aspect in which the algorithm exhibits its robust character.

Fig. 7.8 (a) The surface profile revealed by the structured-light data, (b) the sonar data. The data in both parts are collected from the surface at every 2.5 cm by translating the mobile robot from \((-75, 0)\) to \((75, 0)\). (c) Result of erosion \((n = 8)\) followed by pruning \((n = 1)\). (d) Polynomial fit of order \(m = 10\). (e) Part (a) superimposed with part (d) to demonstrate the fit obtained. \(E_1 = 3.59\) cm, \(E_2 = 0.263\), and \(t_{CPU} = 0.27\) s

### 7.4.2 Experimental Results

The sonars on the Nomad 200 are in a circular configuration. Since the robot has a limited number of sensors which can detect the surface, by moving the robot and rotating its turret, the equivalent of a much larger number of sensors is created synthetically. First, the robot remained stationary and collected data by rotating its turret. However, there were many locations on the surface which could not be detected by the robot if only the turret rotated. On the contrary, pure translation alongside the surface generally provided satisfactory results.
First, several surfaces have been constructed in our laboratory with
different curvature and dimensions, using thin cardboard of height 1.05 m
and length 3.65 m. In these experiments, only the front five sensors have
been activated.

The structured-light data obtained from one of the cardboard surfaces
constructed are presented in Fig. 7.8(a). The sonar data presented in
Fig. 7.8(b) are obtained by translating the mobile robot horizontally over
a distance of 1.5 m along the line \( y = 0 \) and collecting data every 2.5 cm.
The turret is oriented such that both the structured-light and the front five
sonars are directed towards the surface and it does not rotate throughout
the translational movement.

If the sonar data in Fig. 7.8(b) are examined, it can be observed that
there are many points on the surface which have not been detected by the
sonars. Since the reflections are specular, segments of the surface which are
relatively perpendicular to the transducer line-of-sights are easily detected,
whereas those segments which are more or less aligned with the line-of-
sights could not be sensed.\(^\dagger\) This is less of a problem when the surface
is relatively smooth and its curvature is small. The extent to which this
becomes a problem depends on the radius of curvature, as will be evident
when we consider below a second example with less curvature.

As expected, the structured-light data provide a very accurate surface
profile. In the arc map obtained by sonar, there are some arcs which are
not tangent to the actual surface at any point (e.g. the isolated arcs in
the upper right and upper left parts of Fig. 7.8(b)). These correspond to
spurious data due to higher-order reflections, readings from other objects
in the environment, or totally erroneous readings. These points are readily
eliminated by the morphological processing (Fig. 7.8(c)). If the final curve
in Fig 7.8(d) is compared with the structured-light data (Fig. 7.8(e)), it can
be observed that a close fit to the original surface is obtained. The errors in
this case are \( E_1 = 3.59 \) cm, \( E_2 = 0.263 \), and the CPU time is \( t_{CPU} = 0.27 \) s.

Next, a surface with smaller maximum curvature (hence with larger min-
imum radius of curvature), shown in Fig. 7.9(a), is considered. The results
of morphological processing and curve fitting are shown in Fig. 7.9(c)-(e),

\(^\dagger\)Since specular reflections involve negligible scattering and mirror-like reflections, the
transmitted waveform can be received back at the transducer only if some of the rays
emerging from the transducer are perpendicular to the surface. For the case of separate
transmitter and receiver, the surface must be perpendicular to the normal determined
by the transmitting and receiving elements.
resulting in $E_1 = 1.41\, \text{cm}$, $E_2 = 0.156$ and $t_{CPU} = 0.39\, \text{s}$. It is indeed observed that $E_1$ and $E_2$ are reduced significantly with respect to the previous case, as also evidenced by the much better fit seen in Fig. 7.9(e). Several results obtained for this surface are summarised in Table 7.2. All polynomials are of degree $m = 10$. The minimum estimation error $E_1$ is not much larger than the wavelength $\lambda = 6.9\, \text{mm}$.

It is worth noting that the present approach is aimed at estimation of a curve, as opposed to detection of target primitives. When corners or edges are encountered in the environment, the method does not fail but the results slightly deteriorate since the method exploits neighbouring
Table 7.2  Experimental results for the surface given in Fig. 1.9(a) for different morphological operations.

<table>
<thead>
<tr>
<th>morphological operation</th>
<th>$E_1$ (cm)</th>
<th>$E_2$</th>
<th>$t_{CPU}$ (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>thinning ($n = 1$: pruning)</td>
<td>4.98</td>
<td>0.554</td>
<td>0.41</td>
</tr>
<tr>
<td>thinning ($n = 2$)</td>
<td>4.84</td>
<td>0.539</td>
<td>0.41</td>
</tr>
<tr>
<td>thinning ($n = 3$)</td>
<td>4.07</td>
<td>0.452</td>
<td>0.40</td>
</tr>
<tr>
<td>thinning ($n = 4$)</td>
<td>3.28</td>
<td>0.364</td>
<td>0.39</td>
</tr>
<tr>
<td>thinning ($n = 5$)</td>
<td>2.58</td>
<td>0.287</td>
<td>0.37</td>
</tr>
<tr>
<td>thinning ($n = 6$)</td>
<td>1.96</td>
<td>0.218</td>
<td>0.36</td>
</tr>
<tr>
<td>thinning ($n = 7$)</td>
<td>1.63</td>
<td>0.182</td>
<td>0.35</td>
</tr>
<tr>
<td>thinning ($n = 8$: erosion)</td>
<td>1.42</td>
<td>0.158</td>
<td>0.34</td>
</tr>
<tr>
<td>erosion &amp; pruning ($n = 1$)</td>
<td>1.41</td>
<td>0.156</td>
<td>0.39</td>
</tr>
<tr>
<td>erosion &amp; thinning ($n = 2$)</td>
<td>1.50</td>
<td>0.167</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Fig. 7.10  In all three parts, the result of curve fitting (solid line) is compared to the original surface profile as revealed by the structured-light data. Data are collected by rotating the turret of the robot from $-30^\circ$ to $30^\circ$ taking $1^\circ$ steps. (a) $90^\circ$ corner at 80 cm: polynomial fit of order $m = 6$, resulting in $E_1 = 1.90$ cm and $E_2 = 0.217$. (b) $60^\circ$ corner at 80 cm: polynomial fit of order $m = 10$, resulting in $E_1 = 1.96$ cm and $E_2 = 0.243$. (c) $90^\circ$ edge at 60 cm: polynomial fit of order $m = 8$ resulting in $E_1 = 3.83$ cm and $E_2 = 0.414$

relationships and local continuity (i.e. smoothness). Experimental results for $60^\circ$, $90^\circ$ corners and $90^\circ$ edge are presented in Fig. 7.10. These objects are made of smooth wood and are 1.20 m in height. The results indicate that the method works acceptably in such cases as well, the net effect being that the vertices of the sharp edges are rounded or smoothed (i.e. low-pass filtered) into curved edges. This corresponds to the spatial frequency resolving power of the system as determined by the chosen grid spacing.
Finally, we present experimental results obtained by using the front three sonars of the Nomad 200 robot, following the walls of the room in Fig. 7.11(a). The room comprises of smooth wooden (top and left) and plastered (right) walls, and a window shade with vertical flaps of 15 cm width (bottom). Some of the corners of the room were not perfect (e.g. where the shade and the right wall make a corner). A cylindrical object of radius 15 cm is present at a distance of 30 cm from the center of the right wall. In Fig. 7.11(b), the path of the robot and the resulting arc map are given. In Fig. 7.11(c), the result of morphological processing is shown. It is clear from this figure that despite the potential for many higher-order reflections, spurious arc segments have been fairly well eliminated, and we can expect a good polar polynomial fit (or line segment matching).

Even though the method was initially developed and demonstrated for specularly reflecting surfaces, subsequent tests with Lambertian surfaces of varying roughness have indicated that the method also works for rough surfaces, with errors slightly increasing with roughness [Barshan and Baskent 00].

Closing operations were not needed in processing the experimental data because the points were sufficiently dense. If this was not the case, one would first apply closing in order to add extra points to fill the gaps between the points of the original map.
7.4.3 Computational Cost of the Method

The average CPU times are in general of the order of several seconds, indicating that the method is viable for real-time applications. These represent the total time the computer takes to realise the morphological operations starting with the raw TOF data. (The morphological processing has been implemented in C language and runs on a 200 MHz Pentium Pro PC.) For comparison, the time it takes for an array of 16 sonars to collect all the TOF data is \(16 \times 40 \text{ ms} = 0.64 \text{ s}\) which is of the same order of magnitude as the processing time. It should be noted that the actual algorithmic processing time is a small fraction of the CPU time, as most of the CPU time is consumed by file operations, reads and writes to disk, matrix allocations etc. Thus, it seems possible that a dedicated system can determine the surface profile even faster, bringing the computation time below the data collection time.

Another important factor is memory usage. In the simulations, the objects are relatively large and a relatively large number of sensors are employed. This leads to memory usage ranging between 100–750 kB. In the experiments, the targets are smaller and relatively close so that the data files consume about 50–100 kB. Although memory usage depends on the number of sensors used, size of the object, and the grid size, these figures are representative of the memory requirements of the method.

7.5 Discussion and Conclusions

A novel method is described for determining arbitrary surface profiles by applying morphological processing techniques to sonar data. The method is both extremely flexible, versatile, and robust, as well as being simple and straightforward. It can deal with arbitrary numbers and configurations of sensors as well as synthetic arrays obtained by moving a relatively small number of sensors. Accuracy increases with the number of sensors used (actual or synthetic) and has been observed to be quite satisfactory. The method is robust in many aspects; it has been seen that it has the inherent ability to eliminate most of the undesired TOF readings arising from higher-order reflections as well as the ability to suppress crosstalk when the sensors are fired at shorter intervals than that nominally required to avoid crosstalk. In addition, the method can effectively eliminate spurious TOF measurements due to noise, and process multiple echoes informatively.
The processing time is small enough to make real-time applications feasible. For instance, the system can be used for continual real-time map building purposes on a robot navigating in an environment with vertical walls of arbitrary curvature. Two extensions immediately come to mind: First, it is possible for the robot to continually add to and update its collection of arcs and reprocess them as it moves, effectively resulting in a synthetic array with more sensors than the robot actually has. Second, the method can be readily generalised to three-dimensional environments with the arcs being replaced by spherical or elliptical caps and the morphological rules extended to three dimensions. The method was also found successful in determining the profile of surfaces completely surrounding the sensing system. In this case, it may be preferable to reformulate the method in polar or spherical coordinates.

Although the structured-light system has been used mainly as a reference in this study, the fact that its strengths and weaknesses are complementary to the sonar system suggests the possibility of fusing the output of the two systems. The structured-light system provides a very accurate surface profile, but introduces errors increasing with range, as a result of the triangulation technique it employs. On the other hand, sonars yield better range information over a wider range of operation, but are less adept at recognising the contour details due to their wide beamwidth. Despite the possibility of fusion, the method described in this paper may be preferable in many circumstances, since the structured-light system is much more expensive and complex compared to sonar sensors.

Although not fully reported here, a detailed quantitative study of the performance of different morphological operations and the dependence of the error on surface curvature, spatial frequency, distance, sensor beamwidth, and TOF noise can be found in [Barshan and Baskent 01]. In [Barshan and Baskent 00], processing the arc maps using morphological processing is compared to using spatial voting instead.

The essential idea of this chapter—the use of multiple range sensors combined with morphological processing for the extraction of the surface profile—can also be applied to other physical modalities of range finding of vastly different scales and in many different application areas. These may include radar, underwater sonar, optical sensing and metrology, remote sensing, ocean surface exploration, geophysical exploration, and acoustic microscopy. Some of these applications (e.g. geophysical exploration) may involve an inhomogeneous and/or anisotropic medium of propagation. It is
envisioned that the method could be generalised to this case by constructing broken or non-ellipsoidal arcs.

Acknowledgment

This work has been adapted from [Baskent and Barshan 99] (©1999 Sage Publications, Inc.). The authors are grateful for the support of TÜBİTAK under grant 197E051.
This page is intentionally left blank
Chapter 8

Millimetre Wave Radar for Robotics

Graham Brooker

8.1 Background

The term millimetre waves refers to that portion of the electromagnetic spectrum between 30 and 300 GHz, corresponding to wavelengths between 10 and 1mm. For robotic applications, most activity is concentrated around the atmospheric window at 94 GHz that has been used by the military for a number of years [Miley et al 79]. Since 77 GHz (on the periphery of the same window) was allocated for automotive radar, lower cost sensors at the lower frequency have begun to appear from Saabtech Electronics, Epsilon Lambda and Fujitsu [Luo et al 99; Eriksson and Broden 96] amongst others.

The resurgence of interest in millimetre waves primarily for military applications arises from the realisation that there are limitations with what can be accomplished with infrared and optical systems, in particular their disadvantages in dust and fog or for night-time viewing. Niche commercial applications, particularly for industrial electronic distance measurement (EDM) and collision warning have been proposed since the early 1970's [Groll and Detlefsen 97], but it has only been in the last decade or so that the advantages of millimetre wave for autonomous guidance have been realised [Lange and Detlefsen 91].

One of the primary guidance applications has been the development of active and radiometric landing technology [Bui and Morton 91; Koester and Vaillancourt 92]. This technology is discussed later in the chapter.
|
|---|---|---|
|Sensor Characteristic| Millimetre wave| Laser|
|Cost| Poor| Good|
|Tracking accuracy| Fair| Good|
|Classification/ identification| Fair| Good|
|Imaging| Fair| Good|
|Volume search| Fair| Poor|
|Adverse weather performance| Fair| Poor|
|Performance in smoke, dust etc.| Good| Poor|
|Dirty Antenna| Good| Poor|

Table 8.1 Sensor System Comparison

8.2 When to Use Millimetre Wave Radar

For applications where the performance of ultrasonic sensors is marginal either because the range is excessive or because the environment is too noisy or windy, scanning laser and millimetre wave radar sensors are the only viable alternatives. However, before a selection is made, a careful analysis should be undertaken to determine whether using millimetre wave offers sufficient advantages to outweigh its higher cost. In particular, if it cannot be justified through at least one of the last four characteristics listed in Table 8.1, it should not be used.

For robotic applications, particularly those involved with construction, mining or agricultural activities where dust is almost always present, or for long range outdoor activities that require an all-weather capability, millimetre wave is the frequency of choice.

The all-weather capability is best described with the aid of the curves [Bhartia and Bahl 84] in Figure 8.1 that show atmospheric attenuation as a function of frequency. When a comparison is made between the performance of a sensor at 94GHz (3mm) and a typical laser system (900nm) [Riegl], under clear conditions or in light drizzle neither suffer significant attenuation. However, as the rain rate increases, and the median drop size diameter increases into the optical region [Boguish 89] for both bands, attenuation becomes significant for both (about 10dB/km at 25mm/hr).

It is the performance in mist and fog where the water drop diameter is such that the radar is operating in the Rayleigh region and the laser in the Mie (or resonant) region, that the largest differences can be seen. Under
When to Use Millimetre Wave Radar

Fig. 8.1 Atmospheric Attenuation Characteristics for Wavelengths 3 cm to 0.3 μm. The figures are referenced to 20°C, 1 atmosphere pressure, 7.5 gm/mm³ water. The attenuation peaks are due to water H₂O (1), carbon dioxide CO₂ (2), oxygen O₂ (3), ozone O₃(4).

In these conditions the radar attenuation is still insignificant, but the laser suffers an attenuation of around 300dB/km, which makes it totally blind.

A similar analysis can be performed for environments that contain smoke, dust, or even large concentrations of particles of biological origin (such as pollen). It has been shown [Brooker et al 01; Bhartia and Bahl 84] that millimetre wave radar suffers attenuation of about seven orders of magnitude less than a laser under these conditions.
8.3 Millimetre Wave Radar Principles

Radar systems operate by radiating electromagnetic radiation, and receiving an echo return. The angular resolution is determined by the antenna beamwidth and the range resolution by the bandwidth of the transmitted signal. Resolution, $\delta R$, is defined as the minimum separation between two point reflectors of equal size that can be distinguished by the radar after processing.

In angles, this is a fraction of approximately 0.85 of the antenna 3dB beamwidth [(Skolnik 70], while in range it is proportional to the reciprocal of the transmitted bandwidth, $\Delta f$:

$$\delta R = \frac{c}{2\Delta f}$$  (8.1)

8.3.1 Range Resolution

Because the propagation of electromagnetic radiation is faster by a factor of $10^6$ than that of sound through the air, the electronics of a radar system must operate a million times faster than that of a sonar system to achieve the same range resolution. A real pulse with a duration of $\tau = 1\text{ns}$ is required to obtain a resolution of about 15cm. Because it is difficult to produce very high peak powers using solid state amplifiers and because a pulse time-bandwidth product $B\tau \approx 1$ is required for matched filtering of the received echo, the operational range of such systems is very short.

Alternative methods of obtaining high resolution have been developed in the last fifty years. They include pulse compression, continuous or interrupted frequency modulation and stepped frequency techniques. In essence all of these techniques widen the transmitter bandwidth and lengthen the effective pulse width to improve the resolution while offering good range performance.

In environments where multiple radar sensors are in use, the modulation type or operational frequency must be selected with care to ensure that mutual interference between sensors does not occur. Pseudo Noise (PN) is the modulation of choice in this case [Groll and Detlefsen 97]. However this technology is unlikely to be used for robotic applications where the numbers in one location are limited.
8.3.2 **Pulse Compression**

Pulse compression was described for sonar in section 4.3.2 of chapter 4. In sonar the transmitter modulation and the receiver compression were achieved in software; in radar they are normally achieved in hardware. In a pulse compression radar, a very brief pulse consisting of a range of frequencies passes through a dispersive delay line (SAW expander) in which its components are delayed in proportion to their frequency. In the process the pulse is stretched; for example a 1ns pulse may be lengthened by a factor of 1000 to a duration of 1μs before it is up-converted, amplified and transmitted. The echo returns from the target are down converted and amplified before being fed into a pulse compression network that retards the echo by amounts that vary inversely with frequency to reduce the signal to its original 1ns length. The compressed echo yields nearly all of the information that would have been available had the unaltered 1ns pulse been transmitted. A slight sacrifice in range resolution is incurred in reducing the range sidelobes from −13.2dB with no weighting to about −43dB with
The amount of signal-to-noise ratio (SNR) gain achieved can be described by the pulse time-bandwidth product $B\tau$. Even though using surface acoustic wave (SAW) technology to implement the pulse expansion and compression functions limits the maximum $B\tau$ product to about 1000 [Currie 87], it is the most common method in use today because it is both compact and robust.

Pulse compression technology is not suitable for short-range robotic applications because the minimum range is determined by the time taken to transmit the stretched pulse. In the above example this is equivalent to a range of 150m.

### 8.3.3 Stepped Frequency

When using a real or compressed pulse waveform to achieve high range resolution, it is a challenge to receive the acquired data at the required rate (or bandwidth). For 0.15m resolution, both the real pulse and pulse compression systems would have to sample and process data at a peak rate of at least 1GHz for a period of time equivalent to the range extent of the area of interest.

A technique that avoids the data acquisition problems associated with the real pulse mode is the pulse-to-pulse stepped frequency mode. A pulse width $\tau$ is selected to span the range of interest, for example a 100ns pulse will span a range of 15m. The frequency of each pulse is shifted by a small amount $\Delta F$ from that of the previous pulse, where $\Delta F$ is selected to be about $1/2\tau = 5$MHz to ensure that there is no phase ambiguity in the returned signals.

After each pulse is transmitted, the received echo at a particular range is coherently detected (to maintain the phase information), and the amplitude and phase information stored. For a transmit frequency of $F$, the phase of the received echo will be:

$$\Phi = \frac{4\pi FR}{c} \tag{8.2}$$

For a sequence of pulses equally spaced in frequency, there is a predictable pulse to pulse phase shift, $\delta \Phi$ of that is a function of the frequency difference $\Delta F$:
This pulse-to-pulse phase shift appears as a frequency shift, which is a function of the range to the target. If multiple targets appear in the same range bin, then each will produce a unique frequency that can be extracted from the time domain signal using the Fast Fourier transform (FFT) process.

The total unambiguous range after processing is $c/2\Delta F$, and the range resolution is $c/2F_{tot}$ where $F_{tot}$ is the total frequency excursion of the transmitted signal. For a sequence of $N$ pulses $F_{tot} = N\Delta F$. For an ultimate resolution of 15cm, $F_{tot} = 1GHz$, $\Delta F = 5MHz$ and $N = 200$ samples. As with the pulse compression case, a Hamming window will increase this resolution by a factor of 1.3 to 19.5cm.

The primary difficulty with using stepped frequency is to maintain the stability of the transmitter and local oscillator for the whole period that a measurement is being made. In a robotic environment, a further disadvantage is the time to perform a single high-resolution measurement. For example, if the maximum operational range was 300m, as each new pulse can only be transmitted after the previous pulse has returned, the pulse repetition interval (PRI) is at least 2$\mu$s and the total time required to perform a measurement is at least 400$\mu$s for $N = 200$. If all the targets spanning the range from 0m to 300m must be sampled into bins 15m wide, a total of 20 gates will be required and each of them will have to be processed by the FFT to produce range bins. This equates to a significant processing overhead. For example if a 256 point complex FFT is used, then the total processing time required would be $20 \times 256 \times \log_2(256) = 40960$ time units.

A further difficulty that must be considered is that if the target is mounted on a moving vehicle, then Doppler effects will shift the apparent range of the targets.

### 8.3.4 Frequency Modulated Continuous Wave

Most frequency modulated continuous wave (FMCW) radars operate by transmitting a linear frequency chirp $\Delta f$ of long duration $T_d$. At any instant in time, the echo signal received is shifted in frequency from the
transmitted signal by the product of the roundtrip time $T_p$ to the target and the rate of change of frequency $\delta f/\delta t$. If the received signal is mixed with a portion of the transmitted signal and filtered, the resulting output will be a constant beat frequency $f_b$.

Two factors limit the range resolution of FMCW systems. The first is a function of the chirp bandwidth $\Delta f$, and the second is the actual linearity that can be achieved for the transmitted chirp. An exceptionally high-resolution radar system may have a chirp bandwidth $\Delta f = 1000\text{MHz}$ and a sweep time of 2ms. The theoretical range resolution that can be achieved if only the chirp bandwidth considered is:

$$\delta R_{\text{chirp}} = \frac{c}{2\Delta f} = \frac{3 \times 10^8}{2 \times 10^9} = 0.15\text{m}$$ (8.4)

For a rate of change of frequency (chirp slope) $S = \delta f/\delta t$ at any point on the graph, the linearity, $Lin$, is defined as follows:
\[ L_{in} = \frac{S_{\text{max}} - S_{\text{min}}}{S_{\text{min}}} \] (8.5)

The best resolution that can be achieved in this case is the product of the linearity and the range to the target. For a target at a range of 300m and a linearity of 0.001, the resolution is:

\[ \delta R_{\text{lin}} = R L_{in} = 300 \times 0.001 = 0.3m \] (8.6)

The two results are generally combined in quadrature to determine the ultimate system range resolution:

\[ \delta R = \sqrt{\delta R_{\text{chirp}}^2 + \delta R_{\text{lin}}^2} = \sqrt{0.15^2 + 0.3^2} = 0.335m \] (8.7)

The actual beat frequency for a radar with the characteristics defined above, and a target range of 300m is just the product of the round trip time to the target and the chirp slope:

\[ f_b = T_p \frac{\delta f}{\delta t} \approx T_p \frac{\Delta f}{T_d} = 10^6 Hz \] (8.8)

The relationship between the measured beat frequency and the range is:

\[ \frac{R}{f_b} = \frac{T_d c}{2 \Delta f} = 0.0003m/Hz \] (8.9)

In general many targets will be present in the beam, so some form of spectrum analyser must be used to extract all of the frequencies present in the received signal. In modern systems, this function is usually implemented in software using the Fast Fourier Transform (FFT) process preceded by a weighting function to reduce range sidelobes.

If the signal is observed for a time \( T_d = 2\text{ms} \) then the width of the FFT frequency bin \( W = 1/T_d = 500Hz \) (15cm) and main lobe width, as shown in figure 1.3, is twice that. The 3dB bandwidth of the filter produced by the FFT process is \( 0.89 \) bins (13.3cm) for no windowing, increasing to \( 1.3 \) bins (19.5cm) for a Hamming window [Kirsten 79].
To produce a spectrum spanning the whole range from 0m to 300m with a bin size of 15cm requires 2000 bins. These can be produced by a complex 2048-point FFT or a real 4096-point FFT. Assuming that a 2048 point complex FFT is performed, the total processing time is $2048 \log_2(2048) = 22528$ time units, about half as long as the time taken for the stepped frequency processing.

These results show that at short range the range resolution is generally determined by the chirp bandwidth widened by the applied windowing function, while at longer ranges the chirp linearity becomes dominant.

### 8.3.5 Angular Resolution and Antennas

Angular resolution is determined by the antenna beamwidth that is in turn governed by the antenna aperture and the transmitted frequency (see section 3.3 in chapter 3). An empirical formula that relates the antenna diameter ($D$) for a typical parabolic dish to the 3dB beamwidth is:

$$\theta_{3dB} = \frac{70\lambda^o}{D} \quad (8.10)$$

Millimetre wave radar is used for robotic applications, where size and mass are constrained, because a narrower beamwidth can be obtained from a given aperture at high frequencies than at low frequencies. Even at 94GHz, however, a trade-off must be made between antenna size and acceptable beamwidth for most applications. Low sidelobe millimetre wave antennas with diameters of more than about 300mm are difficult to manufacture, fairly heavy and expensive, so for most general collision avoidance or guidance applications, apertures of 150 to 250mm are used. At 94GHz, the beamwidth of an antenna with a diameter of 150mm is 1.5°, and this reduces to only 0.9° if the diameter is increased to 250mm.

Most radar applications operate in the far-field of the antenna where the radiated power density is governed by the inverse-square law and the beamwidth in degrees remains constant. In the near field, the pattern and power density are range dependent. The critical range between the two is generally defined at a range of $D^2/\lambda$. As a first approximation, it can be assumed that the beam diameter equals the antenna aperture in the near field region and only starts to diverge in the far field.

In addition to beamwidth requirements, antenna gain ($G$) is important
for longer range operation. Gain is determined from the aperture area \(A\), the aperture efficiency \(\eta\) and the wavelength \(\lambda\) as follows *:

\[
G = \frac{4\pi\eta A}{\lambda^2}
\]  

(8.11)

Efficiencies from 0.5 to 0.7 are generally achieved except where ultra low sidelobes are required, in which case the efficiency is generally lower because of under-illumination of the lens or reflector.

A final factor that is important is the antenna impedance match. This determines how much of the transmitter power is reflected back into the radar and how much is radiated (or absorbed). Because FMCW radars often transmit and receive simultaneously through the same antenna, as shown in Figure 8.3, it is critically important to have a good match, otherwise reflected power can saturate the sensitive receiver circuitry. For transmitted powers of 10dBm (10mW) or lower, a return loss of at least 20dB is required for proper operation.

As mentioned earlier, millimetre wave systems are generally used in robotic applications where the environment is unpleasant. For this reason most antennas in use today are hermetically sealed horn-lens units similar to the one shown in on the right of Figure 8.4 rather than the more conventional Cassegrain design shown on the left. In addition to offering all weather operation, horn-lens antennas are the more robust of the two and because there is no aperture blockage, their sidelobe levels are generally also lower.

*Comparing this expression with equation 3.2 in chapter 3 we see that the receiving area of a parabolic dish is related to the physical area by the factor \(\eta\).
8.3.6 Scanning and Imaging

Specialist applications require other antenna structures of which the fan-beam antenna is probably the most common. As its name suggests, the fan-beam produces a beam that is narrow in azimuth and wide in elevation. This ensures that the radar continues to illuminate the target even if the vehicle on which it is mounted is rolling and pitching.

8.3.6.1 Mechanical Scanning

For most robotic applications, a real beam is scanned by physically rotating the antenna using a pan/tilt mechanism or by utilising the reflective properties of a finely polished rotating mirror. Such systems will be discussed in more detail in the next section.

8.3.6.2 Electronic Scanning

Because of reliability issues relating to mechanically scanned systems, a number of research institutes are investigating electronically scanned options [Foessel 00; Belov 01]. Electronic scanning using phased array technology at low frequencies is a mature technology and is described in a number of texts [Brookner 85; Skolnik 81]. Phased array antennas produce a narrow beam by transmitting from a large number of small (wide beamwidth) antennas simultaneously. To scan the beam, a progressive phase shift is introduced into each of the antennas. At millimetre wave frequencies, high losses and expense limit their application at this stage, though advances in micro electro-mechanical systems (MEMS) technology is showing promise as a technology that can produce low loss, low cost phase shifters [Barker and Rebeiz 98].
The forward motion of the radar can be used to scan the ground while simultaneously synthesising a narrower beamwidth than that available from the real aperture. Such a technique, known as synthetic aperture radar (SAR), is an expensive option both in terms of hardware and processing requirements, so it is not applied to most robotic applications. It can be utilised for the autonomous guidance of large unmanned aerial vehicles (UAV’s) as the images produced, such as the one shown in Figure 8.6, have extremely high resolution.

Synthetic aperture radar is seldom, if ever, used for ground based robotic applications because the expense cannot be justified in terms of the marginal improvements in cross range resolution over the resolution from a real aperture system operating at the same range.

8.3.6.3 Image Representation

Two-dimensional images of the surface of the earth made at low grazing angles such as those produced by airborne radar systems are generally displayed on a Cartesian or polar grid with each pixel encoded with the reflectivity of the ground at that point. The images produced are similar in some respects to aerial photographs taken from directly overhead with the sun illuminating the terrain from the radars position. The use of a Cartesian grid allows the integration of multiple scans onto the grid by averaging the amplitudes of successive returns [Brooker et al 01]. This technique reduces the effect of speckle (interference from multiple scatterers in a single range

Fig. 8.6 SAR Image of China Lake Airport [Sandia]
bin) and specular returns and works particularly well if the radar is moving and the relative observation angle changes during the integration period.

The following characteristics have been observed in images made using a millimetre wave radar operating at 94 GHz:

1. Smooth horizontal surfaces (water, roads and runways) are specular and reflect most of the power away so they appear dark.

2. The far edges of roads and runways appear brighter than the surrounding region because of the small corner made between the smooth surface and the vegetation growing along the boundary.

3. Grass and cultivated fields scatter fairly uniformly and show texture.

4. Isolated scrub and trees are good scatterers and generally appear brighter than grass and crops.

5. Large trees, walls, banks and buildings cast deep shadows.

6. Buildings, vehicles and other man-made objects are specular reflectors because their surfaces are flat and smooth. However, because even small corners can have significant radar cross-sections, it is difficult to predict the point of reflection.

7. The reflectivity of natural surfaces increases with increasing grazing angle so hills appear brighter as they become steeper.

For navigation or guidance purposes images like these can be correlated with aerial photographs or specific unambiguous features can be extracted from the image and associated with similar features extracted from an aerial photograph. The feature association technique is faster than correlation because only a few features are required, but it requires some skill to examine.
an aerial photograph and predict those features that are likely to be visible on a radar image.

For short-range navigation and obstacle avoidance applications in buildings or underground in mines, an occupancy grid is commonly used. The evidence or occupancy grid method [Martin and Moravec 96; Elfes 89] is one of the most widely used representations for several reasons; it is easy to implement, has fixed memory and computational requirements and allows a statistical representation of the environment (albeit at spatially discrete intervals). † Lange et al [Lange and Detlefsen 91] used a volume cell with only two states (full or empty) to indicate the presence of a target, though most systems now use statistical representations that allocate a target presence probability to each cell. More recently, other probabilistic methods such as particle filters [Dissanayake et al 00; Fox et al 00] and sums of Gaussians [Sajumder et al 00] have been used that allow probabilistic mapping in a non-spatially discrete way.

An alternative to probabilistic mapping is the use of feature primitives. This is where a set of primitives (edge, corner, wall, building, etc.) are defined, extracted from the sensor data and located in Cartesian space. For radar, this method suffers from the fact that primitives can be very difficult to extract, and may be ambiguous. Combinations of the various representations are possible.

8.4 Review of Work Done in the Field

8.4.1 Indoor Applications

8.4.1.1 Technische Universität München

Work undertaken at the Technische Universität München in the early 90's [Lange and Detlefsen 91] was centred around a demonstrator vehicle that contained a millimetre wave radar providing 3-D real time information up to a range exceeding 50m as well as a short range laser scanner, ultrasonic sensors and shaft encoders on the wheels.

The Pulsed-Doppler radar operating at 94 GHz produced a 1.7ns pulse by gating a continuous wave signal using a high-speed PIN switch. It operated with a 168mm diameter antenna to obtain range and angular reso-

†See discussion in chapters 4 and 7 too
Fig. 8.8 Reconstructed 3-D geometry of an Indoor Scene (walls, door, table and locker) based on $20 \times 20 \times 20 \text{cm}^3$ volume cells

olution to measure indoor topology to a volume pixel (voxel) resolution of $0.2 \times 0.2 \times 0.2 \text{m}$. The beam could be directed through $360^\circ$ in azimuth and $\pm 20^\circ$ in elevation using the classical mirror-scanner design. In addition, a combination of Doppler based velocity measurement and split-gate techniques gave the sensor the unique ability for precise range tracking.

Pulsed technology was utilised in preference to FMCW because, at the time, processing speeds were not sufficiently fast to handle the requirements of a real-time FFT. Operating at a pulse repetition frequency of 1MHz, the received signal was sampled after a programmable delay and filtered then digitised at 10kHz. This data-rate was so low that real-time vehicle guidance could not occur.

Problems encountered that are common to all active systems included specular scattering from targets and ghost targets, generated by multiple bounces. A constant false-alarm-rate (CFAR) technique was used to compensate for the variations in signal level and the operating range was restricted to eliminate ghost images.

Figure 8.8 shows some results, using a volume cell representation.

Because of the phenomenal increases in computer processing speed in the last decade, most other researchers have used FMCW techniques [Clark and Durrant-Whyte 97; Foessel 00; Suomela et al 95; Belov 01] for their
systems.

8.4.1.2 St. Petersburg State Technical University

Non-mechanical scanning technology [Belov 01] is being investigated by the Radiophysics Department of the St. Petersburg State Technical University, where a 35 GHz FMCW system has been demonstrated. It uses a fixed horn to illuminate the scene and an electronically controlled linear ferrite Integrated Phased Array (IPA) receive antenna. The IPA will produce a scan angle of \(50^\circ\) \((-5^\circ\) to \(+45^\circ\) from normal) without switching the feed port. This can be extended to \(90^\circ\) if the feed port is switched.

8.4.2 Outdoor Applications

Radar systems designed for outdoor applications generally operate at longer ranges with poorer resolution than their indoor counterparts.

8.4.2.1 Robotics Institute: Carnegie Mellon University

At Carnegie Mellon University [Langer 97], an unscanned FMCW radar operating at 77 GHz was developed to investigate autonomous navigation on highways. This system has a range resolution of 0.5m produced by a swept bandwidth of 300MHz and uses an array of four horns and a cylindrical lens to produce an antenna pattern with an elevation beamwidth of \(3^\circ\) and an azimuth beamwidth of \(12^\circ\). The bearing to a point target is determined by using the received signal phase difference from the target in adjacent channels. This requires careful phase matching of the components of the four receiver chains. Because the spacing of the horns exceeds \(\lambda/2\), grating lobes are generated in azimuth and ghost targets appear. More recent work undertaken at CMU [Foessel 99] aims to produce a non-mechanical imaging radar for field robotic applications. Techniques to produce low cost phased-array-type scanners and to synthesise narrow beams using relative motion between the radar and the target are being investigated. At this time, the program is in its infancy and a conventional scanning radar is being used [Foessel 00]. The existing sensor is a dual antenna 77 GHz FMCW unit with a \(2^\circ\) elevation and \(1^\circ\) azimuth beamwidth, that scans over \(64^\circ\) in azimuth and over \(8^\circ\) in elevation by selecting a sequence of four stacked beams. Operational range is limited to between 1 and 64m. The theoretical range resolution is 0.5m obtained by a swept bandwidth of 300MHz.
over a period of 256μs. Linearization is accomplished by a lookup table, which cannot provide linearity to match those produced using closed loop techniques, but should nevertheless be adequate for the short operational range. An interesting graphical representation shows the log value of each cell in the 3-D occupancy grid as the radius of an opaque sphere.

8.4.2.2 Helsinki University of Technology

A similar radar operating at 93 GHz [Suomela et al 95], was developed by the Helsinki University of Technology. It used a pair of 150mm diameter horn-lens antennas to produce a pencil beam with a beamwidth of 1.5°. Because mirror scanning with the dual beam configuration is difficult to perform, azimuth scanning was achieved by rotating the whole radar. A range resolution of 0.2m for an operational range of 50m was achieved using wide band FMCW techniques.

8.4.2.3 Australian Centre for Field Robotics: Sydney University

The first radar that was developed at the Australian Centre for Field Robotics of the University of Sydney operated at 77 GHz using a closed loop linearised FM sweep of 600 MHz giving a theoretical range resolution of 0.25m. A 2-D mirror scanner with an antenna aperture of 150mm produced a beamwidth of 1.8° that could be scanned over 360° in azimuth. To allow for beacon identification in highly cluttered environments, the unit had a dual polar receiver that could discriminate between specular returns from unwanted clutter (typically containers in this application) and those from strategically placed polarisation-modifying beacon reflectors. Filtering at baseband compensated for the distance dependent reduction in signal power with range. Then the received signal was digitised to 12 bits and a spectrum obtained using a Fast Fourier Transform algorithm running on a high-speed processor. A 4096-point FFT with its associated pre- and post-processing could be achieved in under 4ms. This allowed a new 2-D polar reflectivity image with an angular spacing of 1.4° to be formed every second. Reducing the maximum operating range, and hence the length of the FFT decreased the processing time significantly. Quadratic interpolation, using the bin containing the target with the adjacent bins, was applied to improve the accuracy of the measured range to a single point target. 

†Accuracy should not be confused with the range resolution that remains unaltered
Review of Work Done in the Field

Fig. 8.9 Vehicle Position Estimation Using Radar Beacon Updates

For vehicle navigation [Clark and Durrant-Whyte 98b], an extended Kalman filter running at 20Hz used suitably extrapolated observations of the radar beacons within the range of the radar to constrain the vehicle position uncertainty. A comparison with differential GPS (figure 8.9) shows the system performance over an elliptical path.

This radar technology has been extended at the ACFR to produce an extremely sensitive collision-warning device that is capable of detecting a seated man in the path of the vehicle at a range of 50m (figure 8.10).

A curved mirror spreads the elevation beam to compensate for irregularities in the road surface that can cause the vehicle to pitch or roll. New hardware to extend the operational range of this technology to 500m is being developed at the ACFR [Brooker et al 01]. This has led to the de-
development of a general-purpose 77 GHz module that can be assembled into radar systems of arbitrary complexity. For example, a single module with a Gunn voltage controlled oscillator can be used to produce an open-loop FMCW radar. A second module and a Gunn local oscillator allows for closed loop linearisation with the associated improvement in range resolution. The addition of subsequent modules increases the number of receiver channels to accommodate phased array or dual polar configurations as required. Various antenna sizes, from a standard gain horn up to a 250mm diameter horn-lens, and configurations are also provided so that the radar hardware can be better matched to the application.

The scanner that is being developed for these applications, shown in Figure 8.5, offers a positioning accuracy of 0.2° over 360° in azimuth and 90° in elevation for an antenna aperture of up to 250mm. Most of the applications envisaged for this new hardware involve long range, high-resolution visualisation applications for industrial and mining processes that must continue under conditions of low visibility, or that are too dangerous for human operators.

8.5 Airborne Radar Systems

Millimetre wave imaging radar technologies have been developed for the autonomous navigation of aircraft. This technology can be used to guide long range standoff weapons such as the Matra Apache [Moore 79] or for low visibility pilot aiding or autonomous landing systems [Bui and Morton 91; Koester and Vaillancourt 92]. A system developed by AMS in South Africa [Brooker et al 01] for airborne navigation and landing research offers some interesting technology that is common to such systems.

8.5.1 Imaging Range and Resolution

An airborne radar can either produce images of the ground by using the forward motion of the aircraft and an offset beam or by physically scanning the beam over the terrain and measuring the intensity of the reflected signal as a function of time.

It is difficult to assess the actual resolution that is required to obtain a specific navigation accuracy, as the relationship is strongly dependent on the target characteristics and the correction technique used. Table 8.2
summarises some results.

As an autonomous navigation aid, the radar sensor would have to resolve sufficiently small features over a sufficiently large area to ensure unambiguous association with a previously generated local map of the specific waypoint. A reasonable compromise for feature resolution, based on the data presented in Table 8.2, is between 10 and 20m. Further constraints are that the radar image covers a sufficiently large area to account for aircraft drift from the previous waypoint and the limited size of the local map.

It has been shown [Moore 79] that it is the pixel area and not its shape that defines the effective resolution (out to a 5:1 aspect ratio). This allows a tradeoff to be made between the radar’s range and cross-range resolution while still maintaining a constant target identification capability. For example instead of using a square pixel it is possible to use one with the same area but longer and thinner. Scanning requirements and various other mechanical restrictions limit the actual azimuth aperture available for the antenna to about 400mm. The frequency then determines resolution, as described in Table 8.3. §

To achieve both the equivalent of a 15 × 15m resolution at a range of 3km and comply with an aspect ratio of 5:1 reduces the options to 94 and 140GHz. The former was selected, as the atmospheric attenuation is lower and the technology more mature at the lower frequency.

The required range resolution of 7.7m can be achieved in a number of ways as outlined in table 8.4. It requires a pulse width of about 50ns or a linear frequency chirp of at least 20MHz. Solid state transmitters operating at 94GHz can produce 20W with pulse widths of less than 100ns, or 500mW continuously using a single IMPATT diode. Higher powers can

---

§Target identification only requires a pixel that has the same area as a 15 × 15m² pixel, so we can tradeoff cross range resolution (which is difficult to achieve) to range resolution so long as the total pixel area remains 225m².
Millimetre Wave Radar for Robotics

<table>
<thead>
<tr>
<th>Frequency (GHz)</th>
<th>Beamwidth (deg.)</th>
<th>Resolution for 15x15m equivalent at 3km (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>1.5</td>
<td>80 x 2.8</td>
</tr>
<tr>
<td>94</td>
<td>0.56</td>
<td>30 x 7.7</td>
</tr>
<tr>
<td>140</td>
<td>0.375</td>
<td>20 x 11.5</td>
</tr>
</tbody>
</table>

Table 8.3 Minimum beamwidth as a function of frequency for a 400mm antenna aperture. Resolution is expressed as an area equivalent to that of a square 15m x 15m pixel.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Advantage/Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short real pulse</td>
<td>Peak power limitations</td>
</tr>
<tr>
<td>FMCW</td>
<td>Linearity and antenna VSWR</td>
</tr>
<tr>
<td>Interrupted FMCW</td>
<td>Linearity and processing</td>
</tr>
<tr>
<td>Frequency step</td>
<td>Long sweep duration</td>
</tr>
<tr>
<td>Pulse Compression</td>
<td>Complex hardware and processing</td>
</tr>
</tbody>
</table>

Table 8.4 High Range Resolution Techniques

be achieved, at a price, using a power combiner [Kuno and Fong 79] or using high efficiency frequency multipliers.

To maximise the average power transmitted and hence the operational range an interrupted version of FMCW was implemented. A 50% duty cycle provided a maximum range of 3km with an average power of 250mW. The swept bandwidth was selected such that the frequency chirp that occurred during the transmit time of 20\(\mu\)s exceeded 20MHz by a factor of at least 1.3 to compensate for the windowing function implemented on the received signal. A total chirp of about 140 MHz in 84\(\mu\)s was in fact used.

A picture of the radar set is shown in figure 8.11. Further details of operation are given in [Brooker et al 01].

8.5.2 Results

Figure 8.12 shows a complete image of an airfield, following image integration (which helps to remove speckle). It used an average of 4 scans. The main features that are visible are a section of runway about 130m wide and 500m long followed by a narrow section about 25m wide and 600m long.
Waypoint Navigation Process

Two taxiways lead towards a cluster of hangars that are surrounded by a concrete apron. A wall and clusters of buildings are visible at the narrow end of the image. The limited dynamic range of the printing process requires that the image be clipped with resultant loss of resolution for the bright objects (particularly the hangars.) A single 30° scan took about 7s to complete, which for a forward velocity of 30m/s resulted in a displacement of 200m before the same slice was re-illuminated.

8.6 Waypoint Navigation Process

Features which provide good contrast and can be related to map data are used for navigation using waypoints. Ideal candidates are groups of well-spaced point features or homogeneous structures sufficiently large to cover
many pixels such as lakes or runways that are embedded in a uniform "sea" of vegetation.

The process is illustrated using the airstrip shown in Figure 8.13. First the image is built up on a Cartesian frame pinned to a specific latitude and longitude as determined by the GPS/INS based estimate of the radar position. Because of possible uncertainties in this estimate caused by the lack of GPS data errors of up to 500m must be catered for. The area is analysed using aerial photographs and any other available data to produce a series of reference features that will probably be visible on a mmWave radar image. These include point reflectors with dimensions larger than the resolution of the radar that are usually made up of an ensemble of closely packed reflectors, lines which generally consist of rows of point reflectors closely spaced and finally edges where a change of contrast occurs. The individual features are reduced to simple polar co-ordinate based structures defined relative to a fixed point on the image and stored in a database for the area.

The radar image, which has been normalised by the pixel integration count, is scaled and clipped to enhance the contrast for low signal to noise ratio areas before further processing. It is then passed through a filter matched to enhance the contrast of a specific feature type, before being re-normalised to allow for histogram-based binarisation. Finally the binary image is processed to produce fact features that conform to the same polar format as the reference features do.
8.6.1 Navigation Error Estimation

Each reference feature of a specific type is compared to similar fact features within a given radius and with similar characteristics (length, orientation, area, etc.) to determine whether they match. A list of possible matches for each feature is stored. The distance between all pairs of reference features and the distance between pairs of all their possible matches is then compared and rated according to some tolerance criteria. The rating of each fact feature accumulates asymptotically towards unity during the whole search process. Finally the fact feature with the highest score is considered to correspond to its associated reference feature. The distances between all the references and their associated fact features are then combined in some optimal sense to produce an estimate of the navigation offset with respect to the GPS/INS based position estimate.

8.6.2 Results

The system performance was measured by flying in the direction of the runway, but offset from it and performing the imagine-feature extraction and navigation processes before landing at a known reference. The position estimate generated by the GPS/INS system and that generated by the radar were logged. This process was repeated as often as possible to obtain a statistically significant sample, as shown in 8.14:

This method assumes that there is no drift in the navigation system between making an image and landing, which is patently not the case so the actual performance is superior to that shown. There were also some scaling and alignment errors between the navigation system and the radar.
system that resulted in an offset. However, the standard deviation, $\sigma$, of the errors in each case was better than that provided by the navigation pack alone as shown in Table 8.5 which shows that this is a simple and accurate navigation technique.

### 8.7 Summary

This chapter has described the operation of millimetre wave radar and discussed the factors that affect its performance. The technology is still in its infancy as a robot guidance sensor but it shows promise particularly for long range outdoor applications, or in environments where the atmosphere is opaque to shorter wavelengths. A summary is included of research systems in robotics, and some results presented from the work at the Australian Centre for Field Robotics: Sydney University.
PART III

OPTICAL SENSORS
This page is intentionally left blank
Chapter 9

Optoelectronic Range Sensors

Briseida Deyarina Maytorena Sanchez

9.1 Introduction

This chapter describes optoelectronic sensors and sensing methods which can be used in diverse mobile robot or autonomous vehicle (AGV) applications such as dead reckoning, range-finding, tracking, map building and navigation.

An optoelectronic transducer is an electronic device which can convert a physical quantity by emission and/or reception of photons or quantum energy from or into an electrical quantity. That is, an optical transducer converts electricity into photon emission or photon reception into electricity. Transducers can be active and/or passive. In this context, a passive transducer is one that obtains information from the environment without modifying it, whereas an active one interacts with the environment by sending a signal and registering the response produced by the environment. The processing required for passive sensors is usually trivial, and in this chapter we will only consider active sensors.

The main use of active optoelectronic sensors is range measurement, the topic of this chapter.

9.2 Range-Finders

9.2.1 Introduction

Optical range-finders consist of an optical emitter which transmits a modulated optical wave into the environment, where the emitter can be an light emitting diode (LED) or a laser. The emitted wave arrives at a target where it is reflected back to the sensor. Range from the sensor to the target may be calculated in one of many forms, either by measuring the time the
wave takes to travel to the target and back to the sensor (time of flight TOF), by measuring the phase shift between transmitted and received signal (amplitude modulation AM) by measuring the beat frequency between transmitted and received signal or by measuring the geometric point of arrival of the signal (triangulation). The performance of the sensor will then depend on the quality of the energy arriving at the optical detector of the sensor. The strength of the received signal accounts for the quality of the signal, keeping an acceptable signal to noise ratio ensures accuracy and resolution. Therefore, careful consideration must be given to the optical or radiometric design of optical sensors.

9.3 Radiometric Design

An optical range-finder design is very similar to a communication system comprising transmitter, receiver and processor. A communication system uses an antenna to transmit and receive information; similarly the range-finder uses an emitting diode or laser as a transmitter and a photodiode, photodetector or CCD camera as a receiver. Communication systems use a modulated wave over a high frequency carrier signal to send information. Similarly the optical range-finder emits a modulated signal over the wavelength of the light source. In both systems the modulation can be AM (amplitude modulated), FM (frequency modulated) or pulsed.

In a communication system the transmitted signal contains the desired information and is obtained by filtering out the carrier signal at the receiver. In a lidar range-finder the received signal needs to be compared with the transmitted signal in order to obtain the range information, whereas in a triangulation range-finder the point of arrival of the received signal needs to be recorded. This is done by comparing the modulating voltage applied to the emitter with the receiver voltage on a photodetector. The voltage on the photodetector is proportional to the reflected energy arriving at the detector. The amount of energy that arrives at the detector will mainly depend on the radiometric design of the range-finder.

The optical or radiometric design of the range-finding sensor depends on several factors, including:

- the distance from the sensor to the target
- the output power of the source or emitter
Fig. 9.1 The figure shows how a laser beam striking a surface will have both specular reflection and diffuse reflection, where only a fraction reaches the receiver aperture

- the detector sensitivity
- the reflectivity of the target
- the aperture area of the receiving lens
- the orientation of the aperture with respect to the surface of the target
- the efficiency of the transmission and receiving optics

Some of these effects are internal factors and are a result of the sensor design, such as the output power of the emitter, sensitivity of the detector, aperture area, efficiency, all of which are discussed below. Others are external and depend on the environment, for instance the distance from the sensor to the target, or the nature of the target. However, the most important concern is that sufficient radiation is returned from the target. The type of reflection which occurs depends on the nature of the target and its position relative to the emitted beam, as illustrated in Figure 9.1. For a smooth target, such that any irregularities on the surface are smaller than the wavelength of the beam, there is specular reflection. That is, the angle of reflection will be equal to the angle of incidence \( \theta_i \). For a rough surface the reflection is diffuse or scattered in all directions across the surface of a hemisphere or an area of \( 2\pi r^2 \). However, most surfaces reflect somewhere between these two extremes. The amount of radiant flux or optical power received by the detector from an observation angle \( \theta \) depends mainly on distance and reflectivity of the target. The reflectivity \( R \) of a target varies from zero to one, where an ideal "non-reflecting" surface has a reflectivity
of zero and an ideal mirror surface has a value of one.

The image is the area illuminated by the source or emitter. To determine the amount of power which is transferred to the image, it is necessary to determine the type of source. The source can either be an extended source or a point source, where a point source is one which has dimensions much smaller than the observation distance. If the emitter has a diameter $d << t_0$, where $t_0$ is the distance from the emitter to the target, then the emitter is considered a point source, as is the case for a laser. Most point sources radiate into a solid angle of $4\pi$ steradians, and only a fraction of the emitted energy is directed to image space. However, if a laser of output power $\phi_e$ is used as the source, then all of the transmitted power is transferred to the image. So, the power in the image is $\phi_i = \mathcal{T} \phi_e$, where $\mathcal{T}$ is the transmittance of the system and has a value between zero and one [O'Shea 85; Klein and Furtak 86].

9.3.1 Specular Reflection

Specular reflection occurs when most of the reflected power from the target leaves the target at an angle equal to the angle of incidence. Specular reflection will reach the sensor when the angle of incidence of the emitted beam is small, approximating zero degrees. The amount of power reaching the detector in these cases will yield the maximum optical power reaching the detector.

The optical power due to specular reflection can be determined by the amount of radiance given by the source, in this case the object, the solid angle of this reflection and the solid angle of the collecting optics. The solid angle of the collecting optics is a function of the distance $t_0$ from the object to the collecting optics and is given by:

$$\Omega_c = \frac{A_c}{t_0^2}$$

where $A_c$ is the area of the collecting optics. This would be the area of the receiving lens or mirror as the case may be. The units of the solid angle are the steradian.

When the emitted beam strikes a target and most of the reflection is specular, the reflection leaves the target in a solid angle equal to the one arriving, which is the solid angle of the beam. For a beam with a divergence
\( \theta_{\text{dev}} \) the solid angle is [Boyd 83]:

\[
\Omega_s = 4\pi \sin^2 \left( \frac{\theta_{\text{dev}}}{2} \right)
\]  
(9.2)

For a source element of area \( \Delta A \) emitting a flux \( \Delta \phi \) over a solid angle \( \Delta \Omega \), the *radiance*, \( \Delta L \), is by definition:

\[
\Delta L = \frac{\Delta \phi}{\Delta A \Delta \Omega}
\]  
(9.3)

The radiance expresses the flux per unit area per unit solid angle from the source. The units of radiance are \( \text{Wm}^{-2}\text{steradian}^{-1} \).

In a direction which is at an angle \( \theta \) to the element, the luminance is reduced by \( \cos \theta \).

Now consider a specular surface illuminated by a source of strength \( \phi_e \), as shown in figure 9.3. The radiation is reflected at an angle \( \theta_s = \theta_i \) over a solid angle \( \Omega_s \). From equation 9.3 the radiance is given by:

\[
L_{os} = \frac{\phi_o}{A_o \cos \theta_s \Omega_s}
\]  
(9.4)

The flux from the object is equal to \( \phi_o = T R_s \phi_e \), where \( R_s \) is the specular reflectivity and is a function of the surface of the target and has a value between zero and one.

A quantity related to radiance is the *intensity*, which has units \( \text{Wsteradian}^{-1} \). The intensity of the radiation from the surface is:

\[
I_{os} = L_{os} A_o = \frac{\phi_o}{\Omega_s \cos \theta_s}
\]  
(9.5)

In general the source and receiver are not perfectly aligned with the surface. In this case, shown in figure 9.3 the power reaching the collecting optics (scanning mirror, lens or active detector area), is given by:

\[
\phi_{cs} = I_{os} \Omega_c \cos \theta_o \cos \theta_1
\]  
(9.6)

where \( \Omega_c \) is the solid angle of the collecting optics. From equation 9.5

\[
\phi_{cs} = \frac{\phi_o}{A_o \cos \theta_s \Omega_s} A_o \Omega_c \cos \theta_o \cos \theta_1
\]  
(9.7)
Fig. 9.2  Figure 9.2 shows how a beam striking a surface at an angle of $\theta_i$ illuminates an area $A_o$. The solid angle of the specular reflection follows the same solid angle as the beam from the emitter.

$$R_s T \frac{\phi_c A_o \cos \theta_o \cos \theta_1}{4 \pi \left[ \sin^2 \left( \frac{\theta_{dBK}}{2} \right) \right] t_0^2 \cos \theta_s}$$  \hspace{1cm} (9.8)

This expression can be simplified if the emitted beam and the received optical power follow the same optical path and therefore the observation angle $\theta_o$ is equal to the angle of incidence $\theta_i$, as illustrated in Figure 9.4.
Fig. 9.4  The figure shows how the angle of incidence \( \theta_i \) formed between the normal of the surface of the target and the laser beam, is equal to the observation angle \( \theta_o \) of the collecting optics.

In addition in the specular case, for reflection to reach the collecting optics \( \theta_o \) has to be equal to \( \theta_s \). Therefore \( \theta_o \) and \( \theta_s \) need to be equal to zero or have a value small enough so that the reflection is within the field of view of \( A_c \). Since \( \Omega_l/\Omega_s \) represents the portion of reflected power collected by the lens or receiving optics, then, all of the reflected energy is collected. This is because the beam is from a laser and the solid angle of the beam is very small compared to \( \Omega_l \). Then equation 9.8 is simplified to:

\[
\phi_s = \mathcal{T} R_s \phi_e
\]  

(9.9)

This is the flux entering the receiving optics. If a lens is used the flux is then imaged to the detector. As can be seen from this equation, the collected optical power due to specular reflection is a large fraction of the emitted power. Losses are due to transmission and the nature of the target. Specular reflection is exploited in triangulation and proximity sensors.

### 9.3.2 Diffuse Reflection

When the reflected energy is diffuse, the flux of this reflection covers the area of half a sphere so that the solid angle is \( \Omega_d = 2\pi \). From Eq. 9.3 the radiance of the object is:

\[
L_{od} = \frac{\phi_o}{\Omega_d A_o \cos \theta_o}
\]  

(9.10)

where \( \phi_o = \mathcal{T} R_d \phi_e \) and \( R_d \) is the diffuse reflectivity.
When the flux of the object is due to diffuse reflection, most of the radiation is directed along the surface normal. This flux varies depending on the viewing angle, so that it can be described as a Lambertian source. For this case the intensity of the object is expressed by:

\[ I_{od} = \int L_{od} \cos \theta_o \, dA_o \]  
(9.11)

and

\[ I_{od} = \frac{\mathcal{T} R_d \phi_e}{2\pi} \]  
(9.12)

The optical power reaching the collecting optics is given by Lambert’s Law:

\[ \phi_{cd} = I_{od} \Omega_c \cos \theta_o \]  
(9.13)

substituting \( I_{od} \),

\[ \phi_{cd} = \frac{\mathcal{T} R_d \phi_e A_c \cos \theta_o}{2\pi \Omega^2} \]  
(9.14)

In this case, the collected optical power from diffuse reflection depends on the distance from the object to the collecting optics and the observation angle \( \theta_o \). However, for small values of \( \theta_o \), specular reflection will also reach the collecting optics.

### 9.3.3 The Emitter and Detector

The preferred emitters for range-finders are lasers. The difference between a laser and other light sources is that a laser is almost a perfect coherent source. A light wave is coherent if it has a single frequency (temporal coherence), and the shape of the wavefront does not vary with time (spatial coherence). This translates into a small line-width at any fixed distance from the source [Beesley 72].

The advantage of using a coherent light source over an incoherent one is the higher resolution obtained. This is because in an incoherent light source such as an LED the spot illuminated on the target by the source becomes larger as the distance increases, and the reflected optical power represents a larger area rather than just a point on the target. Also, with

*The radiance of a Lambertian source is independent of the observation angle [Boyd 83]*
incoherent light sources, the possibility of spurious data due to multiple targets increases as the spot size becomes larger. This happens when the spot illuminates more than one target at different ranges.

When a mobile robot is used in human environments, eye safety precautions should be considered when using lasers. Output power, collimating optics and exposure times are all factors which determine whether a laser is safe. LEDs on the other hand have the advantage of being eye safe and performance can be improved with adequate collimating optics.

The average emitted power can be calculated by averaging the instantaneous emitted power. The optical power of the emitter \( \phi_e \) is given by:

\[
\phi_e = \overline{P_T(t)} = P_{To}
\]  \hspace{1cm} (9.15)

The detector is effectively a power meter which converts the input optical power to a proportional current signal; this is described by the sensitivity of the detector. The higher the sensitivity the better the response to low optical power signals. Since the power reaching the detector decreases with range, sensitivity in addition to noise limit the maximum range of the sensor. Therefore it is important to select a detector with a high sensitivity to ensure low power signal detection.

The detector sensitivity depends on the characteristics of the device. Table 9.1 shows the sensitivity of three of the most popular detectors considered for receiver applications.

The higher the sensitivity of the detector, the higher the current signal response to a given optical power signal. Therefore, a photo-multiplier would be the ideal choice for high sensitivity of the optical design. However, although photo-multipliers are high performance detectors, offering high current amplification, low noise, high sensitivity, high bandwidths, and good long term stability, they require high bias voltages, up to 2800 V [Tech:Opto].

Photodiodes have good linearity, large bandwidths, excellent long term stability, low cost and compact size. These characteristics make these detectors suitable for laser receiver applications. The main types of photodiodes are PIN photodiodes and avalanche photodiodes (APD). PIN photodiodes require considerable amplification, but only a few volts for biasing, whereas APDs have internal amplification, and higher voltage requirements (typically 100 V). However, APD modules with biasing circuitry and tem-
Optoelectronic Range Sensors

<table>
<thead>
<tr>
<th>Detector</th>
<th>Order of Photo Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photo-multipliers</td>
<td>up to $100, K A/W$</td>
</tr>
<tr>
<td>PIN Photodiodes</td>
<td>0.4 to $1, A/W$</td>
</tr>
<tr>
<td>Avalanche Photodiodes</td>
<td>up to $1000, A/W$</td>
</tr>
</tbody>
</table>

Table 9.1  Photo Sensitivity of Photo-detectors

Temperature compensation may be obtained from various manufacturers as a package and this simplifies sensor design [Tech:Hamamatsu].

9.3.4  Optical Geometry

The optical geometry has a significant effect on the received optical power. Factors such as detector position, lens area, optical paths and observation angle are of critical importance in the design of the sensor.

Consider the optical geometry of two lidar sensors, Miller and Wagner’s [Miller and Wagner 87a] and Brownlow’s [Brownlow 93] sensors (see figure 9.5). Both sensors keep the same optical path for emitter and detector. In Miller and Wagner’s design this is achieved by placing the collecting lens and detector behind the emitter. In this configuration, the emitter blocks part of the reflected energy entering the sensor. Although this reduces sensitivity, it avoids saturation of the detector for targets at close range. In Brownlow’s design the lens and detector are placed at a $90^\circ$ angle from the emitter. A mirror with a beam-splitter in its centre is used to maintain the same optical path. Saturation from close range targets is avoided in the electronic design.

The disadvantage of using a beam-splitter is that a fraction of emitted power is lost as the beam exits the sensor and another fraction of detected power is lost as the collected power reaches the beam-splitter. This is because, when the emitted beam reaches the beam-splitter, part of the beam goes out of the sensor and the other part is lost inside the sensor. The collected power follows a similar path; while part of it is directed to the detector the rest of it goes through the beam-splitter. The advantage of the beam-splitter is that the same optical path for emitter and detector can be maintained. Also, since the collected optical power is proportional to the area of the mirror or lens, there is less power lost by using a beam-splitter rather than placing the emitter in front of the collecting optics, as in Miller and Wagner’s design.
Fig. 9.5 The top figure shows the optical geometry of Brownlow’s sensor [Brownlow 93]. The emitter passes through a beam-splitter in the centre of a mirror. The beam is directed by the scanning mirror out into the environment. After the beam strikes a target the reflected energy follows the same path as the emitted beam. Part of the reflected energy goes through the beam-splitter and the rest is reflected by the mirror. The reflected energy is then focused to the detector via the lens. The figure below shows the optical geometry from Miller and Wagner’s design [Miller and Wagner 87a]. The emitter beam is directed out to the environment by the scanning mirror. The reflected energy from a target reaches the scanning mirror and is directed to the lens which focuses this energy to the detector. Part of the reflection reaching the mirror will be blocked by the emitter which is placed in front of the detector lens.

An alternative approach is shown in figure 9.6. The laser beam from the emitter passes through the orifice in the centre of mirror $M1$ and strikes the surface of the scanning mirror which directs the beam to the target. Part of the reflected power from the target follows the same path to the scanning mirror, then to $M1$ which directs it to the lens and images this energy to
Fig. 9.6 The figure shows the optical geometry of a lidar. The beam exits the emitter and passes through an orifice in the centre of mirror M1. The scanning mirror M2 directs this beam to the target and also collects a fraction of the reflected energy from the target. Mirror M1 reflects this energy through the lens and to the detector.

The optical power reflected by the object may be specular, diffuse or a combination of both, as illustrated in figure 9.1. The amount of each of these components is a function of the nature of the target, the angle of incidence $\theta_i$ on the target, the observation angle $\theta$, the solid angle of the reflected energy and the solid angle of the lens. The solid angle of the lens determines the portion of all the reflected energy which is collected by the lens. The advantage of this geometry is that only a portion of any specular reflection will reach the detector, thus avoiding saturation. This is because most of the flux will go through the orifice.
Ranging Sensors

Transducer 1

\[ y_1 = \frac{B \sin \alpha \sin \beta}{\sin(\alpha + \beta)} \]

where:

\[ \alpha = \tan^{-1} \frac{y_1}{x_1} \]

\[ \beta = \tan^{-1} \frac{y_2}{x_2} \]

Fig. 9.7 Range can be calculated through simple geometry using the law of sines. Two transducers are used to determine \( x_1 \) and \( x_2 \) (the origin is taken as the centre of the lenses in the transmission and receiving paths). These can both be passive or one active and the other passive.

9.4 Ranging Sensors

9.4.1 Triangulation

This is probably one of the oldest methods used to determine distance, since range can be determined through simple geometry. In the sixteenth century Rojas and Digges [Rojas 51; Digges and Digges 91] describe how an astrolabe\(^{\dagger}\) can be used by triangulation to target a gun.

Sensors which use the triangulation method, determine range using the law of sines. Triangulation range-finders can use either passive transducers or active and passive transducers. Figure 9.7 shows how this method can be used. The transducers can be two cameras (stereo vision) or a light source and a position sensitive device (optical triangulation).

\(^{\dagger}\)An ancient navigation instrument dating back to 170 B.C.
Consider the case of optical triangulation, where one of the transducers is a light source, such as a laser producing a point beam \( \dagger \), and the second transducer a position sensitive device such as a charged coupled device (CCD) or a lateral effect photodiode (LEP). The laser beam exits the sensor at a fixed angle \( \alpha \) and reaches a target at point \( P \), provided that the reflection from the target reaches the camera or photodiode so that \( x_2 \) can be determined, range \( \rho \) from sensor to point \( P \) can be calculated by:

\[
\rho = \frac{B \sin \alpha \sin \beta}{\sin(\alpha + \beta)}
\]  

(9.16)

where \( B \) is called the baseline of the sensor. The accuracy of the measured range depends on how accurately \( x_2 \) is determined by the transducer and the minimum and maximum range depends on the baseline of the sensor.

In optical triangulation the active transducer is called the *emitter* and is usually an LED or a laser which transmits either a continuous or a modulated signal. The passive transducer is called the *detector*. Although a detector is more sensitive to a laser and the laser is easier to focus than an LED, eye-safety precautions should be considered when choosing the power output of a laser. The problems of interference caused by ambient light can be avoided by optical filtering and/or modulating the transmitted signal. These techniques also improves the signal-to-noise ratio.

If the environment is to be scanned by the sensor, the emitted beam needs to be scanned. The scanning mechanism must be such that not only does it direct the position of the emitted beam at each scan point, but also that the detector receives the reflection from the target at each scan point.

Fast data rates may be achieved with a cell-parallel sensor chip [Gruss et al 92] or an optical RAM [Ito et al 90] used as a detector. These detectors consist of arrays or matrix of photo-detectors, called cells. When these detectors are used the light beam from the source must be "swept" across the object with a moving reflector. For each point illuminated in the object there will be a point of reflection in the cell-parallel sensor. In order to calculate range, the angular position of the reflector and the time of incidence in the detector must be known. The speed of these sensors is limited by the sweep rate of the light source. This technique is ideal for

\( \dagger \)Other illumination structures may be used, such as stripe, multi-pattern, multi-point or multi-stripe.
short-range high precision applications.

Scanning the environment can be done by synchronised scanning of the emitter beam and the field of view of the detector. This achieves a one-dimensional scan of the environment. Figure 9.8 shows a diagram of the sensor developed by Livingstone and Rioux [Livingstone and Rioux 86] that uses this scanning technique. The system uses a mirror driven by a galvanometer. The mirror is double-sided and permits synchronised detection and projection.

In order to obtain a two-dimensional scan two mirrors may be employed. Rioux [Rioux 84] proposed a new geometrical scanning arrangement that increased the measuring rate. The system consisted of a six-face pyramidal rotating mirror, two plane fixed mirrors, and a flat tilting mirror. The scanning geometry is shown in figure 9.9 where light from the source is directed to the rotating pyramidal mirror and reflected onto the stationary left-hand mirror. The larger scanning mirror deflects the beam to the target and collects the reflection from the object. The reflected light is transmitted to the stationary right-hand mirror, to the pyramidal mirror and into the detector.

The light source is a laser and the receiver is a lateral-effect photodiode. The major problem with this scanning geometry is the shadow effect, due to the points in the object not illuminated by the light source, or the points in the object that the receiver cannot “see”. Large shadow effects translates to lower resolution and missing parts of the object in the output image.

The amount of reflected signal reaching the detector depends on the collecting optics, the optical power output of the transmitter, the distance to the target and the nature of the target. Therefore an optical sensor which uses a photodiode will have difficulties in detecting reflections from matt dark coloured objects [Probert and Pears 93].

A novel design which eliminates the shadowing effect was proposed by Wolf and Beck [Wolf and Beck 00]. A laser is used as the emitter and the receiver consists of a ring placed around the emitter beam. An array of lenses is used to focus the reflection around the beam where the average output is used to calculate range.

The accuracy and range of a triangulation sensor are limited by the baseline or distance that separates emitter and detector. The triangulation geometry can achieve sub-millimetre resolution over small depths of field, but the resolution falls off rapidly as the depth of field is increased.
9.4.2 **Lidar**

Lidar or *Light Detection And Ranging* technique follows the general principles of radar to determine range. There are three intensity modulation techniques for lidar:

- Pulsed Modulation.
- Amplitude Modulation Continuous Wave (AMCW).
- Frequency Modulation Continuous Wave (FMCW).

All of these techniques require an emitter and detector. The emitter is usually a laser, although LEDs can also be employed. The advantage of the laser is that it can be focused to a small area compared to LEDs. If the spot size is too large, the spot may illuminate multiple targets which can result in spurious data and erroneous results in AMCW and pulsed. This problem is addressed by Adams [Adams 93], who obtains successful results through filtering in the data processing stage of an AMCW sensor.

The detectors can be a wide range of photodiodes that vary in responsivity depending on the optical requirements of the application.

The advantage of using any modulation technique in optical sensors is that the interference of ambient lighting can be eliminated. It is also
possible to combine techniques; for example, AMCW and triangulation are both used in [Probert and Pears 93] to increase sensitivity and eliminate the effects of background light in the sensor.

9.4.2.1 Pulsed Modulation

Sonar and pulsed modulation follow the same principle of time of flight to determine range. A pulse train of light is transmitted by an emitter; when the signal reaches a target, a portion of the reflected signal returns to the sensor where a detector receives the reflected signal. Range is calculated by measuring the time of flight taken for the emitted pulse train to return to the detector as a reflected signal from a target:

\[ \rho = \frac{tc}{2} \]  

(9.17)

The velocity in this case is the speed of light \( c = 300 \cdot 10^6 \text{ m/s} \). The maximum unambiguous range is:

\[ \rho = \frac{c}{2f_p} \]  

(9.18)
where \( f_p \) is the pulse repetition rate. This method is convenient for ultrasonic transducers but the arrival times are so small for optical transducers that very high speed electronics are required (centimetre accuracy requires times of picoseconds to be resolved). The high speed electronics required, as with radar, make this a high-cost technique although it may be suitable for long range applications (see [Gustavson and David 92; Sizgoric et al 86; Guyen and Steton 86]). However, this method offers some advantages. Beam widths in optical pulsed lidar are smaller than in radar. This is because antennae can achieve narrow beam widths of a few degrees whereas lasers can be focused to sub-degree beam widths. A smaller beam width allows for greater angular resolution. Compared to AMCW, pulsed modulation requires higher power levels and therefore it has a better signal to noise ratio which is ideal for long range applications. It is also easier to design eye-safe systems using a laser with this method [Renairi and Johnson 90; Gregor et al 90; Brown 90; Brun 90]. This is because even with high power lasers the exposure times are small enough that it is simple to maintain permissible exposures. Eye-safety represents an advantage which has to be balanced against the fundamental expense of the method. The resolution in this technique is limited by the speed of the electronics.

9.4.2.2 Amplitude Modulation Continuous Wave

The AM communication technique is very similar to AMCW lidar although, in the case of lidar, it is an intensity modulation rather than a voltage modulation. The reflected signal from the target will have a phase shift proportional to range.

Unlike pulsed modulation the emitter transmits a continuous wave. In AMCW the reflected signal reaching a detector will have a phase shift relative to the range of the target to the sensor (see figure 9.10). The phase shift occurs in the modulating signal and range can be calculated by:

\[
\rho = \frac{\theta c}{4 \pi f} \tag{9.19}
\]

where \( \theta \) is the phase shift in the received signal compared with the transmitted signal, and \( f \) is the frequency of the modulating signal. Notice that range becomes ambiguous when the phase angle is greater than \( \pi \) radians. Therefore the maximum range is limited by the modulating frequency. The ambiguity interval can be extended by using multiple modulating frequen-
Fig. 9.10 Amplitude Modulation Continuous Wave. In AMCW range is dependent on the phase shift of the returned signal and thus becomes ambiguous at angles greater than $\pi$.

cies (MFCW) or by phase shifting. MFCW requires a second modulating frequency which increases the complexity of the receiver electronics. In spite of this, multiple modulating frequencies are often used in long range applications; this method is also known as beam modulation telemetry [Wilson and Hawkes 98].

Phase shifting consists of taking two measurements for each range reading where one of the measurements has an additional phase shift of $\pi$ added to the received signal. This allows the maximum range of the sensor to be doubled. Alternatively this technique may be used to double resolution of the sensor by doubling the modulating frequency while maintaining the maximum range of the sensor. However, the increase of frequency can only be achieved if the electronic design has the required bandwidth and signal to noise ratio [Maytorena Sanchez 99].

AMCW, is simpler, less expensive, and offers greater range resolution than that which can be achieved with pulsed modulation. This is because the phase measurement can be done at a lower frequency than the modulation frequency by using heterodyning methods as is done in radar.
9.4.2.3 Frequency Modulation Continuous Wave

There are two methods of FMCW lidar that function under the same principle. When the source or emitter is periodically swept in frequency and the reflected signal is mixed with the signal from the emitter, the result is a difference in frequency (beat frequency). The two methods differ in the way the mixing process is achieved. The mixing process can be done either in the electronics or in the optics.

Figure 9.11 shows the two mixing methods of FMCW. In the top diagram the intensity of the beam is periodically swept in frequency. The reflected RF signal is detected and then mixed electronically with the RF signal from the emitter. The range can be extracted from the difference in frequency from the two input signals to the mixer. Although this method is more complex and the data rates lower than AMCW, it has the advantage of detecting multiple targets since each target will have a singular tone which is proportional to the range of that particular target. This technique can be very precise and suitable for long range applications [Meyzonette et al 87].

In the second diagram of figure 9.11 the wavelength of the beam is periodically swept in frequency. The bandwidth of the frequency sweeps for this method are in the order of tens to hundreds of gigahertz, whereas the electronic mixing has sweep bandwidths below 10 GHz. The mixing is achieved by using beam splitters to combine a portion of the outgoing beam with the reflected optical signal before it reaches the detector. Once the combined signal is received by the detector, range can be extracted by measuring the beat frequency in the same manner as with the electronic mixing process. The restrictions of this method are due to the parameters of the light source, such as the coherence length which limits the range of the sensor, and the difficulty in constructing lasers that allow a wide bandwidth sweep of their optical frequency. FM tuning and the difficulty in maintaining linearity in the frequency sweep also contribute to data errors. Another limitation is the sensitivity to back-reflected light, which

---

5 Some authors refer to the electronic mixing method as chirped Intensity Modulated Continuous Wave rather than frequency modulation [McClure 90]. Since the method of extracting range is the same as the optical method in this thesis it will be referred to as FMCW.

6 Coherence length of a light source is the speed of light divided by the line-width of the beam from the source.
Fig. 9.11 Frequency Modulation. In this technique frequency is varied linearly against time and the displacement in the returned frequency is proportional to range.

contributes to noise and limits the accuracy of the sensor [Slotwinski et al 89]. In general, the complex requirements of the light source make this technique unsuitable for low cost application.

More detailed comparisons are made of these and other techniques in [McClure 90; de la Chapelle et al 89; Koskinen et al 91; Ahlers 89; Meyzonette and Saccomani 88].
9.5 Scanning Range-Finders

9.5.1 Introduction

The scanning system is one of the most important parts of the sensor. Its purpose is to direct the view of the sensor to the area in its environment where it should "look". This section presents a brief description of the scanning methods used for optical steering. Optical scanning methods include reflection, holographic and acousto-optic. A brief description of each of these methods is followed by description of different scanning geometry used for scanning in lidar techniques.

9.5.2 Scanning Methods

For an optical sensor where the light source is an LED or a laser, the simplest way to direct the light source to a certain point is through reflection, which can be achieved with a mirror. The mirror can then be tilted in azimuth or elevation to produce a line scan. The scanning speed is limited by the size of the mirror; the higher the scanning speed, the smaller the mirror. However, a small mirror translates into a small capture area for the reflected energy, limiting the aperture of the detector. Errors related to high speed scanning are noted by Colquhoun and co-workers [Colquhoun et al 91]; they include across scan errors, along scan errors and facet to facet errors. The first is caused by errors in the assembly of the scanner, producing weave, wobble, dynamic pyramidal error and dynamic tracking error. These errors can only be reduced by careful assembly of bearings. Along scan errors are those caused by fluctuations in the scan velocity and they produce distortions in the scanned image. The facet to facet error is caused when the angles between facets are not exactly the same, and also produce distortion in the output image or data. These errors can be minimised by the use of an encoder which provides feedback control to the drive circuitry.

Other scanning methods include acousto-optic and holographic scanners. These are used when the optical path of the receiver differs from that of the emitter. In these cases higher scanning rates are achieved.
9.5.2.1 Holographic Scanners

Holographic scanners are discs divided into sections of diffraction grating film, where the diffraction depends on the angle of incidence of the optical beam. If the film or grating is moved then the beam is scanned. Usually several sections of diffracting film are contained in one disc in order to obtain several scans as the disc is rotated (see Figure 9.12).

Holographic scanners have various advantages over mirror scanners; they have a smaller volume than mirror scanners, can be designed to suit a particular application, and are flexible enough so that different tilt angles or facet sizes can be placed on the same disc. The complexity of the initial design of a holographic scanner is one of the major disadvantages. The production cost can be justified only when a large volume will be produced. Once a prototype has been finished and tested, the production of other scanners is just a contact print of the original hologram. Optical alignment is very critical. If the beam is not aligned properly there is likely to be some power loss in the signal. These type of scanners are frequently used in photocopying machines and supermarket product code readers [Kramer 91; Kramer et al 91].

9.5.2.2 Acousto-Optic Scanners

Acousto-optics scanners work on the same principle as diffraction gratings. A sound wave is transmitted through a material causing the refractive index of the material to change, thus acting as a grating. A beam passing through
this material will be deflected an angle of $\theta_d$, where

$$\theta_d = \frac{\lambda}{2\Lambda} \quad (9.20)$$

where $\lambda$ is the wavelength of the light beam and $\Lambda$ is the acoustic wavelength. The deflection is proportional to the frequency of the sound wave, since $\Lambda = \nu_s/\nu_s$. $\nu_s$ is the frequency of the sound wave and $\nu_s$ the speed of sound in the material. The beam is scanned by sweeping the frequency of the sound wave. The disadvantage in this type of scanning is that the scan angle is limited by the material used and the frequency of the sound wave. Typically one can achieve scans no greater than one degree.

9.5.3 Some Scanning Sensors

9.5.3.1 The Sick Sensor: Pulsed Lidar

A pulsed time of flight infrared range-finder developed by Erwin Sick [Sick 94] uses pulsed modulation to determine range. This sensor, and its successors, is now very popular in robotics as it is robust and well engineered.

The sensor typically has a maximum range of 50 m, with range resolution of better than 20 mm, angular resolution of 0.5° and a scan time of 40 ms. The sensor uses a rotating mirror to produce the scan, which is over 180°. Other information about the sensor such as the type of detector employed was not available. The scanning geometry employed by the sensor is shown in figure 9.13.

The SICK range-finder was used in the Nomad robot in Antartic exploration. The range sensor was reported to distinguish between snow, blue ice and rocks. Range data was used for terrain classification and obstacle detection [Vandapel et al 99]. On clear days the sensor was reported to provide range data to within 2 cm although problems of erroneous range data were reported in bad weather conditions due to blowing snow.

9.5.3.2 AMCW Lidar Sensors

A number of AMCW lidar sensors use scanning mirrors. The scanning system of Nitzan, Brian and Duda [Nitzan et al 77] was one of the first, and used two mirrors joined at a 90° angle, controlled by 2 motors, pan and tilt. This scanning control enabled the sensor to obtain a 128 x 128 pixel image. The sensor measured ranges from 1 to 5 m, with a range resolution
of 1 cm and a modulation frequency of 9 MHz. The light source used was a laser and the receiver a photo-multiplier tube. However the time taken to acquire the scan was 2 hrs. Subsequent variations on this geometry included [Wesolowicz and Samson 87; Klein et al 87].

Rioux's geometry [Riou 84] (figure 9.9) has been developed by [Cameron and Doshi 89; Zukang et al 92; Jarvis 83; Chen and Ni 93] amongst others. who use a rotating polygon mirror to accomplish a line scan and a tilting mirror to complete an XY scan. Beyer, Jacobuc and Pont [Beyer et al 87] use a nodding mirror and a polygon mirror (Figure 9.14). Miller and Wagner [Miller and Wagner 87a], rotate a single mirror to give a radial scan of 360° and provide a conical scan by tilting the mirror. This sensor was used on "Blanche" a tri-cycle cart, for navigation in a factory environment. This sensor was probably the first to do such a wide field of view scan. A very similar sensor was designed by Brownlow [Brownlow 93]. The next chapter describes further development by Adams, also with a three dimensional imaging capability.

9.5.3.3 FMCW Lidar

In most scanning systems the mirrors are moved by motors, but Gielen and Slegtenhorst [Gielen and Slegtenhorst 91] use bending piezo actuators in their range-finder. The disadvantage of these actuators is that the mirror is rotated over a limited angle. The sensor has a $30° \times 30°$ field of view, a range of 1 to 50 m, and an accuracy of 10 cm. The light source is a laser diode which is frequency modulated to give a $1.5 GHz$ chirp (burst
of pulses) with a duration of 1 ms. The high frequency electronics elevate the cost of the sensor, compared to an amplitude modulated equivalent. However the sample rate is very good, with pulse width less than 2 \( \mu \)s to give a bandwidth of 512 kHz.

Meyzonnette, Remy and Saccomani [Meyzonette et al 87] designed a sensor with a working range of 2 km but with a field of view of only 1° x 0.5°, which can be considered appropriate for that range. The accuracy of the sensor was 5 m, using frequency modulation. The frame rate was 1.5 Hz and the frame was composed of 128 x 64 pixels.

9.5.4 Summary

The sections above describe the general characteristics of range-finding methods used for a wide variety of applications.

Triangulation can be fast and accurate over small depths of field; however, resolution falls off rapidly as the depth of field is increased. As with stereo vision, accuracy and range are limited by the baseline.

AMCW is relatively simple compared to FMCW and lower cost compared to pulsed lidar and FMCW. Unlike FMCW lidar, AMCW is independent of the light source and the modulation minimises the environmental lighting effects.

The range of AMCW lidar is limited by the received optical power and modulation frequency whereas in pulsed lidar, range is mainly limited by the received optical power. In FMCW the physical characteristics of the light source limit the maximum range obtainable. An advantage of FMCW is the possibility of detecting multiple targets which pulsed or AMCW can-
<table>
<thead>
<tr>
<th>Technique</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triangulation</td>
<td>simple, high resolution, low to medium cost</td>
<td>range limited by baseline, resolution falls off with range, short to medium range</td>
</tr>
<tr>
<td>Pulsed lidar</td>
<td>capable of eye-safe designs even at long ranges, good signal to noise ratios</td>
<td>high cost, resolution is limited by the speed of the electronics and the nature of the target</td>
</tr>
<tr>
<td>AMCW lidar</td>
<td>low cost, maintains resolution at long ranges, simple</td>
<td>resolution is limited by modulating frequency and nature of the target</td>
</tr>
<tr>
<td>FMCW lidar</td>
<td>resolves ranges for multiple target returns</td>
<td>resolution is limited by the optical characteristics of the emitter, high cost, complex</td>
</tr>
</tbody>
</table>

Table 9.2  Comparison of Pulsed, AMCW and FMCW Lidar

not detect.
This page is intentionally left blank
Chapter 10

AMCW LIDAR Range Acquisition

Martin D. Adams

10.1 Introduction

Many research institutions are now utilising lidar sensors for robot navigational experiments [Buchberger 93; Kweon 91; Borenstein et al 95] and the motivation for the work presented here is from the experience gained using commercially available sensors [Adams and Probert 96; Miller and Wagner 87b]. The first aim of this chapter is to show how the design of the individual modules within a lidar sensor can affect the information extracted from such a device. This information is often sampled from the sensor's outputs, the sensor being assumed to be a sealed black box, and used to attain some form of reliable environmental mapping. A view inside the black box is taken to show how various design factors influence the data interpretation or indeed its misinterpretation.

To place the sensor data analysis into perspective, the chapter begins by summarising the main physical principles used in light detection and ranging and gives examples of both research based and commercially available sensors which use each technique. Indeed during recent years a number of commercially available lidar sensors have appeared on the market. For the reader seeking a review of the currently available lidar sensors used for, or applicable to, mobile robot map building, a recent book by Borenstein et al. is recommended [Borenstein et al 95]. Based on previous work, a justification for favouring the particular measurement technique amplitude modulated continuous wave (AMCW) for the application of mobile robot navigation is discussed.

The intended application is mobile robot navigation, which is reflected within the design considerations, in terms of size, scanning speed and reach-
The lidar, which was designed and tested in this article, is a coaxial sensor, transmitting collimated, amplitude modulated light into the environment via a single scanning mirror. This can be continuously rotated about a vertical axis, and simultaneously swept with respect to a horizontal axis thus giving limited 3D coverage [Adams 99]. Since the device is coaxial, detection takes place in a direction retro (parallel) to that of transmission, thus eliminating the disparity problem associated with triangulation methods [Borenstein et al 95]. Range estimation results from the phase difference between the transmitted and received signals, this being defined modulo half the AMCW wavelength (15m here).

Section 10.2 takes an in depth view of the critical design factors for any lidar sensor, irrespective of the range measurement method, by defining 'worst-case' sensor to target parameters under which lidar range estimation should just be possible. A theoretical relationship between the minimum detectable photo-current, the lidar's transmitter power and the detector aperture is then derived.

Section 10.3 presents the theoretical performance limits, resulting from various noise sources, which can be estimated before electronic construction takes place. In order to focus on a particular application, an overview of the crucial components of an AMCW lidar sensor, which utilises an avalanche photo-diode (APD) as its receiver, is given in section 10.4, the design of which are used to optimise a real engineering implementation and resulting use of such a device.

Section 10.5 explores the causes of non-random range errors, which can cause problems in mobile robot feature detection and matching algorithms. To introduce this section, a summary of previous literature which outlines some of the causes of erroneous range data in lidar sensors is given. An equally important issue which is addressed is that of ambiguous or unreliable data and this section highlights the causes of ambiguous or multiple path reflections within a range scan by considering the physics of the AMCW phase detection process when the transmitted light beam is split between objects of differing range and/or surface reflectivity.

Section 10.6 uses the analysis to derive a correct calibration method for an AMCW lidar. Provided both the range estimate and the signal amplitude are available, the uncertainty of each range reading can be determined. During this derivation, it will be shown that the naive determination of the sensor to target distance as a function of a lidar's output range signal, in general provides a false calibration, a feature not recognised in the practise
of many mobile robotic researchers.

For real-time navigation, the speed at which an optical beam can be scanned and hence independent range samples recorded must be addressed and this is the subject of section 10.7. By modelling the phase locked loop (PLL), a tool often used for phase measurement, as a closed loop control system, an upper limit for the scanning speed and temporally uncorrelated range data acquisition rate is derived.

Finally it is the author’s belief that models and theories should be tested in practice and section 10.8 shows and discusses some results produced by a prototype 3D scanning AMCW lidar.

10.2 Critical Lidar Design Factors

Whichever of the range measurement methods, discussed in chapter 9, is used (TOF, FMCW, AMCW), any lidar system needs to detect the signal it transmitted reliably for accurate range estimation. In view of this, this section presents the analytical relationships between any lidar’s physical parameters such as the power of the transmitter LED/LASER (giving rise to eye safety issues); the sensitivity of the receiver (increasing the cost of the lidar) and the receiver aperture area (and hence overall size of the lidar).

When considering reflection from distant targets, Nitzan et al. presented a calculation of the received power expected at the receiver, based on Lambert’s cosine law [Nitzan et al 77]. When incident upon an opaque surface, a light ray can under go specular reflection according to Fresnel’s laws, and/or diffuse reflection governed by Lambert’s cosine law. In practice, both occur simultaneously and, it is the diffuse component which dominates the range estimate for most indoor surfaces, and which is of interest in lidar design.

Equation 9.14 in chapter 9 showed that the received power resulting from a diffuse reflection is proportional to $R_d \cos \theta_o/t_o^2$, where $R_d$ is the diffuse reflectance ratio of the object being sensed, $\theta_o$ is the sensor axis to surface normal angle of incidence and $t_o$ is the sensor to target distance. Diffuse reflectance ratios can vary between approximately 0.02 for dark objects and almost 1.0 for white surfaces [Nitzan et al 77]. Further, if objects are to be visible to the sensor at incidence angles $0^\circ < \theta_o < 80^\circ$ (i.e. near tangential reflection) and for typical ranges $0.2 < t_o < 15.0$ m ($0.2$ m being being a typical path distance between the actual receiver
and an object touching the sensor housing), the received signal can have a
dynamic range of $1.620 \times 10^6 : 1$ or 124 dB. Nitzan states in his work that
this dynamic range is in practice even higher, since at near normal angles
of incidence, the specular component of the reflection can be detected, since
this component is then reflected in a near coaxial manner with respect to
the transmitted light beam [Nitzan et al 77]. Brownlow, however, states
that at close ranges, the effective field of view of the receiver must be taken
into account and that the received power, predicted by Lambert’s cosine
law, is greatly over estimated, due to the inverse square relationship with
the range $t_o$. In reality therefore, considering diffuse reflections only, the
dynamic range will be less than the above estimate.

Applying a worst case analysis to equation 9.14, it is possible to enter
'pessimistic' values for $\theta_o$, $R_d$ and $t_o$ and see which 'undetermined' variables
remain. Assuming a worst case combination: $\theta_o = 80^\circ$, $R_d = 0.02$ and
$t_o = 15m$, it can be shown that the received power due to diffuse reflection
$\phi_d$ is [Adams 99]:

$$\phi_d = 4.913 \times 10^{-6} \eta A_R P_T$$  \hspace{1cm} (10.1)

where $\eta$ is a constant containing the transmittance of the system, $\tau$, (defined
in chapter 9) and the efficiency of the receiver diode. $P_T$ is the power
transmitted and $A_R$ the area of the receiver.

Equation 10.1 is fundamental to the design of the lidar. It places in
question the correct starting point for the design. For example, if an APD
is used and the aim is to operate it at a low d.c. bias voltage (thus reducing
avalanche multiplication noise and increasing its temperature stability), its
responsivity will be reduced, meaning that a higher value for $\phi_d$ in equation
10.1 will be necessary to produce a detectable photo-current. This implies
that the area of the photo-receiver, $A_R$ and/or the transmitter power, $P_T$
should be increased.

From an alternative stand point, one could argue that the area $A_R$
dominates the size of the overall sensor, and that this parameter should
therefore be chosen first, and the APD's d.c. bias, and hence responsivity,
set accordingly, so that the resulting received power $\phi_d$ produces enough
current for detection by the receiver electronics.

As with many engineering designs, a decision has to be taken so that
the approximate value of a particular value is realistically chosen, and the
other variables in the design are forced to follow suit (or in this case obey
equation 10.1). In the analysis here, a receiver aperture using a commercially available Fresnel lens was used, having an area \( A_R = 8.17 \times 10^{-3} \) m\(^2\). The total radiant power of the transmitter LED, was 1.8 mW, and the chosen APD has an efficiency \( \eta = 0.85 \) at a wavelength \( \lambda = 900 \) nm. The result therefore from equation 10.1 is that the ‘worst case’ received mean power \( \phi_d = 6.142 \times 10^{-11} \) W.

In the design presented here, the selected responsivity of the APD is 65 A/W. This yields a minimum mean signal current \( I_{min} (\text{rms}) \):
\[
I_{min} (\text{rms}) = 65.0 \times 6.142 \times 10^{-11} = 4.00 \text{nA} \tag{10.2}
\]

**10.3 Performance Limits — Noise**

For the above design specifications to be met, the worst case detected current, \( I_{min} (\text{rms}) \) defined in equation 10.2 must be detectable amidst all electronic noise sources. Therefore the sources of noise and their possible reduction within the sensor’s receiver must be addressed. The total noise current is primarily caused by four effects, the detailed calculations and analysis of which can be found in [Adams 99]:

1. A *shot noise* component as a result of the photo receiver’s dark current, \( \hat{i}_{\text{dark-shot}} \).
2. A noise current source due to *avalanche multiplication* (if an APD is used), \( \hat{i}_{\text{apd}} \).
3. A *shot noise* component due to back ground illumination, \( \hat{i}_{\text{bg-shot}} \).
4. A *shot noise* component due to the induced signal current itself, \( \hat{i}_{\text{rec-shot}} \).

The total RMS noise current is then:
\[
\hat{i}_{\text{total (RMS)}} = \sqrt{\hat{i}_{\text{dark-shot}}^2 + \hat{i}_{\text{apd}}^2 + \hat{i}_{\text{bg-shot}}^2 + \hat{i}_{\text{rec-shot}}^2} \tag{10.3}
\]

It now remains to determine the minimum signal current amplitude which needs to be detected and selected from the APD, and ensure that this is much larger than the RMS total noise current defined in equation 10.3. This gives rise to a further question: “How high does the signal to noise ratio need to be for reliable range estimation?” By estimating the nature of the probability distribution of the phase (and hence range) estimate, Brownlow derived an expression for the probability that the error
in a given range measurement is less than a predefined value [Brownlow 93].
As would be expected, this probability value increases dramatically with
increasing signal to noise ratio, and indeed for a 10 MHz modulation index,
it can be shown that to achieve 99% confidence that all range measurements
are within a tolerance of 1 of the maximum range, a minimum signal to
noise ratio of 30 dB is necessary [Adams 99].
In [Adams 99] we show that substituting each individual noise current
estimate into equation 10.3 gives the result:
\[
\hat{I}_{\text{total (RMS)}} = \sqrt{K B + 2qI_{\text{rec}}B}
\] (10.4)
where \(K\) is the mean square noise current per Hz due to the dark current,
back ground illumination and avalanche multiplication and is \(1.485^{-24} A^2/Hz\)
in this case. For a signal to noise ratio of 30dB, the mean received current
must be at least 32 times larger than \(\hat{I}_{\text{total (RMS)}}\) so that:
\[
I_{\text{rec}} \geq 32\sqrt{K B + 2qI_{\text{rec}}B}
\] (10.5)
The maximum quadratic solution for \(I_{\text{rec}}\), which turns the above inequality
into an exact equality, corresponds to the minimum received current neces­
sary to attain the above defined confidence in the range estimate. It is clear
from equation 10.4 that the noise current is significantly reduced by min­
imising the receiver’s bandwidth \(B\). An ideal choice of receiver is therefore
a simple tuned resonant circuit as used in AM radios, with a resonant peak
at 10 MHz [Brownlow 93]. This value must coincide with the minimum
necessary detectable current from the design specifications (4.0nA here).
Hence, it is therefore necessary to proceed with the receiver analysis by:

1. adjusting the design parameters (receiver aperture size and sensi­
tivity, transmitter power), or sensor specifications (maximum range
requirement), such that the minimum detectable current (4.0nA)
is greater than \(I_{\text{rec}}\) this being the solution to inequality 10.5.
2. constructing a low bandwidth receiver capable of selecting this sig­
nal [Brownlow 93].

10.4 AMCW Lidar Modules

The design and noise analysis thus far has made no assumptions on the
range measurement technique used (TOF, FMCW or AMCW). It applies
to any coaxial active range finder. To focus on actual implementations and
results in detail, the chapter now proceeds to examine only AMCW lidars.

The implementation of a usable AMCW ranging system can be con-
sidered to consist of six major components, which are summarised in the
following list.

(1) The transmitter module: An optical transmission system, capable
of transferring a high frequency (10 MHz here) current signal into
an optical signal which is a faithful replica of the current signal, is
necessary.

(2) The receiver module: A photo-detector and current amplifier ca-
cable of selecting the component of the transmitted signal received
after reflection, is required. The amplifier must be capable of a very
high gain, for the detection of signals from poor reflectors and/or
distant targets, whilst maintaining a high noise rejection capability.

(3) The signal compression unit: The received signal can theoretically
have a very large dynamic range, (greater than 120 dB as shown in
section 10.2) meaning that further amplification of this signal must
be handled with care. Indeed large signals resulting from close or
highly reflecting targets may not need further amplification at all,
whereas weak signals may still require amplification before they can
be of use for relative phase measurement. This unit must therefore
be able to compress the dynamic range of the received signal, such
that it can be of use for further range processing, given any am-
plitude of received signal, within the sensor’s design specifications.
It must further maintain a constant (or preferably zero) relative
phase shift between its input and output [Adams 99].

(4) The relative phase discriminator: The AMCW range estimation
technique produces a range estimate based upon the relative phase
shift between the received and transmitted signals. The phase dis-
criminator module has two inputs, firstly the reference signal, de-
verted directly from the transmitter, and secondly the compressed
received signal, output from the signal compression unit. An output
must be generated which contains the relative phase information,
from 0° to 360° between the two input signals.

(5) The output filter stage: This module is required to convert the rela-
tive phase information into a band limited analogue output voltage.
The upper frequency limit of this signal, along with the mechanical
scanning speed, determines the rate at which new range information can be recorded from the lidar.

(6) **The received signal envelope detector module:** A second useful measure which can be provided by a lidar sensor is the *amplitude* of the received signal, responsible for the current range estimate. This module must be sensitive enough to produce a unique amplitude estimate over the entire dynamic range of the received signal acceptable under the design specifications.

Information describing the electronic and mechanical modules within various complete lidar systems is often in the form of technical reports, examples of which can be found in [Jarvis 83; Johnston 73; Krotkov 90; Kweon 91; Miller and Wagner 87b]. Further articles which address particular aspects of the electronics which can be applied to lidar design, can be found in [Fjarlie 77; Hamstra and Wendland 72; Hopkins 84; Radeka 91; Rokos 73] and many of these articles, and others will be referred to in their relevant sections within this chapter.

A point noted in many of the above articles, is that the subsequent processing of the transmitted reference and received signals becomes an easier task if their frequencies are reduced. In response to this, one more stage within the lidar design, should be added to the above 6 components:

(7) **The frequency reduction stage:** This stage is designed to mix both the received, dynamically compressed signal, and the transmitter's reference signal, with a common local signal generator. The mixed outputs contain signals of a much lower frequency than the transmitted and received signals, but which bear the same relative phase information. These signals are then used as sources for modules 4, 5 and 6 above.

The interconnection of each of the above modules to form a fully functional lidar, is summarised in figure 10.1.

### 10.5 Causes of, and Remedies for, Range Errors

#### 10.5.1 Systematic Range Errors

In most lidar systems, systematic range errors are reported to be of greater concern than random errors as suggested by Hebert and Krotkov [Hebert
Causes of, and Remedies for, Range Errors

Fig. 10.1: The interconnection of each functional module for AMCW Range Estimation
and Krotkov 91]. They present four causes of false range data, namely mixed pixels, crosstalk, distortion due to scanning and range drift, which are summarised below:

**Mixed Pixels:** All laser/light transmission systems produce a light beam with a finite cross section which, if circular, in general projects an ellipse on to the target surface which it illuminates. This is sometimes referred to as the footprint of the beam. Every infinitesimal point within this footprint produces its own range and intensity value and the net result of this is a single range/intensity value at the sensor’s output. Hebert and Krotkov further note the problem which occurs when this footprint illuminates a range discontinuity, meaning that reflected energy is received from two surfaces separated by a large distance. The resulting range and intensity outputs from the sensor are false under these circumstances since the range output does not correspond to either of the two objects. This phenomenon is denoted by Hebert and Krotkov as that of mixed pixels, and they state that the occluding edges of scene objects are unreliable and that phantom objects may appear which do not represent either object within the light beam. They further state “This is a problem inherent to direct detection AMCW laser radars, and it cannot be completely eliminated.” [Hebert and Krotkov 91].

**Crosstalk:** Hebert and Krotkov define two causes of range/intensity crosstalk. The first is the fundamental relationship between the range variance and the measured light intensity, derived in [Adams 99]. The second cause is from the small changes in the mean value of the range output voltage, as a function of the received signal amplitude. Both of these problems will be addressed in this section and their solutions are considered as part of the quantitative model of the sensor.

**Distortion due to Scanning:** To record 3D range images, at least one rotating and a nodding mechanism is necessary. The error which can result here is that a correct range value \( t_0 \) can be recorded at falsely interpreted mirror angles \( \theta \) and \( \phi \), meaning that the resulting sensed coordinate \((r, \theta, \phi)\) is false. The main cause of this problem is synchronisation, since the rotating and/or nodding mirror, and the sampling of the range measuring system must be synchronised exactly. Another factor which contributes to this problem is the sampling technique employed by the processor and the electronic timing constraints imposed by the sensor’s electronics.
**Range Drift:** A problem noted with many of the commercially available lidars is that of significant range drift over time. Experiments were carried out by Hebert and Krotkov with the Perceptron sensor, whereby a target was placed at a fixed distance from the sensor (6 metres) and a range value was recorded each minute for 24 hours. During the first 3 hours of operation, the range drifted enormously, by an amount equivalent to the target moving one metre, after which the range output stabilised. Indeed the detailed analysis into the design of the electronics as given in [Brownlow 93; Adams 99] seems to be justified, since poor temperature and amplitude compensation can indeed dominate the sources of error in lidar sensors.

Other problems with data recorded from lidar sensors have been noted by Carmer and Peterson [Carmer et al 96]. They explain and quantify the effect of *speckle*, which produces a grainy image quality caused by the random interference of light waves from a diffusely reflecting object, when illuminated by light with temporal coherence [Dainty 75]. This problem presents itself more so in AMCW heterodyne systems, and less so in direct pulse detection systems. During 3D image production however, speckle degrades the intensity image but, under reasonable signal to noise ratios, has only a very minor effect upon the range image.

Typical systematic errors which can occur in AMCW lidars are demonstrated in figure 10.2.

The left plan shows a simple line model of the environment surrounding the sensor (located at the centre of the triangle). The right scan was recorded from a commercially available AMCW lidar sensor [Miller and Wagner 87b]. In order to demonstrate the effect of the large dynamic range of the received signal, a dark green piece of paper was mounted on the upper pillar between B and C. The remaining parts of the pillar between A and B and between C and D were white. Due to the differing amplitudes of the received signals from each part of the pillar, a clear systematic range error has occurred. This indicates the necessity for some form of reliable amplitude control of the received signal, to ensure linear range estimation throughout the entire specified dynamic range of the received signal, which was shown in section 10.2 to be higher than 120 dB in the design specifications presented.
Fig. 10.2 The left plan shows a simple map of a laboratory environment surrounding a lidar at the centre of the triangle shown. The right plot shows a single 360° scan. Only the unadjusted range data is shown, and each data point is represented as a cross. Curved regions such as FG correspond to out of range depth readings, and are shown as crosses at 2.5 metres from the centre of the mobile robot.

10.5.2 Random Range Errors

Although the receiver noise sources were presented in section 10.3, the propagation of noise through to the range estimate, was not quantified. This is now necessary for lidar calibration purposes, allowing the sensor to provide quantified range uncertainty information for robot state estimation based navigational algorithms [Adams 98].

The photo-diode acts as a current source which produces a time varying current at the frequency of the modulating signal. However, as shown in section 10.3, a noise current $i_{total}$ (RMS) is also produced. The noise in the amplitude of the received signal is not directly of interest in AMCW measurement systems, since the range estimate arises from the relative phase between the zero crossings on the $\omega t$ axis of the received and transmitted signals. Based upon a simple analysis of zero crossing uncertainty, the resulting range variance $\sigma_r^2$ (m²), varies with the received signal amplitude $V_r$ (volts) according to the following equation [Adams 99; Miller and Wagner 87b]:

$$\sigma_r^2 \approx \left( \frac{\lambda \sigma_n}{4\pi} \right)^2 \left( \frac{1}{V_r} \right)^2 + \sigma_e^2$$

(10.6)

where $\lambda$ is the modulation wavelength (30m in the design example here), $\sigma_n^2$
(volts²) is the combined constant variance of the electronic noise sources (quantified in section 10.3), and \( \sigma_e^2 \) (m²) is the additive electronic noise range variance which results after the amplification, mixing and phase comparison stages.

### 10.5.3 Multiple Path Reflections

It can also be seen in figure 10.2 that spurious ranges present themselves somewhat differently at each pillar edge, a problem which can cause range edge detectors to fail [Hebert and Krotkov 91; Adams and Probert 96]. This occurs because the range estimate, at the sensor’s output, results from a combination of the reflectivities, beam to target angles of incidence and ranges from all objects intersecting the projected optical footprint (known as multiple path effects [Adams and Probert 96; Skolnik 62]), and any “ghost” or internal leakage* path within the sensor. Depending on the particular optical footprint position, when the range output is sampled, the resulting range estimate can vary tremendously, as [Adams 99] shows that the received signal amplitude is considerably weakened. If the sample was taken when one of the artifacts such as D or F dominated (figure 10.2, a range estimate geometrically between these artifacts (such as E) results. Alternatively, if the ghost signal is higher than that produced by any artifact within the footprint, at the time of sampling, points closer in range, such as those on the right side of the lower pillar can result.

In order to identify spurious data points, the effects of simultaneous reflection from two targets warrants further investigation. In the paper by Hebert and Krotkov [Hebert and Krotkov 91], an AMCW lidar is used to form two dimensional pixel images. In this work the above effect is noted and referred to as the phenomenon of mixed pixels. To determine an algorithm for identifying these points, the physics involved when an optical foot print is split between two targets is necessary. This is given below, whilst the derivation of the detector can be found in [Adams and Probert 96].

Consider a transmitted reference signal \( V_0 \cos \omega t \) which is incident upon an edge (figure 10.3). An area \( A_1 \) is illuminated on the closer of the two surfaces returning a signal \( V_1 \cos(\omega t + \phi_1) \), whilst an area \( A_2 \) is illumin-

*This results due to either direct electronic cross talk between the transmitter and receiver or an optical path which exists directly between the transmitter and receiver. The detection and removal of these points is covered in [Adams 99; Hebert and Krotkov 91]
Fig. 10.3 The transmitted signal is split into two returned signals of differing phase by an edge. As the beam traverses the edge the illuminated areas and hence returned signal amplitudes will vary with traversal time.

The transmitted signal is split into two returned signals of differing phase by an edge. As the beam traverses the edge the illuminated areas and hence returned signal amplitudes will vary with traversal time.

The signal returned to the sensor will actually be the result of many modulated signals \( \sum_{i=1}^{n} V_i \cos(\omega t + \phi_i) \), each being emitted from a small area \( \delta A_i \) within the infra-red beam cross section. For a small beam cross sectional area the analysis is simplified if it is assumed that during the time the beam traverses the edge, \( \phi_1 \) and \( \phi_2 \) remain constant and that \( V_1 \) and \( V_2 \) change only due to changes in \( A_1 \) and \( A_2 \). Changes in \( \phi \) as the beam moves across areas \( A_1 \) or \( A_2 \) individually are therefore assumed to be negligible. Hence the returned signal \( Y \) is given by:

\[
Y = V_1 \cos(\omega t + \phi_1) + V_2 \cos(\omega t + \phi_2) \tag{10.7}
\]

so that:

\[
Y = [V_1 \cos \phi_1 + V_2 \cos \phi_2] \cos \omega t - [V_1 \sin \phi_1 + V_2 \sin \phi_2] \sin \omega t \tag{10.8}
\]
which can be written as a single sinusoid:

\[ Y = V \cos \phi \cos \omega t - V \sin \phi \sin \omega t = V \cos(\omega t + \phi) \]  

(10.9)

which is the form that is estimated at the sensor outputs — i.e. \( V \) is the output signal strength produced by both targets and \( \phi \) the resulting phase shift.

From equations 10.8 and 10.9 it can be seen that:

\[ \cos \phi = \frac{V_1}{V} \cos \phi_1 + \frac{V_2}{V} \cos \phi_2 \]  

(10.10)

and:

\[ \sin \phi = \frac{V_1}{V} \sin \phi_1 + \frac{V_2}{V} \sin \phi_2 \]  

(10.11)

giving two simultaneous equations in \( V \) and \( \phi \). Eliminating \( \phi \) gives:

\[ V^2 = V_1^2 + 2V_1V_2 \cos(\phi_1 - \phi_2) + V_2^2 \]  

(10.12)

Before proceeding further the relationship between each returned voltage, the sensor to target range and the illuminated area must be determined as generally as possible. The assumption is made that the emitted power is uniformly spread over the cross sectional area of the beam. Therefore:

\[ V_{1,2} = K_{1,2} \frac{A_{1,2}}{F(R_{1,2})} \]  

(10.13)

where the subscripts 1 and 2 refer the reflected signals 1 and 2, \( K_{1,2} \) are constants for each surface and incorporate surface reflectances and beam to target angles of incidence, and \( F(R_{1,2}) \) represents a function of the sensor to target ranges \( R_{1,2} \).

To establish a relationship between \( K_1 \) and \( K_2 \) in equations 10.13, the magnitude of the returned signal strengths when each surface is illuminated independently must be considered. These are denoted as \( V_{e1} \) and \( V_{e2} \). These are referred to as end conditions and in general:

\[ V_{e2} = \eta V_{e1}, \quad \frac{K_1}{K_2} = \frac{V_{e1} F(R_1)}{V_{e2} F(R_2)} \]  

(10.14)

Hence substituting equations 10.13 and 10.14 into equation 10.12 gives:

\[ V^2 = \frac{K_1^2}{[F(R_1)]^2} [A_1^2 + 2\eta A_1 A_2 \cos(\phi_1 - \phi_2) + \eta^2 A_2^2] \]  

(10.15)
Substitution of this function along with equations 10.13 and 10.14 into either of equations 10.10 or 10.11 gives the phase shift estimate $\phi$ as a function of time, as the beam crosses the edge.

When the beam is equally divided between the two surfaces, $A_1 = A_2 = \left(\pi b^2 / 2\right)$, where $b$ is the optical beam radius, thus giving:

$$V = \frac{K_1 b^2 \pi}{2F(R_1)} \left[(1 + \eta^2) + 2\eta \cos(\phi_1 - \phi_2)\right]^{1/2} \quad (10.16)$$

which is the returned signal amplitude. Substitution into equation 10.10 then yields:

$$\cos \phi = \frac{\cos \phi_1 + \eta \cos \phi_2}{\left[(1 + \eta^2) + 2\eta \cos(\phi_1 - \phi_2)\right]^{1/2}} \quad (10.17)$$

For the particular case when $\eta = 1$, i.e. when each surface individually returns the same signal strength:

$$V = \frac{K_1 b^2 \pi}{F(R_1)} \cos \left(\frac{\phi_1 - \phi_2}{2}\right) \quad (10.18)$$

and:

$$\phi = \frac{\phi_1 + \phi_2}{2} \quad (10.19)$$

When the beam is split into two equal areas on two different surfaces which, when illuminated individually return the same signal strength, the range estimate is the average of the two individual ranges.

The above theory can be used for the detection of range readings such as points E and F in figure 10.2 [Adams and Probert 96; Adams 99]

**10.6 Correct Calibration Procedures**

The tools have now been presented which are necessary for correctly interpreting lidar data, and calibrating an AMCW lidar for recording range. This section considers the necessary procedures for determining the three relationships necessary to provide a full calibration of an AMCW lidar, namely: **Calibration 1**: the output range voltage versus actual range; **Calibration 2**: the erroneously, internally induced, electronic phase (and
hence range) shift versus received signal amplitude\(^\dagger\), and **Calibration 3:** the range variance versus received signal amplitude.

**Calibration 1:** To eliminate the varying effect of calibration 2, it is essential that when initially calibrating range-voltage versus range, the returned signal strength is held constant, by using, for example, different coloured targets. This crucial factor is often over-looked by mobile robotics researchers, employing lidar sensors. The left graph in figure 10.4 shows an initial calibration of range-voltage versus actual sensor to target distance. This curve will only approach linearity, if optical and electronic leakage between the transmitter and receiver is minimised [Adams 99]. This graph offers a correct range-voltage versus actual range calibration for a target at any range from the sensor, provided it returns the signal amplitude at which the calibration took place. Initially, this calibration seems useless as, in general, targets will of course return varying received signal amplitudes within the lidar. This is the reason that calibration 2 is necessary.

**Calibration 2:** The data points in the right graph of figure 10.4 show the error caused by the amplifiers within the receiver circuit relative to the initial calibration in the left graph. It is interesting to note that various

\(^\dagger\)Although this effect can be reduced by dynamic range compression [Adams 99], it will not be completely eliminated, meaning that calibration 2 is in general still necessary.
combinations of target reflectance; orientation of target normal relative to the emitted light beam, and sensor to target distance will affect the returned signal strength [Allegre 91; Okada and Rembold 91]. Experiment shows that the factors which affect the returned signal strength are irrelevant as far as modelling the sensor is concerned and it is only the returned signal strength itself which is of importance. An analytical model for the right curve is not necessary here and would provide no general insight into the problem, as similar sensors exist which use other circuits before phase detection [Allegre 91; Miller and Wagner 87b; Maxwell 66; Nitzan et al 77]. This calibration is essential for correcting the range output of a lidar, and to our knowledge is often omitted by mobile robotics researchers, which causes large mapping errors when weak or strong received signals are encountered from the scanned environment.

**Calibration 3:** To establish the range variance as a function of the received signal amplitude, 10,000 independent range measurements were made of fixed targets with the sensor stationary. This must be carried out, whilst adhering to certain sampling time constraints [Adams 99]. The histograms in figure 10.5 have horizontal axes showing the measured range $r$, produced from the left calibration curve of figure 10.4, and vertical axes showing the number density. Note that the distributions are normalised, since the sum of the heights of all the range measurements is constant (10,000 in this case). All of the histograms in figure 10.5 were produced from different targets at a fixed range (3.0 metres) from the sensor. As expected, different signal strength values correspond to different variances within the range values. Note that the distributions are approximately Gaussian\(^4\). Figure 10.5 also shows the changes in the sample mean of the ranges for different signal strengths, which must be compensated for by calibration 2.

These results can be used to determine the unknown constants $\sigma_n$ and $\sigma_e$ in equation 10.6 and hence the numerical relationship between $\sigma_r^2$ and $V_r$\(^5\). The use of calibration 3 is demonstrated in figure 10.6 where the left

\(^4\)For very weak received signals, it can be shown that the distributions are Rayleigh in form [Connor 82; Adams 99].

\(^5\)Note that a TOF lidar will also produce randomly distributed range estimates but the analysis should be based upon the finite rise time of the received pulse as a function of the received signal intensity. In general TOF lidars suffer more than their AMCW counterparts if the received signal is weak as false detection, or no detection at all can result. An AMCW lidar will produce a noisy but consistent range estimate, assuming
Correct Calibration Procedures

Fig. 10.5  Histograms showing the effect of different coloured targets at a given range. All targets were at a true range of 3.00 meters from the sensor. The signal strength values $V_r$, sample means $\mu$ and range standard deviations $\sigma_r$ are shown with each graph. The continuous curves show calculated Gaussian distributions with the same mean and variance as the discrete data.

scan shows the amplitude of the received signal in polar coordinates (radial axis in volts) as a function of the scanning angle (angular coordinate in degrees), and the right scan shows the standard deviation in range (plotted in Cartesian form). The right scan shows lines of length $2\sigma_r$, calculated from the received amplitude in the left scan and equation 10.6, centred on the actual range estimates themselves. The dotted line represents a plan view of the actual environment. The length of the line segments in figure 10.6 gives a quantitative assessment of the uncertainty associated with each range estimate, particularly useful for weighting the influence of each range value for feature detection purposes [Adams 99].

correct calibration [Vuylsteke PO90].
Fig. 10.6 Left: Signal amplitude (radial axis (volts)) versus sensor scanning angle (degrees). Right: Lines of length $2\sigma_r$ centred on the range estimates. The triangle in the right graph shows the position of the mobile robot. The dotted line represents a plan view of the actual environment.

### 10.7 Possible Scanning Speed

This section provides the lidar user with a tool for deriving the speed at which range information can be faithfully sampled from a lidar. In any lidar system, this derivation requires a model of the range estimation electronics, and again here, reference will be made to the phase estimation electronics within the AMCW process.

A reliable tool for producing square waves locked in phase to almost any noisy periodic input signal is the *phase-locked loop* (PLL) (see figure 10.7). To derive the range sampling constraints, the PLL must be examined. A PLL can be modelled as a phase comparator, which produces a digital signal representing the phase difference between its two input signals. This signal is then low pass filtered, and used as an input to a *voltage controlled oscillator* (VCO) which in turn can produce a square wave with frequency proportional to the input voltage [Horowitz and Hill 93]. This can be used in a closed loop system (as shown in figure 10.7) to produce a clean, locally produced signal, with the same frequency as the noisy input signal (derived from the sensor's amplified received signal) and a constant relative phase relationship.

To ensure that the PLL is able to track the dynamic phase variations of the received signal as the sensor scans, it is necessary to analyze the
low pass filter used in conjunction with the phase detector and VCO. The

\[
G(s) = \frac{1 + sT_2}{1 + s(T_1 + T_2)}
\]  \hspace{1cm} (10.20)

where \( s \) is the Laplace frequency variable and \( T_1 \) and \( T_2 \) are time constants, dependent on the components used in the lead-lag filter. If the gain of the phase comparator is \( K_p \) and that of the VCO is \( K_{vco}/s \) as shown in figure 10.7, the overall closed loop transfer function between the phase of the input sinusoid \( \phi_{in} \) and the output square wave \( \phi_{out} \) is given by:

\[
\frac{\phi_{out}}{\phi_{in}} = \frac{K_{vco}K_p(1 + sT_2)}{s^2(T_1 + T_2) + s(1 + T_2K_{vco}K_p) + K_{vco}K_p}
\]  \hspace{1cm} (10.21)

which produces a classical second order response to changes in the input phase \( \phi_{in} \), caused by actual range changes. The denominator of equation 10.21 can be arranged to have the standard form \( s^2 + 2\zeta\omega_n s + \omega_n^2 \) where \( \zeta \) is the damping factor and \( \omega_n \) the natural frequency of the response. With knowledge of the possible speed at which the input phase can change with respect to time (i.e. rate of range change) and the desired settling time for the locally produced VCO output square wave, values for \( \zeta \) and \( \omega_n \) can be calculated and implemented by choosing the correct components in the lead-lag low pass filter. The highest frequency changes in range
which need to be recorded correspond to a change in full range (15 metres) divided by the time necessary for the scanning mirror to rotate through the effective beam width of the light spot. The estimated beam-width of the lidar considered here, is 0.0067 rads, calculated by measuring the optical foot print diameter at maximum range. With a maximum mirror scanning speed of 2.5 revs/s for example, the time between independent range measurements is 0.43 ms. Within this time interval, it is necessary that all transient effects of the second order transfer function of equation 10.21 have reached an acceptable level. Brownlow defined this “acceptable level” as the time period $t = \frac{3\pi}{\omega_n}$ after which any overshoot has reduced to less than 1cm range error [Brownlow 93]. Applying this criterion, $\omega_n$ is approximately 22,000 rads/s, meaning that a bandwidth of almost 3.5kHz is necessary.

The range estimate, as a result of the frequency lock-in detection capability of the PLL, is demonstrated in figure 10.8 where the top graph shows the received (low amplitude, noisy wave) and reference signal from a target at 7.0 metres. The lower graph shows the square wave outputs from the two VCO's running on separate PLLs. The reference ‘square’ wave has been shifted vertically by 2.0 volts to clarify them. The subsequent processing necessary to produce an analogue output proportional to range, simply requires a suitable phase detection circuit with both of these square waves as inputs. To demonstrate the effect of the received noise, figure 10.8 shows the results recorded from the VCO outputs at 5 different time intervals, these being superimposed upon each other in the lower graph. The smaller, highly noise-corrupted sine wave in the top graph is the received signal. It can be seen in this graph that the time axis crossing of the received signal is ill defined due to its low signal to noise ratio. This effect reproduces itself at the outputs of the VCOs in the form of phase noise as shown in the lower graph of figure 10.8. Since this is a very weak signal it can be seen that a large phase uncertainty, and hence ultimately, range uncertainty results.

A “good” reflector placed 7 m away from the sensor, was used for the same experiment in figure 10.9. This time, the larger signal is the received signal in the upper graph, and once again the received signal’s VCO output was recorded at 5 instants in time relative to the reference VCO signal. It can be seen that the time axis crossings are more clearly defined and the phase noise is greatly reduced.
Fig. 10.8 Reduced frequency reference and received signals (top graph) and their corresponding VCO outputs (bottom graph) for a weakly reflecting target at 7.0 m used in earlier experiments. In the lower graph, the output of the received signal’s VCO was recorded at five different times.

An estimate of the range simply results from low pass filtering the output of the phase comparator⁷. Clearly this final block in the range estimation circuitry requires an optimally set cut-off frequency to yield the necessary measurement bandwidth, but at the same time minimise the high frequency phase noise at the output. Applying the same argument as above, the bandwidth of the low pass filter should be chosen to allow for the maximum possible range changes at the maximum mirror scanning speed (2.5 rev/s here). The results of low pass filtering the output of the phase comparator for various targets, again situated at a range of 7.0 m from the sensor, are shown in figure 10.10. The lower curve in each graph shows the amplitude estimate of the signal concerned. The graphs are scaled such that full range in each case corresponds to 0 to 5 volts, which for the range signals corresponds to 0.0 to 15.0 metres, and for the amplitude output corresponds

⁷Detailed methods for phase discrimination can be found in [Adams 99]
to the weakest and strongest received signals. The top graph shows the results obtained from a relatively good diffuse (white) reflector at 7.0 m. It can be seen that the amplitude measurement is only at 2.0 volts, however the 0 to 5.0 volt scale allows for very large signals produced by specular reflectors such as mirrors or shiny metal, which can still produce valid range estimates. Note that the range estimate, situated at 2.33 volts appears almost noise free and, due to the relatively strong signal amplitude gives a reliable, low noise range estimate. In the middle graph, the target is substituted with a darker material, again situated 7.0m from the sensor. The detected signal amplitude is approximately halved, whilst the range estimate remains at an average value of 2.33 volts, although slightly more noise is evident. The lower graph in figure 10.10 demonstrates the case of an extremely weak reflector (black felt), again situated 7.0m from the sensor. This time the amplitude of the received signal is only 1/20 of that of the
Fig. 10.10 Range and amplitude outputs for differing target reflectivities at a fixed range (7.0 m). The maximum range (15m) corresponds to a signal output of 5.0 volts good reflector, even after semi-logarithmic amplification by the dynamic range compression module. The range estimate is almost random over a large region, although it can be noted that given enough time to average the range output, a reasonable range estimate would still result.

10.8 3D Range/Amplitude Scanning — Results

To complete this chapter, some 3D sensor scans, recorded with a newly designed sensor, the design specifications of which were presented in section 10.2, are shown here. The calibration sequence of section 10.6 which is summarised by the two graphs in figure 10.4, were used to produce the range and intensity data. The 3D data can be represented in Cartesian form as in figure 10.11. In this representation one can clearly see the local environment surrounding the sensor (each point shown corresponds to a single range data point). On the lower wall of the corridor, were two swing
doors with glass panes. These can produce multiple path reflections in AMCW lidars.

Another representation of the same scan (this time with a person standing in front of the sensor) is shown in figure 10.12. The top scan in the figure shows the albedo values produced from the sensor, calculated from the amplitude of the received signal. This is considered in various texts to be similar to the quantity measured by the eyes. Mathematically, the albedo is calculated from the measured amplitude $\times$ (measured range)$^2$, which gives the Lambertian reflectivity product $R_d \cos \theta_o$, where $R_d$ is the surface reflectivity and $\theta_o$ the light beam to surface normal angle of incidence. Each pixel in the top scan of figure 10.12 is illuminated proportional to the albedo of each point in the environment. The top row of pixels corresponds to a single 360° revolution of the sensor's mirror. The mirror is then automatically tilted downwards about its horizontal axis through 0.5° and the next row of pixel values is recorded. This process is repeated 70 times to give the full "wrap around geometry" scan shown. The similarity to a black and white photo of the full environment is evident.

The second image in figure 10.12 shows the result of illuminating each
pixel proportional to the range output generated during the same scan. The
darker the pixel, the closer the object is to the sensor. This range image
is important for navigation, as a robot can immediately see which objects
are in its way. Note that one no longer sees the posters on the walls since
they are equidistant from the sensor as the walls themselves. This shows
the linearity of the sensor as it is able to produce accurate range estimates
for almost all reflectivities. The lower image shows each pixel illuminated
proportional to the standard deviation of each range reading, calculated
from equation 10.6. The darker pixels in this image correspond to range
values in the middle image which have a low standard deviation and can be
considered more trustworthy as those which produce brighter pixels in the
lower image. The three images in figure 10.13 show the same type of data
but in a more complex environment. Notice that in the top albedo scan
one can distinguish the monitor frame from the screen (in the right hand
part of the scan) and one can see the poster (and text) on the wall in the
left hand part of the scan along with the cement between the bricks in the
wall itself. In the range image (middle) the monitor frame and glass screen
are indistinguishable as is the poster and the cement from the wall since
they are equidistant from the sensor. Once again the lower image shows
the standard deviation of each recorded range value.

10.9 Summary

The analysis of range and amplitude data production at the electronic hard-
ware level indeed provides a useful tool for modelling the expected perform-
ance of any range measuring sensor. An explanation of the commonly
used lidar range measuring techniques was presented along with applica-
tions employing each method. Irrespective of the measurement method
used, it was demonstrated that the minimum detectable current induced in
the receiver, for a worst case target, can be derived. The limiting factors, in
the form of noise, attributed to the receiver electronics, which oppose the
reliable detection of this current must be addressed. A theoretical quan-
tification relating the power of the transmitter (and hence eye safety); the
sensitivity of the receiver (and hence sensor cost); the receiver aperture
area (and hence sensor size) and the receiver bandwidth (and hence pos-
sible range sampling and scanning speed) can give the user an informed
choice of necessary sensor for a given application.
Fig. 10.12 Each pixel map shows an image with resolution 792 x 70 (i.e. a total of 55440 range/intensity values). The top scan shows the albedo image recorded in a corridor and is similar to a black and white photo showing the "wrap around geometry" of the environment. The brighter the pixel, the higher the reflectivity of the scanned point. The middle image shows each pixel value plotted proportional to range. (Darker pixels are closer to the sensor). Notice that the posters on the walls can no longer be distinguished from the walls since they are at the same distance from the sensor. The lower image shows each pixel value with brightness proportional to the standard deviation of each range value. In general closer and highly reflective objects produce range readings with a lower standard deviation and are shown as darker regions.

The factors which oppose the reliable detection of the reflected light signal were explained in terms of additive noise sources in the receiver. These noise sources were quantified to further define the performance limits of any lidar sensor.

By focusing on the AMCW ranging method, the individual modules
Fig. 10.13 Three further images also showing albedo (top), range (middle) and standard deviation (bottom), recorded from a single scan in a more complex environment

which comprise such a sensor were outlined in section 10.4, along with the critical design factors associated with them in terms of the effects of noise, distortion of the input signals and speed at which changes in range can be measured.

A solution for minimising systematic range distortion as a function of the received signal intensity was presented in section 10.5. For reliable sensor data manipulation in general, an invaluable quantity is the range variance. By considering the physics of noise propagation from the receiver to the output range estimate, a model was derived which produces a unique, provably correct range variance with each range value. An important point often overlooked by robotics researchers is that, in general, the naive determination of the output range voltage from AMCW or TOF lidars, as a function of the actual sensor to target range, provides a false calibration. The further issue of multiple path reflections which occur when the optical
foot print is divided by a change in range and/or reflectivity. It is possible to show that range values result which do not correspond to either of the actual ranges on both sides of the discontinuity. Methods for detecting these false range values were referenced.

Section 10.7 examined the speed at which range scans can be recorded. A simple second order system which models the phase response of PLL systems to actual range changes between the sensor and its target was used to derive the speed constraints of the sensor as a function of the transmitted optical beam-width.

Finally results of 3D wrap-around geometry scans were shown, recorded from a prototype scanning AMCW lidar, developed for autonomous robot navigational experiments. Provided the physics of the sensing method is understood, and the limits within which reliable data can be expected is known, AMCW lidars provide a very attractive and relatively cheap range data source for autonomous robot applications.
Chapter 11

Extracting Lines and Curves from Optoelectronic Range Data

Robert Taylor

In earlier chapters we talked about the difficulties of extracting parametric curves from sonar data and described feature extraction in terms of line and point primitives. In this chapter we describe methods to process optical data. Optical data provides a representation close to the physical surface structure as we have seen in the last chapter. Therefore a natural way to process it is to extract and parameterise surface curves.

This chapter describes how two parameterised curves: straight lines and ellipses can be extracted from optoelectronic range data. The application was developed for a nuclear processing plant, in which there was a complex structure of pipes. These may restrict access to some areas and get in the way of the line of sight in others. There are particular needs for sensing for mapping and for manipulator guidance in processing areas where there is radiation hazard, since these are not accessible to people except through observation windows. Current robotic systems rely on teleoperation, but this may be impeded by a poor field of view and the problems of delay and remote control.

Most vision feature extraction algorithms have been developed to process camera images, for which only amplitude data is available. Conventionally, processing goes through several stages: a high pass filter to find edges, a process of joining nearby edges into segments, and then higher level algorithms to find certain shapes (lines, circles etc). The segmentation process divides the image into the various different components (such as walls, pipes, process plant) which are distinguished through intensity differences. However intensity difference also occurs from shadowing and other lighting artefacts.

Having range data available immediately simplifies the processing, since changes in range are a more reliable indicator of different structural features. Stereo vision finds range through using two camera images, but
problems may arise from calibration difficulties (especially near the edges of the image) and registration (identifying the same object in the two views - again a problem when there are shadows, occlusion or clutter). * The availability of high quality range data makes segmentation much easier.

The building block for finding pipes is ellipse recognition. In this chapter we describe algorithms which process a two dimensional range map to extract segments containing lines and ellipses as the data is acquired. Although more reliable methods would be available using a three dimensional representation, the sensor then needs to be scanned in both x and y directions and the speed of data acquisition is therefore slower.

Provided that it works at an optical wavelength, the hardware of the sensor should be immaterial to the processing algorithms. In practice it affects aspects such as the noise in the images and the maximum range which can be determined. In addition of course the scanning mechanism has affect not only on accuracy but on the dimensionality of the data gathered and the speed with which it is obtained (which can be important when the robot is moving). In this chapter we show data acquired using two sensors: a low power (1mW) two dimensional triangulation sensor, and a high power (50mW) time of flight (AMCW) sensor.

11.1 The Optoelectronic Sensors

11.1.1 The Triangulation (LEP) Sensor

This section describes the triangulation sensor for range finding [Pears and Probert 96]. The emitter is a visible laser (850nm) and the detector is a Lateral Effect Photodiode (LEP), known commercially as a Position Sensitive Detector (PSD) [Hawthorne 89]. It consists of a one-dimensional continuous p-n junction, length P, with a terminal at each end. The position, p, of the incident laser light on the device is given by the normalised difference of the currents at the two terminals, $I_1$ and $I_2$:

$$ p = \frac{I_1 - I_2 P}{I_1 + I_2} $$

Hence range can be determined (sec. 9.4.1 of chapter 9). The sum of the currents gives the intensity of the incident beam, which is related to the

*Hence some advantages of active vision - chapter 12
uncertainty of the range measurement. This is useful for statistical analysis of the results.

A more common detector for triangulation is a charge coupled device (CCD) array. The commonest form of two dimensional array is the video or digital camera, but one dimensional arrays are available too. The CCD requires a more complicated interface and is slower for a spot measurement. In its favour, it is more sensitive and has a higher signal to noise ratio except at very short distance.

Figure 11.1 shows the layout of the sensor used. It follows the synchronised scanning mechanism of Livingstone and Rioux [Livingstone and Rioux 86] discussed in chapter 9. Light is projected from the laser onto the first mirror, a, onto the scanning mirror, b, and onto the projection mirror, c. Reflected light is collected by the detecting mirror, d, reflected onto the reverse of the scanning mirror, and focused by the lens onto the detector. The sensor uses a longitudinal synchronous scanning geometry: the field of view of the detector moves with the scanned beam, which has the effects of improving immunity to ambient light and reducing susceptibility to errors
in scan angle. The sensor has a range of approximately 0.5 – 2.5m, and a field of view of 40°, taking 256 range readings per scan. It can sample at 2.5kHz giving a 10Hz scan frequency.

It is reasonable to assume that the error in orientation is very small. The error in range increases non-linearly with range (a property of all triangulation sensors) and is about 2mm at 1m range. In processing we assume that all the error lies in the range measurement.

11.1.2 The SICK Sensor

As discussed in chapter 9 the company SICK Inc manufacturers a number of optical sensors for measurements and industrial use. The proximity laser scanner is becoming increasingly popular as an optical rangefinder in robotics. The model used had a range of up to 50m although the accuracy is only about 2-3 cm. As in the LEP sensor, it is reasonable to assume that range error dominates over angular error.

Range is determined through time of flight measurement. Details of operation are not available. In the outdoor versions a processing unit includes a software filter to remove effects from rain and snow when the sensor is used outdoors.

11.1.3 Perceptron Laser Scanner

Measurements were taken too with a 3 dimensional sensor, a lidar sensor, produced by Perceptron. This sensor works on the principle of laser radar
described in the last chapter.

The scanning unit consists of a coherent amplitude modulated laser light source projected onto a spinning polyhedron to scan the beam horizontally. Vertical scanning is achieved using a nodding mirror mounted in the upper part of the scanner. An avalanche photodiode is used as the photo-detector which receives reflected light via the same optical path as the transmitted light. The amplitude and the phase of the received signal are digitised and stored, providing perfectly registered range and intensity images. This is an accurate device, designed for inspection tasks. The laser is a class 3B, 50mW, 830nm unit held in an oven to maintain stability. This power is hazardous to the human eye. Perceptron claim a resolution of 200μm on surfaces with reflectivity between 18% and 80%, with an accuracy of 10 bits over the range. For a 10m head this represents an accuracy of approximately 9.7mm, for a 2m head 1.9mm.

The scanner has a maximum horizontal field of view 40° and a maximum vertical field of view of 30°.

11.2 Feature Extraction and Processing

These sensors were used to acquire data over their complete angular field. All work by scanning a spot, and so points come in sequentially. The objective of the feature extraction software is to acquire range data, and parameterise the scene in terms of straight lines and conic sections. These represent the features at the intersection of the plane of laser light with, say, walls and cylinders. Discontinuities (edges) can be detected, as well as curved surfaces, such as pipes.

To start the whole process off, the first few points were taken to lie on a straight line, \( y = mx + c \) and parameters determined as described in section (11.2.2). Each subsequent point acquired by the sensor was initially assumed to belong to the current line. If it did not - i.e. the error to the fit was large - it was placed in a bucket for later processing, along with any other points which could not be matched to a line. The size of the error was determined using a statistical measure as described in section 11.2.2. A new line was initialised. After all the points were acquired, those in the bucket were processed to see if they could be fitted to an ellipse. This process is shown in figure 11.3.
11.2.1 Kalman Filter for Straight Line Extraction

A Kalman filter algorithm was used to extract the straight lines. This was primarily for two reasons:

- it can be implemented recursively, as points are gathered
- it provides an estimate of the likelihood that a point will belong to the current line through the innovation (see section 11.2.4). This means that it can be used to segment data into lines and potential ellipses in real time.

Most algorithms for line fitting published in the literature describe routines to fit points in a Cartesian coordinate system, based on the equation of a straight line, \( y = mx + c \). The sensor, however provides range readings, \( r \), at different angles, \( \theta \). i.e. uses a polar coordinate system. Although it is straightforward to convert between the systems, the result is that the errors in \( x \) and \( y \) are correlated. Standard methods therefore do not return optimal results.

Instead a model was set up to process points in polar co-ordinates.
11.2.1.1 Extended Kalman Filter Equations

The extended Kalman filter was first set up using Cartesian co-ordinates. It uses a two stage process to update its estimate of the current state of a system. First it uses a model to predict the next state from the current one. Then it takes an observation of the state. These two estimates of state are combined to minimise the (least square) error given an estimate of the uncertainty in each prediction. Along with a combined estimate of state results an estimate of the covariance of the state estimate. The notation used here is taken from [Chang et al 86].

To use the Kalman filter we need to set up a model of the system, and we use the straight line model:

\[ y = mx + c \]  \hspace{1cm} (11.2)

We define the state: $^\dagger$

\[ u = \left( \begin{array}{c} c \\ m \end{array} \right) \]  \hspace{1cm} (11.3)

The Kalman filter requires two equations: one updating the state from a model based estimate and the other relating the observation to the state. Since the components of the state are constant for a straight line, the state update equation from sample $k$ to sample $k+1$ is:

\[ u_{k+1} = u_k + v_k \]  \hspace{1cm} (11.4)

where $v_k$ is a sequence of zero mean white Gaussian noise with covariance matrix $Q_k$. The components in $Q_k$ express the uncertainty in the model. They could, for example, be non-zero to express that a line need not be entirely straight. However they were set, more restrictively, at zero.

Let the observation be $z$ such that:

\[ z_{k+1} = H_{k+1} u_{k+1} + w_{k+1} \]  \hspace{1cm} (11.5)

where $w_{k+1}$ is a sequence of zero mean white Gaussian noise with covariance matrix $R_{k+1}$, representing the measurement noise, and $H_{k+1}$ is

$^\dagger$This algorithm differed from the standard use of the Kalman filter to track points on a line, since in this case it was the parameters of the line itself which were uncertain. The more usual implementation assumes the model parameters $(m,c)$ and modifies the positions of each point, $y$. 

the observation matrix. In this case the observation is of the y co-ordinate for a given \( x \), and \( w_{k+1} \) is a scalar, \( w_{k+1} \):

\[
y_{k+1} = \begin{pmatrix} 1 & x_{k+1} \end{pmatrix} \begin{pmatrix} c_{k+1} \\ m_{k+1} \end{pmatrix} + w_{k+1} \tag{11.6}
\]

and

\[
H_{k+1} = \begin{pmatrix} 1 & x_{k+1} \end{pmatrix} \tag{11.7}
\]

\( w_{k+1} \) is determined from the known errors in the sensor (from the total current, \( I_1 + I_2 \) in Eq. 11.1.

11.2.1.2 *Cartesian to Polar Co-ordinates*

The problem with this formulation is that there are errors in \( H_{k+1} \) from the errors in \( x \).

Therefore, to get an error in only one component (because of the scanning mechanism the errors in orientation are small), the equation of the line in \( \{x, y\} \) co-ordinates was converted into \( \{r, \theta\} \):

\[
x = r \cos \theta \\
y = r \sin \theta
\]

Substituting into Eqs. 11.6 and 11.7 we find that:

\[
r_{k+1} \sin \theta_{k+1} = \begin{pmatrix} 1 & r_{k+1} \cos \theta_{k+1} \end{pmatrix} \begin{pmatrix} c_{k+1} \\ m_{k+1} \end{pmatrix} \tag{11.8}
\]

and

\[
H_{k+1} = \begin{pmatrix} 1 & r_{k+1} \cos \theta_{k+1} \end{pmatrix} \tag{11.9}
\]

The \( H \) matrix now contains a term in \( r \), and therefore is still not known without error. The above equation, Eq. 11.8, can, however, be rearranged to give the following:

\[
\frac{1}{r_{k+1}} = \begin{pmatrix} \cos \theta_{k+1} & \sin \theta_{k+1} \end{pmatrix} \begin{pmatrix} \frac{1}{c_{k+1}} \\ \frac{m_{k+1}}{c_{k+1}} \end{pmatrix} \tag{11.10}
\]
which is now in the correct form. The noisy observation is $1/r$. The observation matrix now only contains terms in $\theta$ which is assumed to be without error.

### 11.2.2 Initialisation Phase

Using the above equations, we can find the best estimate of $(c \ m)^T$. It can be shown that [Chang et al 86]:

$$\hat{u}_k = [H'_k R_k^{-1} H_k]^{-1} H'_k R_k^{-1} z_k$$

(11.11)

where the notation $\hat{u}_k$ denotes the expectation of $u_k$, and $H'$ is the transpose of matrix $H$. Then, substituting from Eq. 11.8, and taking $R_k = \sigma_k^2$, we find:

$$\begin{pmatrix} \frac{1}{c_{k+1}} \\ \frac{m_{k+1}}{c_{k+1}} \end{pmatrix} = \frac{1}{\Delta} \left( \begin{array}{c} - \sum \sin^2 \theta \sum \frac{\cos \theta}{r} - \sum \cos \theta \sin \theta \sum \frac{\sin \theta}{r} \\ - \sum \cos \theta \sin \theta \sum \frac{\sin \theta}{r} + \sum \cos^2 \theta \sum \frac{\sin \theta}{r} \end{array} \right)$$

(11.12)

where

$$\Delta = \sum \cos^2 \theta \sum \sin^2 \theta - \sum (\cos \theta \sin \theta)^2$$

(11.13)

and the covariance is:

$$P_k = [H'_k R_k^{-1} H_k]^{-1}$$

(11.14)

These equations are used to initialise a new line from a set of points (typically five) \{r, \theta\}. The line model is then ready to be refined by the next points received.

### 11.2.3 Recursive Implementation

Once a line was initialised, the estimate was improved by other points coming in. The equations Eqs. 11.11 and 11.14 were reformulated to provide a recursive way to update the state estimate and covariance.

We define two terms:

$$S_{k+1} = H_{k+1} P_k H'_{k+1} + R_{k+1}$$

$$W_{k+1} = P_k H'_{k+1} S_{k+1}^{-1}$$
\( S \) is a quantity related to the expected covariance of the measurement. \( W \) is known as the Kalman gain. Then it can be shown [Chang et al 86] that, going from instant \( k \) to \( k + 1 \):

\[
P_{k+1} = P_k - W_{k+1} S_{k+1} W'_{k+1} \quad (11.15)
\]

\[
\hat{u}_{k+1} = \hat{u}_k + W_{k+1} [z_{k+1} - H_k \hat{u}_k] \quad (11.16)
\]

Note that the state update equation (Eq. 11.16) updates the current state using the difference between the actual measurement and the prediction of the measurement, weighted by the Kalman gain. The difference is known as the innovation. For the system above:

\[
v_{k+1} = z_{k+1} - H_{k+1} \hat{u}_k \quad (11.17)
\]

### 11.2.4 Feature Segmentation

Feature segmentation is implemented using this recursive filter to decide whether a newly acquired point lies on the current feature or is part of the next feature. This is done for straight line segments by looking at the normalised innovation squared, called the validation gate:

\[
q_{k+1} = v'_{k+1} S^{-1}_{k+1} v_{k+1} \quad (11.18)
\]

The magnitude of the validation gate can be related, through the \( \chi^2 \) distribution, to the probability that a point lies on the line. For example for a one dimensional measurement using \( q_{k+1} = 2 \) implies that there is a 95.4% probability that the point with a value within that gate lies on the line [Chang et al 86]. If the point lies outside the gate, an edge is flagged and the batch process restarted to initialise the next line segment. To detect curved surfaces, a parameter, \( p \) is introduced, which corresponds to the minimum number of points a straight segment must contain for it to be considered valid. If the validation gate exceeds its threshold before \( p \) points have been sampled in the current segment, the edge is rejected. The segment is flagged as invalid and stored in a temporary buffer. If, after re-initialisation, the above condition occurs again, the new invalid line segment is appended to the buffer. This process repeats until a valid straight
line segment is recognised, or the end of the scan is reached. At this point, the buffer of invalid segments is transmitted to the ellipse fitting algorithm described in the next section. This arrangement allows the line fitting process to fit straight segments and recognise areas that cannot be accurately be approximated by straight lines.

Figure 11.4 shows the results of this algorithm from a typical scan across planar surfaces placed between 1.25 - 2.25 m from the sensor. The sensor's intensity information is additionally used to detect the validity of received data. If the intensity is below a threshold, then the range information is too uncertain, an example being the out of range condition. In such cases, such as segments D and F, the data is ignored, and no line is fitted. It can be seen that the recognised segments (A, B, C, E) are less noisy than the rejected ones, and this allowed the validation gate threshold to be reduced for more accurate edge detection. The hit rate for such straight line segments was over 95% for a variety of segment lengths and orientations.

11.2.5 Elliptical Sections

The remainder of the data was passed through the ellipse detector. Only small sections of the ellipse are seen in the images. Fitting such small sections of data to an ellipse is a notoriously ill-conditioned problem, since
small changes in the data set produce very large changes in the fit which results. A further problem here is due to uncertainty over which points in the image actually lie on the ellipse and which on an edge.

There are a number of methods for ellipse detection reported in the literature [Bookstein 79; Porrill 88; Illingworth 89; Yoo and Sethi 93; Rosin 93; Sampson 82]. The Bookstein [Bookstein 79] algorithm was one of the first to be developed to fit to conic sections. An improved version which pre-normalised the data was implemented on this sensor. Sometimes it gave some good results but they were rather inconsistent. The problem is that it did not differentiate between different types of conic, i.e. parabolae, hyperbolae and ellipses, all of which have the same general equation

\[ F(x, y) = ax^2 + by^2 + cxy + dx^2 + ey^2 + f = 0 \]  \hspace{1cm} (11.19)

Often, due to noise, a parabola might indeed fit the data better than any ellipse using a standard least square approach. However the physical conditions imposed an ellipse as solution. This means that the conic must be bounded. To ensure that it is bounded we need to ensure there are no non-trivial zeros (which correspond to points at infinity).

Eq. 11.19 was reorganised into a leading form:

\[ ax^2 + bxy + cy^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \]  \hspace{1cm} (11.20)

Rewriting this we define matrix \( M \) such that:

\[ M = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \]  \hspace{1cm} (11.21)

Hence

\[ ax^2 + bxy + cy^2 = \begin{pmatrix} x & y \end{pmatrix} M \begin{pmatrix} x \\ y \end{pmatrix} \]  \hspace{1cm} (11.22)

A sufficient condition for this form to contain no non-trivial zero is that \( M \) is positive definite. This means that the matrix can be written as the square of some other symmetric matrix, \( N \):

\[ M = N^2 = \begin{pmatrix} A & B \\ B & C \end{pmatrix}^2 \]  \hspace{1cm} (11.23)

---

\(^1\)The author is grateful to Dr Alison Noble for guidance in this area
where we define A, B and C as the elements of N. We can rewrite equation 11.19
\[ F(x, y) = (A^2 + B^2)x^2 + 2(AB + BC)xy + (B^2 + C^2)y^2 + dx + ey + f = 0 \] (11.24)

This polynomial is then defined by the six parameters \{A, B, C, d, e, f\} and is guaranteed to be an ellipse.

This solution was termed the Bounded Conic Fit. The parameters were found by using a least squares fit to minimise:
\[ \sum |F(x, y)|^2 \] (11.25)
where the summation is over all points on the ellipse. The resulting equation was solved by an iterative non-linear regression routine, known as the Levenberg-Marquandt method [Press et al 84].

A problem we mentioned earlier was how to decide which points belonged to the ellipse. As we see by comparing figures 11.5 and 11.6 it is not always clear, and significant differences may arise in the best match if different set of points are taken. This problem was overcome by pre-processing each incoming point, to test its distance from its predecessor, and its intensity. Points at long distances from their neighbours or with low intensities were rejected.

Outliers, however, would still sometimes creep into the data set on some scans, and so a more robust estimation method was sought, which would reject a degree of outlier contamination. This led to Random Sampling [Fischer and Bolles 81], whereby small subsets of the total data set are taken
at random, and an ellipse fitted to each of them in turn. The minimum subset size for ellipses is five points, and enough subsets are taken to give a 95% probability (on a $\chi^2$ test) that one of the subsets will produce a fit which lies within given criteria.

In experiments, one hundred subsets were taken of five random points per subset, and the six ellipse coefficients ($\{A, B, C, d, e, f\}$ in Eq. 11.24) were calculated from each fit. One idea is to then select the median value for each parameter. This method, however, failed to improve robustness in our case, because of the large number of outliers in the set.

A better way of finding the best fitting conics proved to be slightly different. The conics chosen as being "best" were those which had the greatest number of points in the total set of data, which were below a certain residual threshold. The residual was calculated simply by working out the value of $|F(x, y)|^2$ for each point.

The residual threshold per point was chosen so that on average, around six conics were produced, for which the residual of every point in the data set lay below the threshold. For each of these, the dimensions of the conic were calculated from the six coefficients $\{A, B, C, d, e, f\}$. The one chosen was then that which had the smallest overall sum of residuals of the total set of data. Additionally, prior information about an environment, or knowledge from other sensors, could be used to select which ellipse to use for the final fit. For example, if it was known that all pipe sections had horizontal or vertical axes (hence giving circles when scanned at 90°), then the ellipse section which had a good fit and was most circular could be used.

Further data was acquired from a larger and more complex environment.
Fig. 11.7  A data scan extracted from the Perceptron sensor (a), and one from the SICK sensor (b). The scale is in metres

Fig. 11.8  The image segmented from scan (a) in figure 11.7. The numbers (which can be seen faintly) are the average Validation Gate value for each portion of the scan and show how the fitting algorithm progressed. The centres of the ellipses recognised are marked with a cross

Data and results from the line/ellipse fit are shown in figures 11.7 and 11.8. The robustness of the system was much improved by this random sampling system and the frequency of ellipse detection increased from 65% to just over 80%, over a large number of scans.
11.3 Conclusions

In this chapter we have shown how segments from lines and ellipses can be extracted from optoelectronic data. It illustrated the great strength of this type of sensing which lies in the accuracy of the measurements and the closeness of the representation to the physical surface geometry. The problems lie in the need to scan, which makes acquisition, especially 3-D acquisition slow (although triangulation sensors can be set up to acquire 3-D images with only a 2-D scan, by transmitting a laser stripe), and in lack of sensitivity where there is bright ambient light.

11.4 Acknowledgments

This work was partially funded by BNFL, Sellafield.
Chapter 12
Active Vision for Mobile Robot Navigation
Andrew J. Davison

12.1 Vision for Mobile Robots

Vision is the sensor which is able to give the information “what” and “where” over the fullest region for the objects a robot is likely to encounter. Although we must be somewhat cautious about comparisons between robot and biological systems, it is clear that it is the main sensory aid to navigation for humans and most animals.

By the term “vision” in robotics, we are usually referring to the processing of digitised images obtained from cameras which are sensitive to electromagnetic radiation in approximately the same range of wavelength as the human eye (400–700nm). There are a wide variety of compact camera modules available which can be attached to mobile robot platforms. Cameras typically consist of a lens (fixed or zooming, often with automatic iris control) in front of a CCD array which has anything in the range of a few thousand to several million discrete intensity sensors or “pixels” in a rectangular array. Each pixel records the intensity value of light arriving at the camera from a particular direction in the scene, with a precision of 8 bits (a value in the range 0–255) or more. Until very recently, the process of transferring image data from camera to computer has relied on television technology, where the values of the pixels are encoded sequentially into an analogue signal at 50Hz (PAL system) or 60Hz (NTSC system); this signal is then decoded back into pixel values by a video capture card in a computer. However, digital cameras which are able to transmit signals to a computer without an analogue stage are now becoming commonplace, and have many advantages in terms of picture quality and ease of use, as well as removing capturing frequency restrictions.

Monochrome and greyscale cameras are both often used in robotics; colour cameras simply have three sensor elements at each pixel position.
which are sensitive in the red, green and blue ranges of wavelength respectively and return three times as many bits of data to represent image intensity. However, the advantages of using colour are perhaps not as clear-cut as it might seem from the point of view of a human to whom colour seems very eye-catching: it should be remembered that our brains are able to understand the world very well just from greyscale images, as watching a black and white film proves. In the robot case, using colour images means processing three times as many bytes, and this extra computational effort might be better spent in other ways: for instance in processing the images from three greyscale cameras which are placed in different positions.

In some ways the attraction of vision as a research area is exactly that so many parallels can be drawn with our familiar experience of seeing as humans: we are presenting a robot with the same data as we receive from the human eye, and therefore would hope that one day a computer would be able to do the same job as the brain in processing this data to understand the world.

**12.1.1 Active Vision**

Humans are most certainly in possession of an *active* vision system. This means that we are able to concentrate on particular regions of interest in a scene, by movements of the eyes and head or just by shifting attention to different parts of the images we see. What advantages does this offer over the *passive* situation where visual sensors are fixed and all parts of images are equally inspected?

- Parts of a scene perhaps not accessible to a single realistically simple sensor are viewable by a moving device — in humans, movable eyes and head give us almost a full panoramic range of view.
- By directing attention specifically to small regions which are important at various times we can avoid wasting effort trying always to understand the whole surroundings, and devote as much as possible to the significant part — for example, when attempting to perform a difficult task such as catching a ball, a human would concentrate solely on the moving object.

Active vision can be thought of as a *task*-driven approach. With a particular goal in mind for a robot system, an active sensor is able to select from the available information only that which is directly relevant to a solution,
whereas a passive system processes all of the data to construct a global picture before making decisions — in this sense it can be described as data-driven. Visual images contain a very large amount of information: consider that a 24 bit colour image at resolution 1000 × 1000 is 3 megabytes of data. Processing whole images with a bottom-up, passive approach is often computationally prohibitive as well as unnecessary.

The emerging view of human vision as a “bag of tricks” [Ramachandran 90] — a collection of highly specialised pieces of “software” running on specialised “hardware” to achieve vision goals — rather than a single general process, seems to fit in with active vision ideas if a similar approach is adopted in artificial vision systems. High-level decisions about which parts of a scene to direct sensors towards and to focus attention on can be combined with decisions about which algorithms or even which of several available processing resources to apply to a certain task. The flexibility of active systems allows them to have multiple capabilities of varying types which can be applied in different circumstances.

12.1.2 Navigation Using Active Vision

Active, directable cameras potentially provide a navigating vehicle with the ability to fixate and track objects in the environment over extended periods of time, and wide fields of view. If we can repeatedly make measurements, over a wide range of robot motions, of the positions relative to the robot of a selection of fixed objects in the world, we are able to use this information to estimate the motion of the robot, even if the positions of the objects themselves are not known in advance (the essence of the simultaneous localisation and map-building approach). While it is relatively straightforward to apply fixating vision to tactical, short-term navigation tasks such as servoing around obstacles where the fixation point does not change [Murray et al 96], using serial fixation on a succession of features to provide information for strategic navigation is more involved. However, active vision is well-suited to this task: the ability to measure features over such a wide range means that the same ones can be used as a robot makes a wide range of movements.

Our core work concerns the problem of simultaneous localisation and map-building for a robot with a stereo active head, operating in an unknown environment and using point features in the world as visual landmarks. In general, a “feature” is anything (point, line, plane, corner, etc.) which
can be identified repeatably by vision (or other sensors) and inserted into a map as a simplified representation of part of the environment. Some approaches to visual navigation have not explicitly worked with so-called features, using instead dense pixel-by-pixel maps of depth for instance to represent the geometry of a scene. These approaches are not suitable for building useful maps, however, since the amount of data involved quickly becomes unworkable: our generic definition of feature is just a way to parameterise the complexity of the scene into quantities with small numbers of parameters, and is not limiting.

We have attached importance to producing maps which are useful for extended periods of navigation. Many map-building methods fail on extended runs because they do not have the ability to recognise previously visited areas as such and adjust their maps accordingly. With active cameras, the wide field of view means that we really can expect to re-detect features found a long time ago, even if the area is not passed through along the original trajectory. Maintaining a large, consistent map requires detailed information to be stored and updated about the estimates of all the features and their relationships. This information is computationally expensive to maintain, but a sparse map of landmark features can be handled successfully.

12.1.3 A Robot Platform with Active Vision

Figure 12.1 shows the robot used in this work. A simple 3-wheeled vehicle base was equipped with a member of the high performance "Yorick" [Sharkey et al 93] series of active stereo heads, for which performance characteristics are given in Table 12.1. The head's wide range of movement and high acceleration mean that it is possible to drive the cameras to fixate at a given angle within an almost spherical field of view within a few tenths of a second (and of course small redirections of gaze can happen much faster). The active head has very accurate optical encoders, so at all times it is possible to know its angular position, and therefore the direction of the cameras. We use sophisticated mathematical models of the robot and head to calculate how control inputs lead to motion and how geometry leads to image measurements, and also model the uncertainty in these processes so that everything can be combined in a statistical framework.

All the robot's vision and navigation processing, above the lowest control levels, was carried out on an offboard PC running Linux on a Pentium
Fig. 12.1 (a) The robot used in this work comprises an active binocular head, equipped with two video cameras, mounted on a three-wheeled vehicle. (b) Detail of the active head which has four axes of movement: pan, elevation and left and right vergence.

<table>
<thead>
<tr>
<th>Description</th>
<th>Vergence</th>
<th>Elevation</th>
<th>Pan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axis Range</td>
<td>178.5°</td>
<td>177.5°</td>
<td>335°</td>
</tr>
<tr>
<td>Max Slew Rate</td>
<td>540°/s</td>
<td>470°/s</td>
<td>430°/s</td>
</tr>
<tr>
<td>Max Acceleration</td>
<td>38000°/s²</td>
<td>25000°/s²</td>
<td>20000°/s²</td>
</tr>
<tr>
<td>Backlash</td>
<td>0.0075°</td>
<td>0.0075°</td>
<td>0.0025°/s</td>
</tr>
<tr>
<td>Angle Resolution</td>
<td>0.00036°</td>
<td>0.00036°</td>
<td>0.00018°</td>
</tr>
<tr>
<td>Repeatability (min)</td>
<td>0.00036°</td>
<td>0.00036°</td>
<td>0.00018°</td>
</tr>
<tr>
<td>Repeatability (max)</td>
<td>0.005°</td>
<td>0.005°</td>
<td>0.005°</td>
</tr>
<tr>
<td>Mass</td>
<td>8.5kg (excluding cameras)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Payload</td>
<td>2 x 0.5kg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.338m</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 12.1 Performance and specifications of the Yorick 8-11 active head

processor. In realtime operation, head and robot movement, image acquisition and processing and all map updating were carried out at a frequency of 5Hz.
12.2 Scene Features

The features which will make up the sparse map used for localisation in this work need to be "landmarks" for the robot: they must be reliably and repeatedly detectable and measurable, and preferably from a wide range of robot positions. This differs slightly from the requirements of the features used in more traditional visual structure from motion algorithms [Torr et al 98] where it is not usually intended that the same features (usually "corners" or line segments) be detected and matched over long periods — indeed this is typically not possible because with a passive camera features can go out of view very quickly. What properties do our landmarks need to have?

- They must be features present in a normal environment so that no artificial interference such as positioning of beacons is required.
- They must be stationary in the scene (or potentially have known motions, though this is beyond our current scope).
- They must be easily identifiable from a variety of ranges and angles, and not frequently be occluded.

It was decided to restrict landmarks to being stationary, point features, and to use simple image patches to represent them, with matching to be performed using normalised sum-of-squared-difference correlation. The patches used are larger than those usually representing corner features in structure from motion systems: $15 \times 15$ pixels rather than $3 \times 3$ or $5 \times 5$. The reason for this choice is that with larger patches the chances of mismatches are reduced due to their greater distinguishability. Having larger patches makes processing slower than otherwise, but the relatively small number of features used and the fact that with the active approach only one feature is observed at a time means that this is not a problem. The key to our approach is to have a small number of very reliable landmarks rather than a large number of transient ones.

12.2.1 Detecting Features

To detect good features automatically in an image, the operator suggested by Shi and Tomasi [Shi and Tomasi 94] has been used. This feature detector is very similar to the Harris corner detector [Harris and Stephens 88], but Shi and Tomasi apply it to patches of a large size similar to those in our
Scene Features

work. Its function is to find image patches which will make good features due to their high variation in intensity — these stand out well from their blander surroundings.

To evaluate the "interest" value of a particular patch of an image, first the horizontal and vertical spatial gradients of image intensity, \( g_x \) and \( g_y \) respectively, are calculated at each pixel position. The \( 2 \times 2 \) matrix \( Z \) is then formed, where:

\[
Z = \sum_{\text{patch}} \begin{bmatrix} g_x^2 & g_x g_y \\ g_y g_x & g_y^2 \end{bmatrix}.
\]

The two eigenvalues \( \lambda_1 \) and \( \lambda_2 \) of \( Z \) are found. The patch has a high interest value if the smaller of \( \lambda_1 \) and \( \lambda_2 \) is large. Cases where just one of the eigenvalues is large mean that the patch has a large interest score in one image direction, but not in the perpendicular direction — the patch contains an edge-like feature, which is not useful as a point landmark. (Incidentally, this does mean that the operator can be used as a combined corner and edge detector as in [Harris and Stephens 88].)

To find the most "interesting" patches in an image, therefore, the operator is scanned over it, and the patches giving the largest small eigenvalues of \( Z \) are chosen. The algorithm is inexpensive: only first-order image gradients are required, and an efficient scanning implementation can be achieved — to form the sums \( \sum g_x^2 \), \( \sum g_x g_y \), \( \sum g_y g_x \) and \( \sum g_y^2 \) at a new patch position, incremental changes can be made to the results from neighbouring locations.

Figure 12.2 shows the best patches found by the operator in different views of the laboratory. The central region only of each image has been searched, and the five best non-overlapping patches are shown. However, it is easy from manual inspection of the images to say that some of the patches will make good landmarks while others will not: they clearly correspond to parts of the image which contain depth discontinuities, meaning that a point landmarks position is not defined, or in other cases to reflections which are not stationary references because they will move as the robot does (although interestingly reflections in planar mirrors do behave as stationary features).

Some previous work, such as that by Shilat et al. [Shilat et al 97] has attempted to identify "bad" features like these at the detection stage. Their approach is to reject patches which contain very sharp "step" edges, since
Fig. 12.2 The best five feature patches automatically detected in the central region of different views of the laboratory. The views show potential problems with occlusion boundaries (b, c) and reflections (c, d) which can lead to the selection of patches not corresponding to stationary point features.

these are most likely to occur with depth discontinuities or reflective highlights. In our work, and that of others like [Shi and Tomasi 94], however, no attempt is made to discern good or bad features at the detection stage: the strategy used is to accept or reject features depending on how well they can be tracked once the robot has started moving. Patches which do not actually correspond to stationary, point features will quickly look very different from a new viewpoint, or will not appear in the position expected from the vehicle motion model, and thus matching will fail. These features can then be deleted from the map, as will become clearer in our discussion of experiments later.
12.2.2 Searching for and Matching Features

The patch detection algorithm described above is run only to find new features. The image values of the selected patches are saved, those features are inserted into the robot’s map, and then when it is necessary to find a particular feature in a subsequent image, a search is carried out using normalised sum-of-squared-difference.

In our localisation system, since we propagate full information about the uncertainty present in the map as will be explained later, whenever a measurement is required of a particular feature, a region can be generated in the left and right images within which the feature should lie with a high probability. This means that only that area of the images need be searched. This is the crux of the active methodology used in this chapter. Search regions are typically ellipses centred in the image: some examples are shown in Figure 12.3.

To search such an elliptical region for a match to a known feature patch, at each position in the region (scanned in left-to-right, top-to-bottom “raster order”), each patch-sized part of the region is compared with the patch saved to represent the feature being searched for using the following normalised sum-of-squared-difference measure:

\[
C = \frac{\sum_{\text{patch}} \left[ \frac{f_1 - f_0}{\sigma_1} - \frac{f_0 - f_0}{\sigma_0} \right]^2}{n},
\]  

(12.1)
where $I_0$ and $I_1$ are the image intensity values at corresponding positions in the saved patch and image patch respectively, and $\bar{I}_0$, $\bar{I}_1$, $\sigma_0$, and $\sigma_1$ are means and standard deviations of the intensity across the patches. $n$ is the number of pixels in a patch. This expression has the value 0 for a perfect match, with higher values as the difference increases. It was considered essential to use a measure which was normalised with respect to overall intensity changes, since this gives improved matching over the long periods over which the system is required to work, when lighting conditions might change, and also since when working in stereo the two cameras will vary somewhat in their output. What the expression actually measures is the average difference in standard deviation above the patch mean intensity between corresponding pixels in the patches being compared. The best (lowest) value of $C$ found in the search region is accepted as a match if it lies below a threshold (set manually at around 0.9).

Figure 12.4 shows matches obtained of some features, giving an impression of the surprising range of viewpoints which can be matched successfully using the simple patch representation of features. However, clearly matching can only be expected to succeed for moderate robot motions, since the patch representation is intrinsically viewpoint-variant — features look different when viewed from new distances or angles. Therefore, we have defined a criterion for expected matchability based on the difference between the viewpoint from which the feature was initially seen and a new viewpoint. Figure 12.5 shows the situation: $h_{\text{orig}}$ is the vector from the head centre to the feature when it was initialised, and $h$ is that from the head centre at a new vehicle position. The feature is expected to be visible if the length ratio $\frac{|h|}{|h_{\text{orig}}|}$ is close enough to 1 (in practice between something like $\frac{5}{7}$ and $\frac{7}{5}$) and the angle difference $\beta = \cos^{-1}(\langle h \cdot h_{\text{orig}} \rangle / (|h||h_{\text{orig}}|))$ is close enough to 0 (in practice less than 45° in magnitude). In our localisation algorithm to be described later, we are in a position to estimate both of these vectors before a measurement is made, and so attempts are made only to measure features which are expected to be visible. The result is a region of the robot’s movement space defined for each feature from which it should be able to be seen. A feature which fails to match regularly within this region should be deleted from the map, since the failures must be due to it being an essentially “bad” feature in one of the senses discussed above.

It should be noted that the patch representing a certain feature is not updated when a successful match is made and the feature is re-found —
one strategy would be to replace the saved patch with the region matched to it in the latest image. This would certainly improve the range of robot movement over which the feature could be tracked, since matching would always be between images taken from closely-spaced viewpoints. However, the problem with this approach is that the patch tends to drift away from its original position over the feature: a small error in locating the feature at each step (due even to just the pixelation limitation) can lead to a large patch movement over many frames. It is essential that landmarks are stationary features.

12.2.3 Other Feature Types

Our map-building methodology make no assumptions that features must be point-like objects, and two further feature types have been considered
Planar patch features: several authors [Wiles et al 97; Shi and Tomasi 94] have started to use a representation of point features as small planar regions in the 3D world rather than 2D image patches. If this assumption is made, when an attempted measurement is made of a feature, the estimated transformation between the patch plane and the image plane can be used to warp the saved patch representation into the shape and size in which it will appear from the new viewpoint. Clearly this method has the capability to greatly extend the range of robot motion through which particular features can be seen and measured, and this will benefit localisation systems, where it is always an advantage to see the same landmarks for a long time.

- Using line features is another possibility. Line features would require new active strategies since it is not so clear where to fixate the camera when making measurements of them. Line features are now commonly used in structure from motion systems [Schmid and Zisserman 97; Zhang and Faugeras 92]. Having this capability would make a wider range of environmental objects usable as features, and perhaps lead to improvements in autonomous naviga-
tion capabilities since line features frequently form the boundaries of obstacles, such as the edges of a door frame.

12.3 Fixation

Fixation means directing a camera or cameras to point directly at a feature in the scene: the camera optic axes will pass through the feature, which will be imaged at the image's central "principal" points. Fixation is one of the main tools of active vision: it means that image processing always takes place near the centre of images, where discrepancies are most easily dealt with — there is little danger of objects of interest being lost outside image boundaries. Many vision tasks can be simplified with simple fixated processing. A fixated camera acts as a "pointing stick" towards a feature. In addition, image processing near to the centre of an image reduces reliance on calibration parameters, as will be shown in Section 12.3.2.

In our work, fixation is used whenever possible to make measurements of the positions of features.

12.3.1 Acquiring Features

Acquiring a new landmark consists of detecting a good patch in an image, then driving the active head to fixate the feature so that a measurement can be obtained. The feature detector of Section 12.2.1 is applied in the left image, and an "epipolar line" is generated in the right image: this is the image in the right camera of the line in the scene upon which the feature is known to lie from its image location in the left image (which defines a ray in space), and it can be determined from our head model and knowledge of the head angles. A search for a match of the feature is carried out in the close vicinity of this line (several pixels either side of the line). If the match is successful, the 3D location of the point relative to the head centre is calculated via our head model. The head angles necessary to fixate the feature are then determined, and the head is driven to this position. The feature should now be centred in both images: searches for the feature in circular regions near to the principal points of both are carried out. If the feature is found within a very small radius of the principal point in both images, the fixation is successful; otherwise, its position and new head angles are re-calculated, and fixation is re-attempted. If fixation fails
several times, the feature is abandoned: a mismatch is usually to blame.

### 12.3.2 The Accuracy of Fixated Measurements

In image measurements, we can expect to detect features to an accuracy of ±1 pixel in normal circumstances. How does this translate into angular errors in fixation? Referring to Figure 12.6 which shows a simple view of imaging geometry in one plane, we can relate the angle to a scene feature $U$ to its image position $u$ via the equation:

$$\tan U = \frac{u - u_0}{f k_u},$$

where $f$ is the focal length of the lens and $k_u$ is the pixel density on the image plane measured in pixels $m^{-1}$. Differentiating this to determine how small changes $\delta u$ in $u$ relate to small changes $\delta U$ in $U$, we can deduce that for our cameras a one pixel uncertainty in image space relates to an angular uncertainty of

$$\delta U \approx 0.006\text{rad} \approx 0.3^\circ.$$

Compared to this, angular errors introduced by the active head, whose axes have repeatabilities with a maximum error of $0.005^\circ$, are negligible. Therefore, in all head measurements, an uncertainty of approximately $0.006\text{rad}$ is assigned to angular values obtained from head/image measurements. These uncertainties are directly incorporated into our localisation and map-building filter.

Translating this into uncertainties in the positions of features estimated using fixated stereo measurements is revealing. Table 12.2 shows the depth...
Fixation

<table>
<thead>
<tr>
<th>Depth $z$ (m)</th>
<th>$\gamma_R = -\gamma_L$ (radian)</th>
<th>$\Delta x$ (m)</th>
<th>$\Delta z$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.6987</td>
<td>0.0014</td>
<td>0.0019</td>
</tr>
<tr>
<td>0.5</td>
<td>0.3241</td>
<td>0.0022</td>
<td>0.0070</td>
</tr>
<tr>
<td>1</td>
<td>0.1664</td>
<td>0.0044</td>
<td>0.026</td>
</tr>
<tr>
<td>2</td>
<td>0.08380</td>
<td>0.0086</td>
<td>0.10</td>
</tr>
<tr>
<td>3</td>
<td>0.05594</td>
<td>0.013</td>
<td>0.22</td>
</tr>
<tr>
<td>4</td>
<td>0.04198</td>
<td>0.017</td>
<td>0.40</td>
</tr>
<tr>
<td>5</td>
<td>0.03359</td>
<td>0.022</td>
<td>0.64</td>
</tr>
<tr>
<td>6</td>
<td>0.02799</td>
<td>0.026</td>
<td>0.92</td>
</tr>
<tr>
<td>7</td>
<td>0.02400</td>
<td>0.030</td>
<td>1.2</td>
</tr>
<tr>
<td>8</td>
<td>0.02100</td>
<td>0.034</td>
<td>1.6</td>
</tr>
<tr>
<td>9</td>
<td>0.01866</td>
<td>0.038</td>
<td>2.0</td>
</tr>
<tr>
<td>10</td>
<td>0.01680</td>
<td>0.042</td>
<td>2.6</td>
</tr>
<tr>
<td>15</td>
<td>0.01120</td>
<td>0.064</td>
<td>5.6</td>
</tr>
<tr>
<td>20</td>
<td>0.008400</td>
<td>0.084</td>
<td>10.0</td>
</tr>
</tbody>
</table>

Table 12.2 The uncertainties of the positions of fixation points measured by binocular cameras with inter-ocular spacing 0.338m and angular vergence uncertainties of 0.006rad

Stereo cameras
0.336m baseline

0.2m 0.5m 1m 2m 3m 4m 5m

Fig. 12.7 In this scale diagram ellipses show the uncertainties (1 standard deviation) in recovered point positions from stereo fixation up to depths of 5m.

($z$ coordinate) and transverse ($x$ coordinate) uncertainty in estimated feature locations when measurements are made of features lying at varying distances directly in front of the robot. Figure 12.7 shows uncertainty ellipses to scale for distances up to 5m. It is clear that while the transverse uncertainty remains small, growing only to 8cm at a distance of 20m, the depth uncertainty grows rapidly as the feature distance increases. This is because at large depths the two cameras are nearly parallel, and the vergence angles change very little for large distance changes: the small uncertainties in vergence measurements translate into large depth uncertainties.
Active Vision for Mobile Robot Navigation

It will be seen later how this uncertainty distribution affects navigational capabilities.

12.4 Localisation and Map-Building

We use the simultaneous localisation and map-building methodology using the Extended Kalman Filter described earlier in the chapter 2 and in [Davison and Kita 00; Davison 98] and similar to that of various authors [Castallanos 98; Chong and Kleeman 99a; Durrant-Whyte et al 99; Kwon and Lee 99]. Specifically, since our robot moves in 2D and makes measurements of point world features in 3D, the state vector and covariance matrix have the following form:

\[
\hat{x} = \begin{pmatrix}
\hat{x}_v \\
\hat{y}_1 \\
\hat{y}_2 \\
\vdots
\end{pmatrix}, \quad P = \begin{bmatrix}
P_{xx} & P_{xy_1} & P_{xy_2} & \cdots \\
P_{y_1x} & P_{y_1y_1} & P_{y_1y_2} & \cdots \\
P_{y_2x} & P_{y_2y_1} & P_{y_2y_2} & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{bmatrix}.
\]

(12.2)

\(\hat{x}\) has \(3(n + 1)\) elements, where \(n\) is the number of known features. \(P\) is symmetric, with size \(3(n + 1) \times 3(n + 1)\). \(\hat{x}\) and \(P\) will change in dimension as features are added or deleted from the map. \(\hat{x}_v\) is the robot position estimate, and \(\hat{y}_i\) the estimated 3D location of the \(i\)th feature:

\[
\hat{x}_v = \begin{pmatrix}
\hat{z} \\
\hat{x} \\
\hat{\phi}
\end{pmatrix}, \quad \hat{y}_i = \begin{pmatrix}
\hat{X}_i \\
\hat{Y}_i \\
\hat{Z}_i
\end{pmatrix}.
\]

At this point we refer the reader to the discussion of map-building in chapter 2, and in particular would like to recall the importance of carrying the cross-coupling of estimates which happens in the mapping process and how the EKF algorithm used here is able to fully capture this aspect. The effect of this methodology in a real system will be seen clearly in the following sections.

12.4.1 An Extended Experiment

In typical navigation, the robot will not solely make use of landmarks detected from its starting position: prominent landmarks will be initialised, measured and sometimes rejected from the map at all stages of the journey.
Later we will consider the automation of these steps, but in this section an experiment is presented where the robot moved through a long distance, building up and using a realistic map, and stopping at frequent step intervals so that ground-truth measurements of its position could be made manually. The key aim of the experiment was to verify the localisation algorithm over a long run, and in particular check its ability to re-find features after long periods of neglect and "close the loop" on a round-trip journey.

From the robot’s starting position, just two features were identified and initialised as landmarks. The landmark identification was carried out largely automatically, but some human input at this stage helped reliable features to be chosen. For each movement step (of approximate size 40cm), one feature was chosen, and 10 measurements were made of it during the motion. After each step, the robot position was measured, and a new feature was chosen for the next step. New features were initialised when the existing ones went out of view.

15 features were used altogether as the robot moved twice up and down the gridded corridor-like area of the laboratory. The robot steered sharply at the ends of the corridor to turn through 180°. The farthest point reached was about 6m from the origin, so the total distance travelled was around 24m.

Some of the features used are displayed and labelled in the order in which they were initialised in Figure 12.8. The progress of the robot can be followed in Figure 12.9, where an overhead view of the state at 12 numbered

---

**Fig. 12.8** Fixated views of the first 8 features used in the extended experiment
Fig. 12.9 State snapshots from an extended experiment. Ground truth robot and numbered feature positions in black are tracked closely by the robot's estimates, in grey with ellipses representing feature uncertainty.

step counts is displayed, ground truth being overlayed for those steps for which it was available. Numbered features are labelled in the diagrams shortly after they are first seen by the robot and initialised in the map, usually with large covariances. Some features, such as 3 and 5, were deleted
'Localisation and Map-Building'

automatically from the map because measurements had failed and they were “bad” in the senses described in Section 12.2 (it can be seen in Figure 12.8 that feature 3 for instance happened to lie on an occlusion boundary and was therefore not a real point feature). Feature 4 never shows up in the state snapshots because it was rejected very shortly after initialisation.

The robot moves up the corridor for the first time through steps (2), (7) and (10), before starting its first turn at step (14). It can be seen that the features initialised when the robot is at this end of the corridor, such as as numbers 5–9, have estimated positions which are much more uncertain than those initialised from near the origin such as 1 and 2 — their uncertainty ellipses are quickly reduced to a size which is too small to show up on the diagrams. This reflects a fundamental characteristic of map building: these features are far (in terms of the number of measurements implicitly compounded to estimate their locations) from the place where the coordinate frame is defined; the robot location itself is uncertain when they are measured, and so their locations are as well.

In step (22), the robot gets its first view of the back wall of the laboratory, and initialises feature 10 which is the corner of a window. Step (27) is key, because it provides an opportunity to re-measure an early feature: number 2 comes back into view. The result of this is that the estimate of the robot’s position immediately improves — it had been drifting slightly up to step (22), but locks closely back on in (27). Also, the estimates of the locations of features initialised since measurements were last made of feature 2 improve, and their uncertainties reduce, due to their coupling to the robot state in the Kalman Filter — the uncertainty ellipses of features 6–9 and especially 10 visibly shrink. A further slight improvement occurs in step (32) when feature 1 is re-found.

Close to the origin, the robot turns in step (35) and heads back up the corridor. Far fewer new features need to be initialised now because the old ones are visible from this previously-followed route. Turning again at the far end of the corridor in (53) and back at the origin in (69), the robot halts at step (72), close to its starting position and orientation. The map of features generated sequentially through the experiment has now reached a fairly stable state, and there is no essential difference between landmarks detected early or late in the run.

To give a numerical idea of the localisation algorithm’s performance, at step (14), where the robot was starting to turn for the first time at the far end of the corridor, the true and estimated robot locations (in metre and
radian units) were:

\[
x_v = \begin{pmatrix} z \\ x \\ \phi \end{pmatrix} = \begin{pmatrix} 5.17 \\ -0.15 \\ -0.39 \end{pmatrix}, \quad \hat{x}_v = \begin{pmatrix} \hat{z} \\ \hat{x} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} 5.25 \\ -0.02 \\ -0.34 \end{pmatrix}.
\]

The covariance of the robot position estimate was:

\[
P_{xx} = \begin{bmatrix} 0.0035 & 0.0003 & 0.0002 \\ 0.0003 & 0.0069 & 0.0033 \\ 0.0002 & 0.0033 & 0.0017 \end{bmatrix}.
\]

Here, the robot is far from its starting position, and the large covariance accounts easily for the discrepancy within one or two standard deviations: the robot’s \(x\) coordinate, for example, was incorrectly estimated by 13cm, and the square root of the (2, 2) element of \(P_{xx}\) was \(\sqrt{0.0069} = 0.08\): \(\hat{x}\) had a standard deviation of 8cm. (Note that here the non-zero value of the cross-terms of the covariance matrix reveals the coupling between the estimates of the different parameters.)

We will also examine the later step (32), where the robot had returned to near the origin and re-measured early features. Here the true and estimated robot locations were:

\[
x_v = \begin{pmatrix} z \\ x \\ \phi \end{pmatrix} = \begin{pmatrix} 0.78 \\ 0.31 \\ 2.76 \end{pmatrix}, \quad \hat{x}_v = \begin{pmatrix} \hat{z} \\ \hat{x} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} 0.77 \\ 0.35 \\ 2.75 \end{pmatrix}.
\]

The covariance of the robot position estimate was:

\[
P_{xx} = \begin{bmatrix} 0.00096 & -0.00001 & -0.00052 \\ -0.00001 & 0.00011 & 0.00002 \\ -0.00052 & 0.00002 & 0.00035 \end{bmatrix}.
\]

Feature 1, one of the features very close to the origin which was therefore known very accurately, had true and estimated positions:

\[
y_1 = \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} = \begin{pmatrix} -1.27 \\ 1.29 \\ 0.72 \end{pmatrix}, \quad \hat{y}_1 = \begin{pmatrix} \hat{X}_1 \\ \hat{Y}_1 \\ \hat{Z}_1 \end{pmatrix} = \begin{pmatrix} -1.25 \\ 1.27 \\ 0.71 \end{pmatrix}.
\]
and covariance

\[
P_{y_1y_1} = \begin{bmatrix}
0.00016 & -0.00007 & -0.00006 \\
-0.00007 & 0.00004 & 0.00004 \\
-0.00006 & 0.00004 & 0.00010
\end{bmatrix}.
\]

It can be seen that after “closing the loop” and refining features close to
the robot’s starting position, the estimates were very accurate: to within a
few centimetres for both the robot and feature positions. We consider that
this is impressive localisation performance after a total journey of around
12 metres.

12.5 Continuous Feature Tracking

In the remaining parts of this chapter, we will look at the ways in which
the accurate localisation system already presented was brought to life in
autonomous navigation in which the robot moved with no external input
at all.

The high-performance active head gives the robot the capability to track
at fixation one feature in almost any viewing direction whenever the robot is
moving, making repeated measurements of that feature as it does so which
provide continuous navigation information. Tracking a feature consists of
making repeated fixated measurements of it as the robot moves past. This
task is of course similar to that of tracking a moving target from a stationary
robot viewpoint, a problem which has been studied in depth already in
active vision. The situation is somewhat simplified now, however, since
the relative movement between robot and target is provided by the robot:
the robot is under its own control, so its motion is at least approximately
known. Tracking a freely manoeuvring target is difficult because although
a model of its possible motions may be known, the “control inputs” are
not. This is not the case when a robot tracks a stationary target.

Further, the details of our robot’s control system mean that in our robot
vehicle motion model [Davison 98] we are able to describe robot movements
directly in terms of the two control inputs velocity and steering angle —
knowledge of these allows an estimate of a new position to be calculated
from the previous position estimate without extra parameters. This might
not be the case in a robot with more complex dynamics: suppose it was
like a normal road car, where the control inputs are the steering angle and
the position of the accelerator pedal, which is more closely related to the acceleration of the vehicle than its actual speed (as discussed by Maybank \textit{et al.} [Maybank \textit{et al} 96] with regard to visual tracking of externally viewed cars) — in order to produce a new position estimate, it is necessary to know the current velocity as well as the current position. The velocity must therefore be part of the state vector of estimated quantities. Other vehicles may require higher-order derivatives to be estimated as well.

For these reasons, our tracking algorithm can proceed quasi-statically. This means that continuous tracking can be performed with the same filtering approach as would be used if the robot was moving in a step-by-step fashion, stopping each time a measurement was to be made. Assuming a certain inter-measurement interval, as the robot starts a movement, the prediction step of the filter produces an estimate of its new position at the next measurement point, and the expected relative position of the target feature is calculated. The active head is then driven to fixate on where the feature should be found, and the robot waits to make its measurement as the interval checkpoint on its trajectory is passed. Vehicle odometry is used as the trigger for measurements: the trajectory is divided into steps of equal numbers of wheel encoder counts. This is better than using strict time intervals, since the distance travelled is what really matters, and possible fluctuations in velocity (particularly when the robot is just starting or stopping) will be unimportant. Because of this, the measurement frequency is not strictly constant, but in current implementation tracking is performed at approximately 5Hz (the odometry measurement intervals being dynamically set proportional to the robot velocity to keep this constant).

The actions taken in a step of the tracking loop, during which the robot drives continuously, are therefore:

1. Based on the current control inputs, perform the prediction step of the filter and estimate the new robot position.
2. Calculate the estimated head angles for fixation on the feature, and move the head to this position in readiness.
3. Wait until the robot has reached the correct "new" position according to its odometry.
4. Obtain new images from the cameras.
5. Perform correlation searching for the feature as in Section 12.2.2.
6. If the feature is found in both images, perform a measurement update of the filter.
Figure 12.10 shows an image sequence obtained from one of the robot’s cameras in a period of fixation tracking.

12.6 A Fixation Strategy for Localisation

The active head allows tracked features to lie within a very wide field of view, meaning that the same one can be tracked as the robot makes large movements, or alternatively that a variety of different features can be tracked at different stages of a motion. The fact that choice is available means that a strategy for selecting features to be observed is required. As discussed in Section 12.2.2, only a subset of features in the map will be viewable from a certain robot location, due to constraints of viewing direction and distance, but which should be observed at a given time?

First, it is necessary to consider in which senses one feature could be “better” than another to make measurements from:

- If measurements of that feature would provide more useful information.
- If less effort is required to make the measurements.

That some measurements can be more valuable than others in terms of information content is quickly apparent: depending on which components of the current estimated quantities, in particular the robot’s estimated position coordinates $\hat{z}$, $\hat{x}$ and $\hat{\phi}$, are currently the least certain, it will be more useful to make measurements of features lying in certain positions. If for instance the robot’s location is uncertain primarily in the forward/backward direction, there will be little use in making a measurement of a feature which lies directly ahead at a large range, since the large depth uncertainty of this measurement would provide little extra constraint on the robot position.

Active fixation switching is very evident in human and animal naviga-
tion — the human eye is rarely at rest as it rapidly shifts between points of interest. During a period of walking, for example, attention will be divided between the ground surface, obstacles, near and far landmarks and other objects of interest. Blind navigation is unreliable and dangerous, as is easily demonstrated by the discomfort felt during an attempt at walking with closed eyes: the urge to re-open the eyes and reorient is very strong. Further, though, staring at just one feature in the world for an extended period while moving can be disorientating — it is necessary to pay attention to a variety of landmarks to be fully confident of safe navigation. Investigating feature selection in this section casts some light on these natural behaviours as well as providing a key part of the robot navigation system, where the full statistical information stored about the estimates of robot and feature positions means that we are in a position to make reasoned decisions about which features to measure at a given time.

12.6.1 Choosing from Known Features

We consider here the case of making a choice between known and currently visible features for an immediate measurement. (Further considerations, such as choosing between newly initialised features and taking into account the “cost” of active measurements, are tackled in [Davison 98].) Because it is intended for long-term use, what is important in the map-building system is the integrity and consistency of the whole map and the estimate of the robot’s position within it. We aim to maximise this with our choice of feature, rather than specifically aiming to improve the estimate of the robot position or that of the chosen feature — these are quantities which live in the world coordinate frame, which is arbitrarily chosen. The relative locations of robot and features are what matter for navigation. Furthermore, these estimates are all intricately coupled by previous measurements — it would be somewhat meaningless to try to optimise just a subset of them. However, we will see below that choice of feature on this basis usually does provide the best improvement in the explicit world-frame estimate of the robot location.

The basis of the approach used is to make a measurement where the ability to predict is lowest, as discussed and derived in recent work by Whaite and Ferrie [Whaite and Ferrie 97] in their paper about active exploration to determine the shapes of surfaces. Speaking generally, given a choice of measurements in a system where the uncertainty in estimates of
the parameters of interest is known, it makes sense to make the one where we are least certain of the result, since this will in a sense “squash” the total uncertainty, viewed as a multi-dimensional ellipsoid, along the longest axis available. If there is any part of the estimated state vector which is particularly poorly known, this choice of measurement will act to reduce that uncertainty, smoothing off the estimation’s rough edges.

In our system, whenever the robot is to make a measurement of a feature, a predicted measurement $h$ is formed and the innovation covariance matrix $S$ is calculated. This matrix describes how much the actual measurement is expected to vary from the prediction in the $\alpha$, $e$, $\gamma$ angular measurement coordinates. The size and shape of the ellipsoid represented by $S$ will reflect the amount of uncertainty in the estimated relative position of the feature and the robot; measurement noise is constant in measurement coordinates so it only provides a constant addition.

To produce scalar numbers to use as the basis for decisions about which feature to observe, the volume $V_S$ in angular $(\alpha, e, \gamma)$ measurement space of the ellipsoid represented by $S$ at the $n_\sigma \sigma$ number of standard deviations level (again, $n_\sigma = 3$ is normally used) can be calculated for each visible feature: eigenvalues $\lambda_1$, $\lambda_2$ and $\lambda_3$ of $S$ mean that the ellipsoid has axes of length $n_\sigma \sqrt{\lambda_1}$, $n_\sigma \sqrt{\lambda_2}$ and $n_\sigma \sqrt{\lambda_3}$; so

$$V_S = \frac{4}{3} \pi n_\sigma^3 \sqrt{\lambda_1 \lambda_2 \lambda_3}. \quad (12.3)$$

This quantity is evaluated for each of the features which is currently visible, and the one with the highest value is selected for measurements.

12.6.2 Experiments

The most striking consequence of this criterion seen in experiments is that it demands frequent changes of measured feature. Once a few measurements have been made of a particular one, the criterion tells the robot that there is not much more information to be gleaned from it at the current time, and it is best to switch attention to another. This is a result of the way that tracking one point feature, even with perfect measurements, does not fully constrain the robot’s motion — uncertainty is always growing in some direction.

The following experiment shows this clearly: five widely-spaced features were initialised by the robot from its starting point at the origin. One
was arbitrarily selected, and tracked continuously while the robot made a forward and backward motion of around 2m, ending up about 30cm in front of where it started. The situation at the end of this motion is displayed in Figure 12.11(a), where it is feature 0 which has been tracked. The five features were then evaluated according to the criterion above — the values of the innovation covariance volume $V_S$ for each were:

*Note that $V_S$ depends on the uncertainty in a predicted measurement - see Section 12.6.1 - and so although the uncertainty of feature 2 is small in the world - see figure 12.11, the predicted uncertainty in a measurement after the robot has made a movement is large. The feature is close to the robot and hence a small uncertainty in robot movement gives a large uncertainty in measurement space.
0. \( V_S = 0.00004 \)
1. \( V_S = 0.00046 \)
2. \( V_S = 0.00127 \)
3. \( V_S = 0.00049 \)
4. \( V_S = 0.00040 \)

The clear conclusion is that there is little merit in making yet another measurement of feature 0. Of the alternative features, 2 is the best, with the others being roughly equal.

Figure 12.11(b, c, d) show the estimated state after an extra measurement of features 0, 1 and 2 respectively. The conclusion of the feature choice criterion proves to be successful: making the extra measurement of feature 0 in (b) does little to improve the robot position estimation, which has drifted along the direction ambiguous to measurements of that feature; (c) and (d), however, show the robot estimate locked back onto the ground-truth. There is little difference to be seen between these latter two from the diagrams or between their estimate of the robot state (both of which are within a centimetre or two of the truth); the robot state covariances in the two cases, however, are:

\[
P_{xx}(1) = \begin{bmatrix}
0.00035 & 0.00008 & -0.00013 \\
0.00008 & 0.00024 & -0.00009 \\
-0.00013 & -0.00009 & 0.00010
\end{bmatrix}
\]

\[
P_{xx}(2) = \begin{bmatrix}
0.00010 & 0.00005 & -0.00003 \\
0.00005 & 0.00021 & -0.00010 \\
-0.00003 & -0.00010 & 0.00009
\end{bmatrix}
\]

The second, after a measurement of feature 2 which was favoured by the \( V_S \) criterion, is noticeably smaller, showing better confidence in the new robot state estimate. Feature 2 lies much closer to the robot than 1, 3 and 4: compared to these, just the small change of viewpoint in the robot motion provides more chance to look at 2 from a different angle and get some different information than provided from the initial measurement.
12.7 Steering Control and Context-Based Navigation

By context-based navigation, we mean making movements in relation to what can be seen in the near surroundings in order to reach local goals. Useful journeys can be formed by combining movements of this kind. The obstacle-avoidance manoeuvres of [Murray et al 96] fall into this category. What is difficult is how exactly to combine contextual movements towards a total goal. With a form of topological map of a known journey, as perhaps humans have, one small movement leads to another, meaning that complex routes can be travelled without absolute position knowledge.

A robot system can, and therefore perhaps should, do better than this. Position-based navigation, with its requirements of accuracy and fast processing of a lot of data, is exactly the sort of thing that computers do very well. Context-based navigation cannot help in all situations: sometimes position-based is better — when striking out into an open region between two known areas for instance. How can the two types of navigation be linked?

A way in which this can be achieved in our landmark-based system is to attach "labels" to certain features, which provide information on their context in the world and how local navigation should proceed with respect to them. We can foresee how this could be performed automatically: for instance, if a free-space-finding vision module reports that an obstacle lies ahead which has a large amount of free-space to the left side, a landmark feature lying on the left obstacle boundary could be labelled as "a feature which must be steered around to the left side". If no features had already been found in this region, an active attempt could be made to initialise one there.

Another way of attaching a label to a feature is to specify the feature at the start of motion, and to affix the label at that stage with human assistance. We will give an example in the following autonomous experiment.

12.7.1 Steering a Twisting Course

An autonomous experiment was conceived where two known features were specified on the inside corners of a zig-zag route to be followed along corridors in the laboratory over a total distance of about 12 metres. These features had positions which were known accurately in the world, and they were initialised into the map manually as prior knowledge (they appear in
The estimated trajectory and frames cut from a video as the robot navigated autonomously around two known landmarks and out of the laboratory door. The navigation knew the locations of features 0 and 1 as prior knowledge, along with information on their status as obstacles.

The "journey plan" given to the robot consisted of the following:

1. Go forward to waypoint \((z, x) = (2.0, 0.0)\).
2. Steer to the left of feature 0.
3. Steer to the right of feature 1.
4. Go to waypoint \((z, x) = (8.5, -3.5)\) and stop.

The robot set off, now using all the techniques developed in this chapter and moving fully autonomously, performing map-building and localisation and steering as described above. The results of the experiment are shown in Figure 12.12, where the estimated trajectory generated is pictured next to stills from a video of the robot safely navigating the course.
Steering control in this experiment involved making transitions between various waypoint-steering and obstacle-steering modes. Switching out of a waypoint-steering stage is trivial, because the waypoint is said to have been reached when the robot is within a certain small distance of it. Making the transition from obstacle-steering is more involved, but necessary to stop the robot from continuing in a never-ending orbit. The point at which to leave depends on what the robot is going to do next — some obstacles will cause the robot to deflect its trajectory only slightly in a short period of avoidance, while some, such as the inside corner of a 90° turn in a corridor, will require relatively long steering manoeuvres. The solution is to consider the next guidance point in the motion — either waypoint or obstacle — and compare the current steering action with that which the next point would be demanding in the absence of the current obstacle. The transition to the next guidance point should be made when the two instantaneously agree: the current point will have done its job and control can pass to the next. This will only happen when the current obstacle has been cleared sufficiently that the robot can start making for the next waypoint or obstacle-avoidance without danger of hitting it. Figure 12.12 shows the smooth transitions occurring between waypoint and obstacle steering modes which this criterion produces.

Note that in this experiment, steering around the known obstacles took place on a positional basis. The robot steered so as to avoid the known obstacles based on its current position estimate, even before it had first measured them. The automatic feature-selection criterion decided when it was necessary actually to measure the known features, and in the experiments this proved to be as soon as they became visible, in order to lock the robot position estimate down to the world frame. The point when a first measurement of known feature 0 is made can be clearly seen in Figure 12.12 as a small kink in the robot trajectory: actually measuring the feature corrected the robot’s drifting position estimate and meant that the steering angle was changed slightly to correct the approach. After this, the obstacle feature was fixated on only when it again became the best measurement to make. Otherwise, attention was paid to improving the map of automatically-acquired features.
12.8 Summary

In this chapter we have discussed the role that vision, and particularly active vision, can play in robot localisation and navigation, and presented a full implementation of autonomous localisation incorporating real-time stereo feature tracking, automatic feature initialisation and measurement selection, simultaneous localisation and map-building and intelligent steering control. The accurate vision-based localisation provides a solid basis for future developments in navigation using vision or other sensors.

The C++ software implementing all of this research is available as part of an open-source, generalised package for simultaneous robot localisation and map-building at:

http://www.robots.ox.ac.uk/~ajd/Scene/

MPEG videos depicting aspects of the work are also available from this site.
PART IV

ACTIVE SENSOR ADAPTATION
Chapter 13

Strategies for Active Sensor Management

Stephen Reece

13.1 Introduction

This chapter* focuses on the need to adapt and schedule signal processing software and sensors for image interpretation tasks. A framework is introduced by which a set of sensors and/or algorithms can adapt to different environments and contexts. The chapter addresses the structure applied to signal processing software.

Our ability to devise safe, reliable signal processing software is critical in an age where sensor platforms are increasingly sophisticated. System robustness is hard to guarantee as it is virtually impossible to anticipate all the rich, subtle and complex operating environments the system may encounter. For the sensor data to be interpreted correctly, significant experience or knowledge of the physics of the sensing process must be available. For example, Fig. 13.1 shows how lighting and temperature conditions can affect infrared images. The human system designer, who may be familiar with the intricacies of vision, may not be completely familiar with the behaviour of sensor modalities such as infrared and ultrasound. Such modalities are best described using very unfamiliar, sensor-centred concepts. For example, a recent ontology† to emerge for describing specular images obtained by scanning, time-of-flight ultrasound sensors contains the "Region of Constant Depth" (RCD) concept [Leonard and Durrant Whyte 92] (introduced in chapter 3 section 3.5.1), which is a feature common to all sonar

*Effort sponsored by the Defence Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory, Air Force Material Command, USAF, under agreement number F30602-98-0056.

†An ontology is the set of entities presupposed by a theory.
Fig. 13.1 An Oxford park setting (including bench, tree and path) seen through a 4.0 to 6.0 micron infrared camera on (1, 2 and 3) a bright day, (4) a dull day, (5) a rainy day, and (6) at night on January 27, 2000. Images were obtained with an Inframetrics, InfraCAM infrared camera, courtesy British Aerospace.
images involving specular reflectors but does not occur in vision. There are many sensor-centred concepts, such as the thermal shadow in infrared sensing.

In general, a sensor ontology contains fundamental and compound concepts. A fundamental concept is a basic sensor-centred feature found in many environment images (the RCD described above, for example). Compound concepts are invariant within an environment but may change between environments. For example, cities, parks and airports are compound concepts of urban environments whereas forests, lakes and mountains are compound concepts of natural environments. Compound concepts are described by combinations of fundamental concepts and their behaviours.

It is difficult for the system designer to furnish the system with a complete ontological model for interpreting sensor images. Even when an ontology is sufficient it need not necessarily be the most appropriate and transforming to a new ontology could be the only way to proceed. For example, circles are best described using the polar coordinate system ontology as opposed to the Cartesian ontology. Thus, there is a need for algorithms which adapt ontologies to system tasks and environments. Further, these algorithms should be equipped with sensor management strategies for selecting appropriate sensor configurations for the ontology and adapting signal processing software to interpret and classify the sensor data [Hager and Mintz 91; Kam et al 97].

This chapter is not concerned with hardware adaptation, for example active vision or sensor placement. Much work has been done in this field [Hager and Mintz 91; Huang 95; Reed and Allen 00]. Some areas are discussed in this book. For a review of the different sensing strategies for visual object detection, object recognition and scene reconstruction, see [Konstantinos et al 95]. Neither is this chapter concerned with the detection and recovery from various kinds of sensing faults including hardware failure and bandwidth limitations. Instead, this chapter emphasises the need for flexible signal processing software which is able to select the most appropriate signal processing tools for the task at hand, select model parameters and structure each tool appropriately and then determine explicitly an ontology for describing the behaviour of various multi-tool configurations in various environments. The signal processing software reconfiguration problem is similar to the sensor placement problem and it is not hard to believe that both share similar solutions including the synthesis of analytical relationships, expert systems, model-based simulation and generate-and-test
methods.

(a) Training Stripes  (b) Spatial Segments

Fig. 13.2 Aerial image segmentation of Edinburgh showing roughly sea, coast, suburbs and parks. The training stripes are shone on the left. On the right the areas around the stripes are shaded according to the region identified

This chapter develops information theoretic sensor software management strategies for infrared aerial image interpretation. The purpose is to determine an image ontology from unsegmented images and then configure sensors to discriminate the concepts within each image. The sensor system is supplied with a repertoire of simple generic signal processing tools. These tools could, in principle, be used in any sensing environment. Each tool determines some statistic of a local image region such as, for example, the average spectral energy or texture entropy. Tool configurations are then identified which exhibit the required image feature recognition coverage and also minimal redundancy. Our approach is able to determine alternative tool configurations in order to recover from software tool failure and it allows dynamic scheduling of the most informative tools by anticipating their informativeness using knowledge of the scene obtained from accumulating observations.

The next section introduces some simple signal processing tools for near-infrared, multi-spectral, aerial image interpretation. Section 13.3 explores the impact that sensor software reconfiguration can have on feature discrimination rates. Sections 13.4 and 13.5 introduce concept formation and
signal processing tool selection algorithms. Finally, Section 13.6 describes a dynamic on-line sensor scheduling algorithm.

13.2 Simple Signal Processing Tools

This section describes some simple signal processing tools which are used to interpret aerial images obtained by a near-infrared, multi-spectral camera. This sensor records reflected radiation at a ground resolution of 20m in three, discrete wave bands - 0.50 to 0.59 microns (green band), 0.61 to 0.68 microns (red band) and 0.7 to 0.8 microns (very near infrared band). Fig. 13.2 shows the 512 × 512 grey scale rendering of a near infrared image of Edinburgh.† We aim to construct an automated image segmentation and segment recognition algorithm. It is envisaged that the algorithm would be supplied with a large corpus of simple signal processing tools. The algorithm would then select a subset of these tools appropriate for its task. The task is to partition the image into spatial segments of roughly equal size so that when the sensor system traverses the same scene it is able to determine when a spatial segment boundary has been crossed. This is the localisation problem.

Simple or complex signal processing tools are used which measure statistics (mean, mode, variance, entropy, histogram etc) of observables (luminosity, texture, shape, size etc). In the aerial image interpretation domain image statistics are determined from the reflected radiation energy J(w) at pixel w. The statistics are evaluated over a sliding window of area 40 × 40 pixels. Each tool determines some statistic for J(w) over subsets of the discrete spectral bands and outputs an ordering (>) of the various spectral bands - green (G), red (R), near-infrared (I) and non of these (B). The tool ordinal output is determined by some statistic S which is calculated for each of the spectral bands a, b ∈ {G, R, I}. In general, a > b whenever $S_a > S_b$.

Some tools operate on the binary spectral domination image $B(w)$. The binary domination image for a spectral band is the set of pixels at which the intensity of the band exceeds that for all other bands. For all a ∈ {G, R, I}:

$$B_a(w) = \begin{cases} 1 & \iff (\forall b \in \{G, R, I\}) \ J_a(w) \geq J_b(w) , \\ 0 & \text{otherwise} . \end{cases} \quad (13.1)$$

†This image was obtained from the European space agency satellite SPOT.
Fig. 13.3 (a) Binary domination images from the SPOT near-infrared multi-spectral data of Edinburgh. Also, (b) binary domination sub-images and corresponding texture histograms

The white areas in Fig. 13.3(a) depict the binary domination images for our image of Edinburgh.

**Luminosity Spatial Extent Tool (LD):** the luminosity spatial extent tool returns an ordering of \( \{G, R, I, B\} \). A spectral band is preferred if it exhibits the greatest intensity over the majority of the pixel window. For \( a \in \{G, R, I\} \) the tool determines the extent of domination of each band:

\[
S_a = |\{w \in W : B_a(w) = 1\}|. \quad (13.2)
\]

For \( a, b \in \{G, R, I, B\} \):

\[
a > b \iff S_a > S_b. \quad (13.3)
\]

**Luminosity Variance Tool (LV):** the luminosity variance tool returns an ordering of \( \{G, R, I\} \). A spectral band is preferred if the variance of its intensity value over the window \( W \) exceeds that of other bands. For \( a, b \in \{G, R, I\} \):

\[
a > b \iff \sigma_{w \in W}(f_a(w)) > \sigma_{w \in W}(f_b(w)). \quad (13.4)
\]

**Luminosity Mean Tool (LM):** the luminosity mean tool returns an ordering of \( \{G, R, I\} \). A spectral band is preferred if the mean luminosity value over
the window $W$ exceeds that of other bands. For $a, b \in \{G, R, I\}$:

$$a \succ b \iff E_{w \in W}(J_a(w)) > E_{w \in W}(J_b(w)).$$  (13.5)

Texture tools are used to evaluate the distribution of structure sizes within an image, including empty space (i.e. black) structures. The texture tools are based on Galloway’s Primitive Length Encoding [Galloway 75] and operate on the binary domination images. The image is examined for neighbouring pixels of the same type (either filled or empty) which fit some structuring element - in this case a square of size $d$ pixels. The number of image pixels $N(d)$ with neighbourhoods that can contain the structuring element are counted for each $d \in [1, 512]$. A histogram over values of $d$, called the texture histogram, records the values $\mathcal{V}(d) = N(d) \times d^2$.

Texture Entropy Tool (TE): the texture entropy tool returns an ordering of $\{G, R, I\}$. This tool operates over the individual spectral band texture histograms. The band with the greatest normalised histogram entropy is preferred. For $a, b \in \{G, R, I\}$:

$$a \succ b \iff \text{Entropy}(\mathcal{V}_a(d)) > \text{Entropy}(\mathcal{V}_b(d)).$$  (13.6)

Texture Mode Tool (TM): the texture mode tool returns an ordering of $\{G, R, I\}$. This tool operates over the individual spectral band texture histograms. The band with the greatest histogram mode is preferred. For $a, b \in \{G, R, I\}$:

$$a \succ b \iff \arg\max_d(\mathcal{V}_a(d)) > \arg\max_d(\mathcal{V}_b(d)).$$  (13.7)

To illustrate the operation of the five tools described above, Fig. 13.3(b) shows the binary domination images and texture histograms for a 40 x 40 spatial segment drawn from near the centre of the Edinburgh image at location (300, 300). The luminosity spatial extent tool yields 45%, 48% and 7% domination for each of infrared, red and green frequency bands. The band ordering is $R \succ I \succ G \succ B$ which experimental data suggests could indicate a suburban area.

The qualitative sensor output is called a sensor cue and is a transfiguration of a set of quantitative observations made by the sensor. The sensor cue is a tuple comprising the tool descriptor, a representation of the observations and a label denoting the interpretation of the representation.
So, for example, a sensor cue for the luminosity spatial extent tool output above, described in the language LISP, is:

\[
(:tool 'LD :rep '(R I G B) :interp 'IR-R-G-B-ordering) .
\]

### 13.3 Reconfigurable Sensors and Signal Processing Tools

Sensor selection is necessary when the number of sensors on platforms is prohibitively large for data processing [Doyle et al 92]. Further, the preponderance of available signal processing tools requires that the sensors themselves are tailored to various tasks by selecting appropriate signal processing tools [Henderson 84; Henderson et al 88]. The sensor system should reconfigure itself so that it is complete (i.e. a-priori no objects are excluded) and this can be achieved in one of three ways: by learning when to reconfigure according to context; by recognising that a sensor (or tool) is not performing to specification or by reasoning about the problem domain.

*Learning* which sensors to use in various situations requires only basic statistical information obtained by observing the likely success rates of sensors. Such methods are generic, although they do require guaranteed complete experience of the problem domain. A more sophisticated statistical approach is to reason explicitly about each sensor but in such away that the reasoning is independent of the problem domain. Many sensors have generic (domain independent) performance signatures and it is possible to recognise when a tool is under performing. For example, some indication of the performance of the Kalman filter can be determined by continuously monitoring its innovation profile. Doyle's SELMON system [Doyle et al 92] uses a number of sensor-independent information measures to determine a *sensor importance score* - an importance ordering on the set of sensors. These scores, along with causal models, determine when and which sensor output is presented to the human system monitor. In contrast, Kastella [Kastella 97] proposes a method for object discrimination which chooses which sensor to apply based on the likelihood of an object given previous observations. The method chooses that sensor which maximises the Kullback-Leibler information (or cross-entropy) between the current estimate of the object probabilities and the expected estimate obtained from a sensor observation. In Hager's task-directed sensor planning system for geometric sensing problems [Hager and Mintz 91] both the discrimination
gained and the cost of gathering and fusing information are figured into their method for selecting sensors and their placement. Deliberative methods require an understanding of the problem domain in order to anticipate the behaviour of sensors and their tools in unfamiliar situations. Deliberative systems are appropriate when insufficient prior experience is available to construct probabilistic models of the problem domain. We will not consider deliberative systems here (see [Reece 00]).

Our approach learns to recognise when to adapt tool sets using probabilistic models of the problem domain. There are two adaptation tasks for such a system. Firstly, models of the image concepts are adapted by (re-)clustering cues. Secondly, signal processing tools are reconfigured in order to distinguish these clusters. The concept formation and tool selection algorithms will be described in subsequent sections. The remainder of this section explores the relationship between features and their sensor descriptors and asks the question “How should cues be clustered and which features in an image singled out for optimal feature discrimination”?

In order to answer this question we need to consider multi-sensor systems and, for this purpose, introduce the idea of the aspect descriptor. A feature may look different when viewed in different orientations or under different weather conditions. An aspect descriptor is a multi-sensor description of a feature from one such vantage point. The complete set of all possible aspect descriptors for a feature constitutes a cue-centred model of the feature.

Definition 13.1 Assume access to sensors $S$ and signal processing tools $T$. Each sensor and tool is identified by a numeric index.

- A simple qualitative sensor cue from signal processing tool $t \in T$ is denoted $q_t$. All possible qualitative outputs from a sensor-tool $T(i)$ are collectively denoted $C_{T(i)}$.
- An aspect descriptor $ss$ is an $n$-tuple formed from the qualitative output from $n$ signal processing tools $T \subseteq T$ ($|T| = n$): $ss \in C_{T(1)} \times \ldots \times C_{T(n)}$.
- A feature model $S$ is a set of aspect descriptors $S = \bigcup_i ss_i$.
- The function $\text{tools}(S)$ returns the signal processing tools which observe the cues in feature model $S$:

$$\text{tools}(S) = \{ t : q_t \in S \}.$$  \hspace{1cm} (13.8)
An observed cue admits only a subset of all possible aspect descriptors. When cues are randomly distributed between aspect descriptors the number of aspect descriptors admitted by \( n \) observed cues is an exponential decreasing function of \( n \) in general (see Fig. 13.4(a)). Fig. 13.4(b) shows the average number of feature models admitted by ensembles of aspect descriptors when the aspect descriptors are randomly clustered into feature models. Combining these two graphs yields the number of feature models admitted by various numbers of observed cues (see Fig. 13.4(c)) and it is plain to see that the number of admitted feature models undergoes a phase transition. The reason for this phase-effect is that cues, when shared between different admitted feature models, are uninformative. The phase transition occurs when the aspect descriptors with a preponderance of intersegment shared cues are eventually filtered. The phase effect phenomenon is not restricted to randomly clustered feature models.

**Example** Binary sensors detect the presence or absence of a property and a binary aspect descriptor is a list of \( n \) truth values indicating the presence or absence of properties from \( n \) sensors. Suppose that all aspect descriptors have equal observation likelihood and no sensor is used more than once. From an initial ensemble of \( N \) aspect descriptors, the number of descriptors consistent with \( t \) observations is:

\[
d_t = (N - 1)0.5^t + 1 .
\]  

(13.9)

Suppose that the aspect descriptors are initially assigned randomly to \( m \) feature models. An expression for the probability that \( n \) feature models are consistent with \( d \) randomly chosen aspect descriptors can be obtained iteratively:

\[
Pr_d(n) = \left[1 - \frac{n-1}{m}\right] Pr_{d-1}(n-1) + \frac{n}{m} Pr_{d-1}(n) .
\]  

(13.10)

Thus, the average number of feature models consistent with \( d \) aspect descriptors is:

\[
\bar{n}_d = \frac{1 - K^d}{1 - K}
\]  

(13.11)

where \( K = 1 - \frac{1}{m} \). Combining Eqs. 13.9 and 13.11, the number of feature
models consistent with $t$ observations is:

$$\bar{n}_t = \frac{1 - K^{[(N-1)0.5^t+1]}}{1 - K}. \quad (13.12)$$

Naturally, for feature discrimination tasks the concept formation and tool selection algorithms should configure the system software to exploit the phase-phenomenon to the full. To do this the algorithms should endeavour to ensure maximal dissimilarity between feature models and maximal similarity within feature models. An optimal Bayesian concept formation algorithm is developed in the next section.
13.4 A Sensor-Centred Image Segmentation Algorithm

A segmentation is valuable only if the signal processing tools can distinguish different segments. Since the choice of tools is unknown by the human system designer a dynamic, sensor-centred segmentation method is required. But there is a problem. The most appropriate segmentation depends on the type of signal processing tools used and the most appropriate tools depend on the segmentation itself. To overcome this reflexive problem each tool votes on the most appropriate segmentation. The segmentation with the most votes wins. Of course, a segmentation which is heavily subscribed to will exhibit some redundancy in the signal processing tools. The segmentation algorithm developed in this section includes a method for filtering redundant tools.

There are two possible approaches to unsupervised clustering. The standard K-means approach clusters spatial segments directly based on neighbourhood and cue variation relationships. Alternatively, the sensor cues themselves can be clustered and then spatial segments can be formed which correspond, in some way, with the highly auto-correlated cue clusters. The latter method is used here as it is more efficient; there are 142 different spatial segment types (i.e. aspect descriptors) to cluster in the Edinburgh image but only 39 different cues.

The clustering algorithm takes as input a joint probability distribution $Pr(z, z')$ over the full range of observable cues $z$ and $z'$. This distribution is the expected likelihood that two sensor cues are observed together in an ensemble of local neighbourhoods within the image. The ensemble of local neighbourhoods is defined by a sliding window $W$ which traverses the image along a number of stripes (our experiments use a window of size $40 \times 40$ pixels and the training stripes shown in Fig. 13.2).

$$ Pr(z, z') = E_W Pr(z, z' | W) $$

$$ = E_W \left( \frac{n_z(W)n_{z'}(W)}{\sum_{z \in W, z' \in W} n_z(W)n_{z'}(W)} \right). $$

(13.13)

(13.14)

where $n_z(W)$ is the number of occurrences of cue $z$ in window $W$.

Cues are clustered in such a way that each cluster, $G$ say, is as distinct from its background $\neg G$ as possible. Using Bayes' Rule the relative likelihood of $G$ and $\neg G$ given (conditionally independent) observations $z_1$ and
\[ z_2 \text{ is:} \]
\[
\frac{Pr(G \mid z_1, z_2)}{Pr(\neg G \mid z_1, z_2)} = \frac{Pr(z_1 \mid G) Pr(z_2 \mid G) Pr(G)}{Pr(z_1 \mid \neg G) Pr(z_2 \mid \neg G) Pr(\neg G)}. \tag{13.15}
\]

Optimally distinct clusters \( \mathcal{S}_1 \) are those which maximise the expected relative likelihood of \( G \) and \( \neg G \) over all observations and clusters \( G \). When \( \mathcal{C} \) is the set of all possible cue clusterings then:

\[ \mathcal{S}_1 = \operatorname{argmax}_{\mathcal{G} \in \mathcal{C}} E_{G \in \mathcal{G}, z} \left( \log \frac{Pr(z \mid G)}{Pr(z \mid \neg G)} \right). \tag{13.16} \]

The optimal clustering \( \mathcal{S}_1 \) maximises the difference between the average cross-entropies of \( Pr(z \mid G) \) and \( Pr(z \mid \neg G) \) over all clusters and the average entropies of \( Pr(z \mid G) \) over all clusters. Thus, optimally distinct clusters maximise inter-cluster variation while minimising internal cluster variation as required by Section 13.3.

When extra constraints are applied to the segmentation task, such as image spatial extent limitations, the cue clusters can be manipulated further to conform to these constraints. The aerial image task specifically requires that segments are observed with approximately equal likelihood so that features are not too localised making them hard to find and also not too expansive making localisation difficult. The probability \( Pr(G) \) is an approximate measure of the likelihood of encountering the spatial segment associated with cluster \( G \). When \( N \) such segments are encountered with equal likelihood \( \sum_G Pr(G) \log Pr(G) \) is maximised. Incorporating the spatial segment constraint into Eq. 13.16 gives:

\[ \mathcal{S}_2 = \operatorname{argmax}_{\mathcal{G} \in \mathcal{C}} E_{G \in \mathcal{G}, z} \left( \log \frac{Pr(z \mid G)}{Pr(z \mid \neg G)} - K \sum_G Pr(G) \log Pr(G) \right), \tag{13.17} \]

where \( K \geq 0 \) (in our experiments \( K = 1 \)). Rewriting:

\[ \mathcal{S}_2 = \operatorname{argmax}_{\mathcal{G} \in \mathcal{C}} \sum_{G \in \mathcal{G}, z} Pr(z, G) \log \left( \frac{Pr(G \mid z)[1 - Pr(G)]}{[1 - Pr(G \mid z)]Pr(G)^{1 + \frac{K}{\log N}}} \right), \tag{13.18} \]

where \( Pr(G \mid z) = \sum_{z' \in \mathcal{G}} Pr(z, z') \sum_{z''} Pr(z', z'') \) and \( Pr(G) = \sum_{z' \in \mathcal{G}, z} Pr(z, z') \).

The image is segmented in one-to-one correspondence with the cue clusters. All pixels in the image assigned to a cue cluster are deemed to belong to the same spatial segment. A pixel is assigned to a cluster \( G \) when, for
some neighbourhood $W$ about the pixel:

$$G = \arg\max_{G \in \mathcal{G}} \sum_z \Pr(z \mid W) \log \frac{\Pr(z \mid G)}{\Pr(z \mid \neg G)}.$$ \hfill (13.19)

Fig. 13.2 shows the result of unsupervised clustering of the Edinburgh image. The image is segmented into sea, dock area and city and includes the large scale park structures in the middle of the city.

Straightforward variations of the segmentation algorithm exist for when more control over the segment properties (e.g. number, composition) is required. For example, the algorithm can include a constraint fixing the number of clusters which can, in turn, be used to produce a hierarchical decomposition of the image by applying Eq. 13.16 or Eq. 13.18 recursively to the individual segments themselves. Further, when the image is (partly) partitioned a-priori or multi-sensor associations exist between a-priori segmented and unsegmented images, $\Pr(z \mid S)$, the conditional probability of cue $z$ within some spatial segment $S$, can be determined by a window which covers each segment $S$ in turn. In which case, $\Pr(z \mid S)$ replaces $\Pr(z \mid G)$ in Eq. 13.19.

### 13.5 Signal Processing Tool Selection Strategies

Cue-clusters can be reduced in size because sensor cues may be ignored when they are either uninformative or they are redundant.

**Inter-cluster discrimination:** ignore cue $z$ which has a weak fidelity $\frac{\Pr(G \mid z)}{\Pr(G)} \approx 1$ with all clusters $G$.

$$\frac{\Pr(G \mid z)}{\Pr(G)} \approx 1 \Rightarrow \Pr(z \mid G) \approx \Pr(z \mid \neg G) \quad \text{[or } \Pr(G) \approx 1 \text{]}.$$ \hfill (13.20)

**Intra-cluster cue redundancy:** ignore cue $z_1$ when it is significantly corre-

---

\(\text{§The raw quantitative sensor data is initially subjected to COMOC morphological filtering to reduce the number of different sensor cues.}\)
lated with some other cue $z_2$ in the same cluster:

$$\frac{Pr(z_1, z_2)}{Pr(z_1)Pr(z_2)} \approx \min \left\{ \frac{1}{Pr(z_1)}, \frac{1}{Pr(z_2)} \right\}.$$ (13.21)

When all cues from a sensor are ignored then the sensor is redundant and can be deselected from the sensor configuration. Intra-cluster cue redundancy often offers many alternative ways in which the sensor configuration can be reduced in size. Denote as $S_i$ the set of all possible reduced cue clusters for cluster $i$ obtained by filtering redundant cues and filtering cues with weak fidelity. Members of $S_i$ are called reduced cue clusters.

**Definition 13.2** When $a$ and $b$ are two cue clusters, $S_a$ and $S_b$ are alternative reduced clusters for $a$ and $b$ respectively and $t$ is a set of tools $t \subseteq \mathcal{T}$ then $t$ is a sufficient tool suite $STS$ for distinguishing $a$ and $b$ when:

$$STS(t, a, b) \iff (\exists s_a \in S_a, s_b \in S_b) t \supseteq \text{tools}(s_a) \cup \text{tools}(s_b).$$

A sufficient tool suite $t$ is a minimal tool suite for distinguishing two clusters if there is no other sufficient tool suite that is a subset of $t$.

**Definition 13.3** When $a$ and $b$ are two cue clusters, $t \subseteq \mathcal{T}$ and $t' \subseteq \mathcal{T}$ are sufficient tool suites, $STS(t, a, b)$ and $STS(t', a, b)$, then $t$ is a minimal tool suite when:

$$\neg (\exists t') t' \subset t.$$  

Two heuristics are available to choose between possible reduced clusters. They are used when minimal adaptation of tools is required between clusters or when static, minimal but globally sufficient tool ensembles are required.

**Definition 13.4** A suite of tools $t$ is a global tool suite for cue clusters $C$ if:

$$(\forall a, b \in C) STS(t, a, b).$$

When $t$ and $t'$ are global tool suites then $t$ is a minimal global tool suite when:

$$\neg (\exists t') t' \subset t.$$ 

*The effect on $Pr$ by removing a sub-cluster $C_2$ from a cluster $C = C_1 \cup C_2$ is:

$$Pr(C \mid z) - Pr(C_1 \mid z) = Pr(C_2 \mid z)[1 - Pr(C_1 \mid C_2)].$$
The **minimal change tool suite** is the minimal set of tools which can be used to distinguish image segment transitions and also require minimal change between segments.

**Definition 13.5** When \( \tau = \{ "S_a \rightarrow S_b" \} \) is the set of all possible cluster transitions, \( t_{ab} \) and \( t'_{ab} \) are minimal tool suites for the transition \( S_a \rightarrow S_b \), then \( t \in \mathcal{T} \) is a **minimal change tool suite** if \( t = \bigcup_{a,b} t_{ab} \) and for all other tool suites \( t' = \bigcup_{a,b} t'_{ab} \):

\[
\sum_{"S_a \rightarrow S_b" \in \tau, "S_b \rightarrow S_c" \in \tau} |t_{ab} \setminus t_{bc}| + |t_{bc} \setminus t_{ab}|
\leq \sum_{"S_a \rightarrow S_b" \in \tau, "S_b \rightarrow S_c" \in \tau} |t'_{ab} \setminus t'_{bc}| + |t'_{bc} \setminus t'_{ab}|.
\]

To illustrate the efficacy of the reduced sensor suite, Fig. 13.5 shows a tri-segmented image of Brighton. In part (a) the full range of signal processing tools taken from section 13.2 (i.e. LD, LV, LM, TE and TM) are used to discriminate the clusters. In part (b), a reduced set of tools (i.e. LD, TE and LM) exhibits a near identical segmentation of the image to that of part (a).

![Fig. 13.5 Segmented image of Brighton, including segments 1 (city), 2 (South Downs) and 3 (English Channel): (a) full sensor suites for 1 = \{LD, LV, LM, TE, TM\}, 2 = \{LD, TE, TM, LM\} and 3 = \{LD, TE, LV\} and (b) reduced sensor suites for 1 = \{LD, TE, LM\}, 2 = \{LD, TE, LM\} and 3 = \{LD\}](image_url)
13.6 Dynamic Signal Processing Tool Scheduling

As observations accrue segment likelihoods change and segment hypotheses can be filtered simply by the fact that they are extremely unlikely given the observations to date. The remaining possible segment hypotheses can be ranked in order of likelihood. Such a ranking can be used to anticipate and select the most informative signal processing tools from the minimum tool suite. However, simply using the most informative sensor is often not enough as it is necessary to ensure that all likely clusters are observable.

Suppose that observations $Z$ have been obtained by any sensor configurations and correspond to some (as yet) unknown cluster $G$. The information $I_S$ that is expected from $n$ new sensor observations $\{z_1, \ldots, z_n\}$ from a specific sensor configuration $S$ of $n$ sensors $\{s_1, \ldots, s_n\}$ is:

$$I_S = \sum_{z_1, \ldots, z_n, G} Pr(G \mid Z) Pr(z_1, \ldots, z_n \mid G, s_1, \ldots, s_n) \log \frac{Pr(z_1, \ldots, z_n \mid G)}{Pr(z_1, \ldots, z_n \mid \neg G)}$$

Assuming $\{z_1, \ldots, z_n\}$ are conditionally independent then:

$$Pr(z_1, \ldots, z_n \mid G, s_1, \ldots, s_n) = \prod_{i=1}^{n} Pr(z_i \mid G, s_i) = \prod_{i=1}^{n} \frac{Pr(z_i \mid G)}{Pr(s_i)}$$

and:

$$Pr(z_1, \ldots, z_n \mid G) = \prod_{j=1}^{n} Pr(z_j \mid G). \quad (13.22)$$

Thus:

$$I_S = \sum_{z_1, \ldots, z_n, G} Pr(G \mid Z) \left[ \prod_{i=1}^{n} \frac{Pr(z_i \mid G)}{Pr(s_i)} \right] \left[ \sum_{j=1}^{n} \log \frac{Pr(z_j \mid G)}{Pr(z_j \mid \neg G)} \right]$$

It can then be shown that:

$$I_S = \sum_{G} Pr(G \mid Z) \left[ \sum_{i=1}^{n} \sum_{z_i} Pr(z_i \mid G) \frac{Pr(z_i \mid G)}{Pr(s_i)} \log \frac{Pr(z_i \mid G)}{Pr(z_i \mid \neg G)} \right] \quad (13.23)$$

The value of $Pr(z_i \mid G)$ can be obtained from the joint probabilities $Pr(z_i, z_j)$ provided $Pr(s_i)$ is the rate at which sensor $i$ is sampled. When $n$ sensors are used an equal amount to gather the training data for $Pr(z_i, z_j)$ then $Pr(s_i) = \frac{1}{n}$. 
The most informative sensor suite $S$ satisfies:

$$S = \arg\max_S \left\{ \frac{I_S}{n_S} \right\}$$

(13.24)

where $n_S$ is the number of sensors in the sensor suite. The denominator is included as sensor suites may vary in size and, in such cases, it is prudent to maximise the expected information offered by each observation.

The efficacy of sensor scheduling for feature discrimination is illustrated by Fig. 13.6. Each trajectory in Fig. 13.2(b) is traversed by two sensing systems; the first employs the scheduling algorithm described above which selects tools with the greatest expected utility while the second algorithm chooses a random tool reconfiguration (each time). In both cases low probability cue clusters are filtered and the tools are chosen from minimal global tool suites for the remaining clusters. Within each spatial segment along a trajectory the selected tools measure observations $z$ and the average value of $\log \frac{Pr(z|G)}{Pr(z|\neg G)}$ for these tools is recorded. The sum $\sum$ of these averaged values is recorded as more observations are made as the sensing systems traverse the spatial segment. Fig. 13.6 represents the expected value of $\sum$ over all spatial segments and trajectories. The graphs show the "probabilistic" counterpart $Pr(G)$ for $E(\sum)$ where $Pr(G)$ satisfies $\log \frac{Pr(G)}{Pr(\neg G)} = E(\sum)$.

![Fig. 13.6 True-positive segment identification probabilities $Pr(G)$ after sampling "# locations" many locations of the spatial segment. Solid and dashed lines show the true-positive rates for the dynamic tool selection algorithm and the randomly chosen global minimum tool algorithm respectively.](image)
13.7 Conclusions

Our research is aimed at developing a sensor management and monitoring system which is able to guide the reconfiguration and interpretation of sensors and their image processing tools. This chapter has developed information theoretic sensor management strategies for infrared aerial image interpretation. An image ontology is determined from a-priori unsegmented images and then infrared sensors are configured for optimal discrimination of the concepts within each image. Image concept primitives are obtained by a clustering algorithm which clusters significantly correlated sensor cues to maintain maximum (Bayesian) discrimination between clusters. Tool configurations are then identified which exhibit the required image feature recognition coverage and also minimal redundancy. The approach allows dynamic scheduling of the most informative tools by anticipating their informativeness using knowledge of the scene obtained from accumulating observations.
APPENDIX A

CONTACT DETAILS OF AUTHORS
Bibliography


B. Ayrulu and B. Barshan. Identification of target primitives with multiple decision-making sonars using evidential reasoning. *International Journal*
Bibliography


Bibliography


Bibliography


Bibliography


D. Langer. *An Integrated MMW Radar System for Outdoor Navigation*. PhD


G. L. Miller and E. R. Wagner. An optical rangefinder for autonomous robot cart navigation. *SPIE, Mobile Robots II*

Bibliography

Opto Technology, INC. *Optoelectronics Data Book, Survey of Photosensitive Materials and Devices*.


Riegl. 3-D laser mirror scanner. http://www.rieogl.co.at/.


This page is intentionally left blank
Appendix A: Contact Details of Authors
Appendix A: Contact Details of Authors

Dr. Penny Probert Smith  
Robotics Research Group  
Department of Engineering Science  
University of Oxford  
Parks Road  
Oxford OX1 3PJ  
UK

Tel: +44-1865-273141  
Fax: +44-1865-273908  
e-mail: pjp@robots.ox.ac.uk  
http://www.robots.ox.ac.uk/ pjp/

Dr. Martin David Adams  
Division of Control and  
Instrumentation  
School of Electrical and Electronic Engineering  
Nanyang Technological University  
Nanyang Avenue  
Singapore 639798

Tel: ++65 790 4361  
Fax: ++65 792 0415  
e-mail: eadams@ntu.edu.sg

Dr. Billur Barshan  
Department of Electrical Engineering  
Bilkent University  
TR-06533 Bilkent, Ankara  
Turkey

Tel: +90-312-2902161  
Fax: +90-312-2664192  
e-mail: billur@ee.bilkent.edu.tr

Graham Brooker  
Australian Centre for Field Robotics  
School of Aerospace, Mechanical and Mechatronic Engineering  
University of Sydney  
Sydney, NSW, 2006, Australia

Tel: +61 2 9351 4023  
Fax: +61 2 9351 7474  
e-mail: gbrooker@acfr.usyd.edu.au  
http://www.acfr.usyd.edu.au

Dr. Andrew J. Davidson  
Robotics Research Group  
Department of Engineering Science  
University of Oxford  
Parks Road  
Oxford OX1 3PJ  
UK

Tel: +44-1865-273149  
Fax: +44-1865-273908  
e-mail: ajd@robots.ox.ac.uk  
http://www.robots.ox.ac.uk/ ajd/
Appendix A: Contact Details of Authors

Dr. Briseida Deyarina Maytorena Sanchez
Robotics Research Group
Department of Engineering Science
University of Oxford
Parks Road
UK
Tel: +44-1865-273154
Fax: +44-1865-273908
e-mail: deya@robots.ox.ac.uk
Oxford OX1 3PJ

Dr. Herbert Peremans
Universiteit Antwerpen
Prinsstraat 13
B-2000 ANTWERPEN
Belgium
Tel: +32-(0)3-2204212
Fax: +32-(0)3-2204799
e-mail: Herbert.Peremans@ua.ac.be
http://www.ufsia.ac.be/ hperemans/

Dr. Stephen Reece
Robotics Research Group
Department of Engineering Science
University of Oxford
Parks Road
Oxford OX1 3PJ
UK
Tel: +44-1865-282186
Fax: +44-1865-273908
e-mail: reece@robots.ox.ac.uk
http://www.robots.ox.ac.uk/ reece/

Robert Taylor
Credit Suisse First Boston
One Cabot Square
Canary Wharf
London E14 4QJ
UK
Tel: +44-20-7888-9891
Fax: +44-20-7888-3924
robert.taylor@csfb.com
*This work was carried out in the Department of Engineering Science, University of Oxford
This page is intentionally left blank
Index

3-D sensors 103,189,194,217,241
absorption coefficient 83
acousto-optic scanner 187
active sensors 7
active sensor management 271,289
active vision 240
adaptive
  adaptive signal processing tools 271,289
  sensing ontologies 271,289
aerial image interpretation 271,289
airborne radar 156
albedo 218
  image 220,221
ambiguous data 194
AMCW - see amplitude modulated continuous wave
amplitude modulated continuous wave lidar 180,182
calibration 208
modules
  amplitude detector 200
  frequency reduction 200
  phase discriminator 199
  range filter 199
  receiver 199
  signal compression 199
  transmitter 199
scanning mechanisms 188,194
angular resolution
optical 172
pixel 157
radar 146
sonar 44,54,79
antenna
  array 32
  directivity 23,25
  efficiency 147
  gain 147
APD - see avalanche photodiode
aperture
  of antenna 26
  shape 29
array
  radar 148,153,156
  sonar 32,89,93,97,103,115,122
architecture 3
  architecture: blackboard 4
  reactive 6
  sensor 7
  subsumption 6
aspect descriptor 279,281,282
attenuation
  atmospheric attenuation of radar 139
  attenuation coefficient 38,84
avalanche photodiode 173,196,197
azimuth angle 106
backscattering coefficient 62,69
backscattered power 38,73
bandpass filter 87
baseline (triangulation) 178
Bayesian image segmentation 282,284
beam
  pattern - see radiation pattern
  splitter 174
  steering 32
binary domination image 275
biomimetic sonar 104
biosonar 81
blackboard architecture 4
Bookstein algorithm 234
bounded conic fit 235
Cassegrain antenna 147
CCD see charge coupled device
  CCD camera 239
charge coupled device 178
circular
  aperture 26
  sonar array 122
classifiers 75
cluster fidelity 284
cochlear processing 87
collecting optics 168
colour camera 240
coherece length 184
communication system 166
computational cost
  119,133,143,146,241
context based navigation 266
continuous feature tracking 259
continuous time frequency
  modulation
    radar 143
    sensor 59
    sonar 56
corner 131
corner target 47,63,85,90,94,98
correlation receiver 52,55,81
correspondence see data association
crosstalk 202
CTFM - see continuous time
  frequency modulation
curved reflectors 101,117
curve fitting
  straight line 223
  ellipse 228
curves 119
cylinder reflection model 36
data association 9,31,65
decision criterion 93,97,101
diffuse reflection 33,171
dilation operator 117
directivity 23,25,146
  circular aperture 27,29,83
  rectangular aperture 30
Doppler
  radar 151
  sonar 59
dual polar receiver 154
dynamic range 50,159,196,199
  compression 199
echo
  formation 22,82
  measurement 51,90,94,98,105
  modelling 89,93
  threshold detection 42
date 131
date target 47,63,86,94,98
EKF see Extended Kalman filter
Extended Kalman filter 10,18,254
  introduction 10
  in edge detection 228
  initialisation 231
  recursive form 231
  in mapping and localisation
    18,65,228,254
electronic scanning: radar 148
electrostatic transducer 24
ellipse fitting 233
energy distribution from rough
  surface 76
epipolar line 251
erosion operator 117
erro measure
  extended Kalman filter 66,257
  morphological processing 118,120
evolutionary robotics 6
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eye safety</td>
<td>173</td>
</tr>
<tr>
<td>false alarms</td>
<td>152</td>
</tr>
<tr>
<td>far field</td>
<td>38,146</td>
</tr>
<tr>
<td>relation to aperture illumination</td>
<td>29</td>
</tr>
<tr>
<td>fast Fourier transform resolution</td>
<td>59,143,145</td>
</tr>
<tr>
<td>feature</td>
<td></td>
</tr>
<tr>
<td>based maps</td>
<td>112</td>
</tr>
<tr>
<td>choice</td>
<td>262</td>
</tr>
<tr>
<td>extraction</td>
<td>76,227</td>
</tr>
<tr>
<td>matching</td>
<td>247</td>
</tr>
<tr>
<td>primitives</td>
<td>151</td>
</tr>
<tr>
<td>quality</td>
<td>261</td>
</tr>
<tr>
<td>type</td>
<td>13,242</td>
</tr>
<tr>
<td>FFT - see fast Fourier transform</td>
<td></td>
</tr>
<tr>
<td>field of view</td>
<td>196</td>
</tr>
<tr>
<td>filter model</td>
<td>83</td>
</tr>
<tr>
<td>fixate</td>
<td>241</td>
</tr>
<tr>
<td>fixation</td>
<td>251</td>
</tr>
<tr>
<td>fixation strategy</td>
<td>261</td>
</tr>
<tr>
<td>FMCW - see frequency modulated continuous wave</td>
<td></td>
</tr>
<tr>
<td>focused radar</td>
<td>33</td>
</tr>
<tr>
<td>focused sound</td>
<td>33</td>
</tr>
<tr>
<td>frequency modulated continuous wave</td>
<td></td>
</tr>
<tr>
<td>interrupted</td>
<td>158</td>
</tr>
<tr>
<td>lidar</td>
<td>180,184</td>
</tr>
<tr>
<td>scanning</td>
<td>189</td>
</tr>
<tr>
<td>radar</td>
<td>143,144,153</td>
</tr>
<tr>
<td>sonar</td>
<td>56</td>
</tr>
<tr>
<td>frequency modulated sonar</td>
<td>54</td>
</tr>
<tr>
<td>frequency reduction electronics</td>
<td>200</td>
</tr>
<tr>
<td>Fresnel's law</td>
<td>195</td>
</tr>
<tr>
<td>gain</td>
<td></td>
</tr>
<tr>
<td>antenna</td>
<td>25</td>
</tr>
<tr>
<td>relationship to receiving area</td>
<td>23,147</td>
</tr>
<tr>
<td>geometric primitives</td>
<td>47,63</td>
</tr>
<tr>
<td>ghost target</td>
<td>66,152</td>
</tr>
<tr>
<td>global tool suite</td>
<td>285</td>
</tr>
<tr>
<td>heterogeneous surface</td>
<td>68</td>
</tr>
<tr>
<td>higher order reflection</td>
<td>120</td>
</tr>
<tr>
<td>holographic scanner</td>
<td>187</td>
</tr>
<tr>
<td>horn antenna</td>
<td>147,153,154</td>
</tr>
<tr>
<td>hybrid occupancy grids</td>
<td>112</td>
</tr>
<tr>
<td>IID see inter intensity difference</td>
<td></td>
</tr>
<tr>
<td>image patches</td>
<td>244</td>
</tr>
<tr>
<td>information rate: sonar</td>
<td>80</td>
</tr>
<tr>
<td>infra-red</td>
<td></td>
</tr>
<tr>
<td>aerial images</td>
<td>272,274,276</td>
</tr>
<tr>
<td>sources</td>
<td>125,188,206</td>
</tr>
<tr>
<td>wavelength</td>
<td>9</td>
</tr>
<tr>
<td>initialisation</td>
<td>255</td>
</tr>
<tr>
<td>innovation</td>
<td>232</td>
</tr>
<tr>
<td>integrated phased-array antenna</td>
<td>153</td>
</tr>
<tr>
<td>inter intensity difference</td>
<td></td>
</tr>
<tr>
<td>contours</td>
<td>107</td>
</tr>
<tr>
<td>inter-aural intensity differences</td>
<td>105</td>
</tr>
<tr>
<td>maps</td>
<td>108</td>
</tr>
<tr>
<td>interference of light</td>
<td>203</td>
</tr>
<tr>
<td>interrupted FMCW</td>
<td>158</td>
</tr>
<tr>
<td>IPA- see integrated phased array</td>
<td></td>
</tr>
<tr>
<td>joint mapping and localisation</td>
<td></td>
</tr>
<tr>
<td>problem</td>
<td>15</td>
</tr>
<tr>
<td>KASPA sensor</td>
<td>30</td>
</tr>
<tr>
<td>K-nearest neighbour</td>
<td>75</td>
</tr>
<tr>
<td>Lambertian reflectivity</td>
<td>218</td>
</tr>
<tr>
<td>Lambertian source</td>
<td>172</td>
</tr>
<tr>
<td>Lambert's law</td>
<td>172</td>
</tr>
<tr>
<td>landmark based mapping</td>
<td>46,244</td>
</tr>
<tr>
<td>landmarks</td>
<td></td>
</tr>
<tr>
<td>sonar</td>
<td>46,47</td>
</tr>
<tr>
<td>visual</td>
<td>244</td>
</tr>
<tr>
<td>laser diode</td>
<td>172,180,195</td>
</tr>
<tr>
<td>lateral effect photodiode</td>
<td>178,224</td>
</tr>
<tr>
<td>sensor</td>
<td>224</td>
</tr>
<tr>
<td>lead-lag filter</td>
<td>213</td>
</tr>
<tr>
<td>LED - see light emitting diode</td>
<td></td>
</tr>
<tr>
<td>lenses</td>
<td>33</td>
</tr>
<tr>
<td>LEP - see lateral effect photodiode</td>
<td></td>
</tr>
<tr>
<td>lidar</td>
<td>180,226</td>
</tr>
<tr>
<td>amplitude modulated</td>
<td>182,188,195ff</td>
</tr>
<tr>
<td>comparison of techniques</td>
<td>191</td>
</tr>
<tr>
<td>design factors</td>
<td>195</td>
</tr>
<tr>
<td>frequency modulated</td>
<td>184,189</td>
</tr>
<tr>
<td>Perceptron scanner</td>
<td>226</td>
</tr>
<tr>
<td>pulse modulated</td>
<td>181</td>
</tr>
<tr>
<td>SICK scanner</td>
<td>188,226</td>
</tr>
</tbody>
</table>
light emitting diode 172,180,195
linear array: sonar 121
line features 250
line fitting 228
initialisation 231
polar co-ordinates 230
line target 47,84,90
localisation 9,17,254
localisation strategy 261
local planner 9
loss
atmospheric 38.139
spreading 38
transmission 84
mapping
.. and localisation 241
batch and sequential 14
experiments 64,132,254
heterogeneous environment 72
representation
feature based 63ff
landmark based 244ff
occupancy grid 44,112,151,157
rough environment 72
smooth environments 66
matched filter 49,52,140
as correlator 52
measurement density 45,46
mechanical scanning
optical 188
radar 148
memory requirement 133,151,241
millimetre wave
antenna 25
attenuation 37
interference effects 21
practical sensors
radar - see radar: millimetre wave
sonar - see sonar
target properties 33
minimal tool suite 285
mixed pixels (lidar) 202,205
mixer electronics
CTFM (FMCW) sonar 58
lidar 200
radar 144
modulation
index 198
types 51,141,142,180,182
monochrome camera 240
morphological filtering 47
as tool for mapping 113
computational cost 119
description 117
dilation operator 117
erosion operator 117
error measure 120
pruning 118
results 122
thinning 118
multi-aural sensing 88
multiple echoes: sonar 48,114
multiple path reflections
lidar 205,207
sonar 194
navigation error estimation
67,161,252,258,262
near field 25,146
neural network 75
neural transduction 87
noise
phase 214
Nomad mobile robot 46,122,125
NTSC video standard 239
occupancy grid 44,112,151
volume cell representation 152
optoelectronic sensor
beam radius 208
design 166
detector 173,178
emitter 165,172,178
flux 171
foot print 202
glomery 165,174
lidar 180
optical range sensor 8
optical reflection model
diffuse 169
specular 167
power 169,172
radiometric design 166
RAM 178
reflection models 168
scanning methods 186
transducers 165
triangulation 177,225
orientation measurement 49,50,64,79,90,94,98,101,105
orientation signature 50
PAL video standard 239
patch antenna 32
patch operator 245
Perceptron laser scanner 226
phase
discriminator 199
noise 214
phenomenon 279,281
shift
as function of time 208
dynamic 212
photodetector 166
photodiode 173
photo- multiplier 173
photo sensitive detector 224
piezo actuator 189
PIN photodiode 173
pixel: CCD 239
pixel resolution 157
planar reflector characteristics 35,36,63,84,94,98
planar target see planar reflector
point features 250
corners and edges 47,63,90,94,98,99,105
Polaroid sensor 24,41,79,114
potential field planning method 45
primitive length encoding 277
processing speed 118,133,143,146,154
pruning 118
pulse chirp 54
pulse compression
radar 141
sonar 54
Pulsed-Doppler radar 151
pulsed modulation lidar 181
pulse to pulse stepped mode 142
qualitative sensor cue 277,279
radar
equation 22
wavelength dependence 8,9,21
radar: millimetre wave
atmospheric attenuation 139
features 160
image characteristics 150
resolution 140
scanner 156
systems in field 151,153,154
transducers 24
when to use 138
radiance 168,169
radiation pattern 25,28,114,182
circular aperture 26
KASPA 31
parabolic dish 146
rectangular aperture 30
radiometric design 166
random sampling, RANSAC 235
range
drift 203
errors in lidar 204,205
filter 199
images and maps
lidar 218,220,221,237
radar 149,150,155,159
sonar 48,49,51,67,73,76,75,108,116,123ff,1
structured light system 132
measurement 64,79
resolution 59,143,144,156,179
signature 76
ray model 84
RCD - see region of constant depth
reactive architectures 6,7
reflection
diffuse 33,171,195
Lambertian rule 172,218
simultaneous 205
specular
            33,34,44,84,129,152,168,195,196
region of constant depth 34,47,116
resolution
    angular resolution 21,79
    FFT resolution 59,143,145
    radar angular resolution 146
    required (radar) 156
    stereo measurements 252
RoBat 104
rough surface 48,51,68,171
    model 35,69
SAR - see synthetic aperture radar
    122,149
saturation
    optical 175
    sonar 50
scanning
    design for radar 148,152
    distortion 202
    electronic 32,148
    mirror 175,179,180,187,189,190
        piezoactuator 189
        synchronised 179,202
    optical geometries 226
        acousto-optic 187
        holographic 187
    speed 212
    systems 186
scattering cross section 22,36
segmentation 223,232,274
sensing strategy 10,271
sensor architecture 7,9
sensor fusion 13
Sensus
    sonar ranging system 126
    structured light system 127
sequential mapping 16
sequential probability ratio 101
SICK sensor 188,226
side-lobe 28,32,147
signal compression 199
signal processing tool scheduling
    287,288
signal to noise ratio 197
simultaneous reflection 205
single return sonar 41
smooth surface - see reflection,
    specular
solid angle 168
sonar
    arc 116
    array 32,115
    echo formation 22,81,114
    electrostatic transducer 24
    equations 23,83
    general characteristics 8,21,44,79
    modulation 51
    resolution 44,54
    ring 43
    systems in field
        biomimetic sensor head 104
        KASPA sensor 30
        Polaroid sensor 24,41,79,114
        Sensus sonar ranging system
            126
        Witt sensor 59
speckle 150,203
spectrogram 88
specular reflection
            33,34,44,84,129,152,168,195
spherical spreading 83
state 10
steering control 266
stepped frequency radar 142
steradian 168
stereo
    bio-sonar head 104
    camera head 243
    KASPA sonar sensor 32
    stochastic mapping 18
    structured light surface profile 129
    surface profile extraction 111,128
    synchronised scanning 179
    synthetic aperture radar 122,149
target
  geometry coefficient 63
  position: IID 108
  recognition 50,72,91,95,99,103
task driven approach 240
temporal coherence 203
texture 72,75,276
  classification 75
thinning 118
threshold error 42
time-bandwidth dependence 54,140
transmitter
  electronic 199
  radiant power 197
triangulation 46,177,224
ultrasonic - see sonar
uncertainty ellipse 66,257
validation gate 232
viewpoint dependence 248
virtual transmitter 95
volume cell representation 152
waypoint navigation 160
world model 4
wrap around geometry 218,220
Yorick head 242
This book describes recent work on active sensors for mobile robots.
An active sensor interacts with its surroundings to supply data on demand for a particular function, gathering and abstracting information according to need rather than acting as a generic data gatherer. Details of the physical operation are hidden.

The book deals mainly with active range sensors, which provide rapid information for local planning, describing extraction of two-dimensional features such as lines, corners and cylinders to reconstruct a plan of a building. It is structured according to the physical principles of the sensors, since to a large extent these determine the function of the sensors and the methods of processing. Recent work using sonar, optoelectronic sensors and radar is described. Sections on vision and on sensor management develop the idea of software adaptation for efficient operation in a changing environment.