Underwater Shape Reconstruction in Two Dimensions

by

Bradley Alfred Moran

S.M. Naval Architecture and Marine Engineering (February 1990)
S.M. Mechanical Engineering (February 1990)
Massachusetts Institute of Technology

A.B. Mathematics Yale University (May 1985)

Submitted to the Department of Ocean Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1994

Author .....
Department of Ocean Engineering
March 18, 1994

Certified by .....
Chryssostomos Chryssostomidis
Professor of Ocean Science and Engineering
Thesis Supervisor

Accepted by .....
A. Douglas Carmichael
Chairman, Departmental Committee on Graduate Students

Underwater Shape Reconstruction in Two Dimensions

by

Bradley Alfred Moran

Submitted to the Department of Ocean Engineering
on March 18, 1994, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Panoramic sweeps produced by a scanning range sensor often defy interpretation using conventional line of sight models, particularly when the environment contains curved, specularly reflective surfaces. The combination of multiple scans from different vantage points provides the necessary geometric constraints to solve this problem, but not without introducing its own difficulties. How do features in the data respond to the motion of the vantage point and to what features in the world to they correspond? Existing multiple scan implementations, for the most part, ignore the correspondence issue. We pose the data interpretation problem as one of tracking sequences of observations that occur chronologically, in one scan after another.

The algorithm developed to address these issues contains three parts. The first part applies a model of the sensor, first to define a canonical observation and then to detect instances of it in the raw scans. The second part uses a model-based multiple hypothesis filter to track observations. In doing so, it identifies topological features in the world and assigns measurements to them. The tracking filter provides low order estimates of local geometric shape in an efficient manner suitable for real time applications. The final part synthesizes observations that share a common origin into higher order representations which approximate the global geometry.

We apply the algorithm to real scans generated during a series of experiments performed for this investigation. The experiments consist of interrogating a variety of two dimensional prismatic objects, standing on end in a 1.2 m deep fresh water tank, from multiple vantage points using a 1.25 MHz profiling sonar system.

Our results reflect the validity of the algorithm under the initial assumptions, and its gradual degradation in performance when these assumptions fail to characterize the environment. We close with recommendations which detail extending the approach to handle more natural underwater settings.

Thesis Supervisor: Chryssostomos Chryssostomidis
Title: Professor of Ocean Science and Engineering

To Alfred Moran, the finest person I shall ever know.

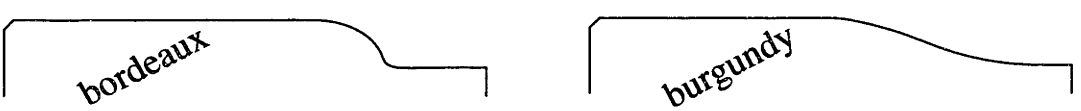
Acknowledgements

The completion of this dissertation ends my time as a graduate student—a long time it has been. Lest I am caught thinking that no more challenges face me, William Faulkner admonishes that

> no battle is ever won. ... They are not even fought. The field only reveals to man his own folly and despair, and victory is an illusion of philosophers and fools. *The Sound and the Fury.*

Yes I have finished my graduate studies, but I still have much to learn.

I'll never forget my first encounter with Professor Chryssostomos Chryssostomidis in his official capacity as my advisor. In our first meeting he proved himself a peerless educator, teaching me a most important lesson that I now share.



I owe my career to you, Chrys, for whatever you saw in me.

I would like to thank my thesis committee. My work reflects the insights, encouragement, and support of Profs. Nicholas Patrikalakis and Henrik Schmidt and Dr. John Leonard. I am especially grateful to John, who as a mentor kept me on track despite my best (and regular) efforts to derail myself and as a friend taught me many things in the process.

The successful completion of my experiments and thesis results from the support and patience of my colleagues, whom I thank: Mike Drooker in the Design Laboratory, Dave Barrett in the Towing Tank, Dr. Tom Consi, Jared Bibler and Jean Nam in the Sea Grant Underwater Vehicles Laboratory, and Dr. Ingemar Cox and Matt Miller at the NEC Research Institute in Princeton, New Jersey.

Most importantly, I thank my family (who are my best friends) Brian, Carol & Michael, Gary, Ronni & Jenna, Mom & Dad, and my friends (who are my extended family) James D., Tanya, Maggie & Bob Schwartz, Helen & Tom Maley, Séamus Tuohy, Re Re Quinn, and Clifford Goudey. You kept me (in)sane through the (best)worst of it.

Funding for this research was provided by the MIT Sea Grant College Program under contract NA90AA-D-SG424, by the C. S. Draper Laboratory under contract MDA 972-88-C-0040, and by the Office of Naval Research under contract N00014-91-J-1014. The original idea (and in$piration) for the goal of deep ocean object retrieval came from Capt. William Shotts, USN, and Mr. John Freund.

Contents

List of Figures

List of Tables

Nomenclature

Symbol	Definition or Meaning
a, b	calibration model parameters
$A(\hat{r}, r)$	geometric spreading factor
AUSS	Advanced Unmanned Search System
AUV	Autonomous Underwater Vehicle
BEM	Boundary Element Method
BIM	Boundary Integral Method
CW	continuous wave
c	phase velocity of compression waves in fluid or elastic medium
DI	transducer directivity index
EKF	Extended Kalman Filter
EL	echo level
$G_\omega(\boldsymbol{r}, \boldsymbol{r}_0)$	Green's function
$\boldsymbol{G}$, $\boldsymbol{G}_m(k)$	Jacobian of measurement model
$g[k, \boldsymbol{z}]$, g_m	model for predicting a measurement
$\boldsymbol{H}$, $\boldsymbol{H}_m(k)$	Jacobian of *inverse* measurement model
$h[k, \boldsymbol{x}]$, h_m	model for predicting a measurement
I, I_i, I_r	acoustic intensity, incident, reflected
IF	intermediate frequency in superheterodyne demodulation
K-best	number of hypotheses generated in MHT filter
k	wavenumber or discrete time, depending on context
MFP	Matched Field Processing
MHT	Multiple Hypothesis Tracking
m	density ratio between two media
N-scan	depth of tree in MHT filter
n	index of refraction between two media
$\boldsymbol{P}(k\|j)$	approximate mean-squared-error of $\hat{\boldsymbol{x}}(k\|j)$
p, p_i, p_r	acoustic pressure, incident, reflected
$\boldsymbol{p}_i$	position vector of i^{th} point in global face representation
R	reflection coefficient
R_{ij}	Rayleigh reflection coefficient between adjacent layers
$\boldsymbol{R}$	covariance of random measurement noise
R_p	radius of circular piston transducer
r	range value of observation
$\boldsymbol{r}$	position vector

Symbol	Definition or Meaning
r_m	minimum likelihood ratio in MHT filter
r_s, r_c	sonar range and corrected range, in calibration model
RCD	Region of Constant Depth
ROV	Remotely Operated Vehicle
S	salinity, parts per thousand (‰)
$\boldsymbol{S}(k)$	covariance of prediction
S_0	surface on which boundary condition applies
t	continuous time
t_d	time of detection
T	temperature, °C.
TL_1, TL_2	outbound, inbound transmission loss
TOF	Time of Flight
TVG	Time Variable Gain
V_0	fluid volume
$\boldsymbol{v}$	particle velocity vector
w	minimum width of an RCD
$\boldsymbol{w}(k)$	random noise present in measurements
$\boldsymbol{x}(k)$	true state at time k
$\hat{\boldsymbol{x}}(k\|j)$	estimate of $\boldsymbol{x}(k)$ given measurements up to and including time j
$Z(k)$	set of measurements (scan) taken at time k
Z^k	cumulative measurement set at time k, $\bigcup_{j=1}^{k} Z(j)$
z	water depth, meters
$\hat{\boldsymbol{z}}(k\|j)$	predicted observation at time k given information through time j
$\boldsymbol{z}_i(k)$	an individual measurement of the set taken at time k
α	bearing angle of observation
$\alpha_{\log}$	logarithmic absorption coefficient for fluid medium
γ	size of validation gate
ζ	characteristic acoustic impedance
λ, λ_{Al}	acoustic wavelength in water, aluminum
λ_h, λ_d, λ_e	modified log likelihood of hypotheses
(ϕ, d)	plane model
ϕ, ϕ_0	phase of complex reflection coefficient
ϕ', ϕ_0'	derivative of phase angle with respect to ω
ϕ^{ML}	maximum likelihood partitioning of observations
ρ	density of fluid or elastic medium
ρ_c	characteristic range
Θ_l^k	hypothesis at time k
$\theta(\phi)$	event that ϕ correctly partitions the current measurements
θ_i, θ_r, θ_t	angles of incidence, reflection, and transmission
(ξ, η)	vertex model
(ξ, η, ρ)	sculpt model
ξ_S	boundary value for acoustic impedance
τ	RCD range threshold
ω	circular frequency of acoustic vibration

Chapter 1

Introduction

On 23 January 1977 the M/V *Lucona* mysteriously disappeared somewhere near the Maldive Islands in the eastern Arabian Sea, taking with her all but a few crew members. Cargo owner Udo Proksch, a prominent East German immigrant to Austria, subsequently filed an $18 million insurance claim that purported the loss of a valuable uranium processing mill. Challenging his claim, however, eyewitnesses reported that the actual loss was merely a discarded wheat mill and newly painted scrap metal—worthless junk. Nevertheless, the politically well connected Proksch remained free and unaccused for quite some time.

Following a political turnover in Austria, the authorities finally issued an arrest warrant in 1988 and apprehended Proksch as he transited the Vienna airport in 1989, by then a fugitive having altered his appearance through plastic surgery. The accusation that Proksch had sabotaged the vessel with military explosives arose. His side countered with the charge that a torpedo was responsible, either Russian or American, to prevent the cargo from reaching its destination. As the highly sensitive case progressed through the Austrian Court, Judge Hans-Christian Leiningen-Westerburg concluded that the only way to achieve justice was to "locate and photograph (as well as possibly recover pieces of) the *Lucona*."

1.1 Deep Ocean Object Retrieval

How does one go about recovering objects from the ocean floor? First, one must find the object. The traditional approach employs a deeply towed sensor array, typically equipped

with sidescan sonar, to produce a bathymetric image that human operators interpret, looking for possible contacts. After visiting each contact to evaluate it, a manned submersible or remotely operated vehicle (ROV) performs the actual retrieval. This technique led to the successful recovery of the cargo door of a Boeing 747, for example. Detached from United Airlines Flight 811 on 24 February 1989, the door sank to a depth of over 4.2 km off the coast of Hawaii.

In July 1990 a search team deployed the *ORION* towfish, equipped with side scan sonar, to traverse a predetermined area and image the bottom. The team identified several contacts, the one believed to be compatible with the missing door being marked with an acoustic beacon [19]. Two months later, another team inspected the debris field around the beacon using the manned submersible *SEA CLIFF*. The most likely contact identified earlier turned out to be merely a portion of the luggage container. The recovery of the actual cargo door, however, subsequently found nearby, successfully concluded the operation [80].

The previous example contains the three steps commonly found in underwater object retrieval:

1. conduct a broad area search using a towfish in order to enumerate possible contacts;
2. visit each contact with a manned submersible or video equipped ROV in order to identify desired objects;
3. recover each object with a manipulator and bring it to the surface.

Use of a tethered platform in step 1 limits the speed that the search covers ground because of the enormous hydrodynamic drag and poor maneuverability. In step 2, ROVs suffer from the same burden and they also require large support ships that are expensive to operate. Worse yet, manned submersibles require even larger ships and pose a risk to human life as well. How can we improve on this approach?

1.1.1 AUV Object Retrieval

The Naval Ocean Systems Center introduced the Advanced Unmanned Search System (AUSS) in the mid 1970's to enhance the the Navy's broad area search capabilities [79]. Improving on the traditional approach outlined above, this system performs the first two

steps with an autonomous underwater vehicle (AUV). The AUSS concept still relies on skilled operators to interpret the initial side scan and final video images, both transmitted to the surface through a low-bandwidth acoustic link [47]. Free from the burden of a tether, however, an AUV allows higher ground speed and better maneuverability. Coordinated from a single surface vessel, *multiple* AUVs achieve even higher efficiency [22]. For the first time, we find an autonomous system meeting the dual criteria of broad area search and final contact evaluation dictated by the first two object retrieval steps. Why not add the third as well?

The extension of the AUSS concept to its next logical step requires that we develop an AUV equipped with a manipulator. A vehicle with this capability could attach a cable to objects that allow it, enabling winch retrieval, or even grasp objects directly, retrieving them itself. Though merely an incremental step in concept, the implied technology dictates the need for advancement in many active fields of research: control of a sensitive system fraught with complex dynamic interactions; underwater grasping with a dextrous manipulator; and autonomous interpretation of sensor data.

This dissertation focuses on the last of the three areas. Before successfully performing a recovery, an autonomous system must know the location and shape of an object in order to determine where and how to grasp it.

1.1.2 Additional Applications

An algorithm capable of autonomously interpreting the data from an acoustic sensor potentially fulfills many needs. The shape information it provides finds use in map building and feature based navigation. The Navy has identified several missions whose requirements are uniquely satisfied by AUVs [68]. Of those mentioned, feature based navigation improves the ability of an AUV to perform shallow water surveillance, covert mapping, and mine identification.

The commercial sector benefits as well. We expect that when AUVs become proven and commercially available they will replace the current ROVs used to inspect offshore fields and pipelines. As offshore industry moves into deeper water this switch becomes even more economically attractive.

Returning to the case of the *Lucona*, Eastport International, Inc., used the traditional approach to broad area search and contact evaluation to provide all of the answers needed by the Austrian Court [56]. The *Magellan 725* ROV, following up on leads generated by the *Explorer 6000* towfish, produced video images that confirmed both criminal charges. Explosives in the hold had indeed blown the *Lucona* apart, and the cargo was quite worthless. Recovering pieces of wreckage proved unnecessary.

Our story concludes with a remark, made by the president of Eastport, that deep water can no longer cover for crime. Proksch received a fine equal to the cost of mounting the search and generating the documentation it provided—the trial proved the most expensive in Austrian history—along with a 20 year prison sentence for his fraud and murder convictions.

Whether or not our vision of AUV use in the future leads to more expedient justice remains to be seen. However we expect that the increased human presence that will be afforded by the use of AUVs will turn the phrase "lost at sea" into an anachronism.

1.2 A Commercially Available Sensor

In this work we decided to use a commercially available sensor. By doing so, we free ourselves from the difficulties associated with developing custom hardware. Which sensor to use depends on several factors.

The *Odyssey* class vehicles, small oceanographic survey AUVs, offer a context to keep in mind. Developed in the Underwater Vehicles Laboratory of the MIT Sea Grant College Program, the *Odyssey* vehicles provide a mobile platform which inexpensively delivers deep ocean access to address the needs of a variety of scientific applications. The vehicle, with a length of 2.2 m and diameter 0.6 m, displaces less than 200 kg. Two 17 in. diameter spheres, made of 1/2 in. thick borosilicate glass, provide the pressure hulls. The vehicles have an operational depth of 6.7 km, certified by hydrostatic tests on the spheres.

Bellingham *et al.* [8, 9] offer a more detailed description of the purpose and performance of the *Odyssey* class AUVs, but there is one consideration worth mentioning here. The hydrodynamically efficient shape of the external fairing gives the vehicle a maximum range of approximately 180 km, assuming alkaline batteries and a hotel (non-propulsive) load of

50 W. Since the dedicated vehicle computer consumes about half of the available power, approximately 25 W remain for sensing systems. When integrating a sensor system with an *Odyssey* vehicle, we either use a miserly system or pay the price in terms of reduced operating range.

After reviewing the different commercially available systems in light of the above considerations, we have chosen a Tritech ST1000 Profiling Sonar. The compact design and relatively low power requirements of 14 W make this mechanically scanned 1.25 MHz pencil beam a good choice for integration to an *Odyssey* class vehicle, although its maximum rated depth of 3000 m limits the amount of ocean floor accessible [16, 75]. The relatively high frequency ensures sufficient range accuracy for shape reconstruction, which further supports it as a good choice. We would prefer a wider beam than the ST1000 provides, but we found it adequate to prove the point we were trying to make in this research.

1.3 Document Layout

In the next chapter, we introduce the specific problem addressed by this research—shape reconstruction using a profiling sonar. We make assumptions on the environment in which this problem applies, and we define the criteria that will determine its successful solution. We then describe the three-step approach that we developed to address this problem. Chapter 2 closes with a discussion of the methods found in the literature that address similar problems.

Why does a sonar scan look the way it does? In Chapter 3, we address the factors that contribute to the structure of profile scans. We look at how an individual return is formed and how multiple returns combine into the discernible features found in complete scans. We incorporate a model for the expected *data* features into an extraction filter that generates polar (bearing and range) observations of the *world* features found in the environment. In the end of Chapter 3, we present a quantitative model for active sonar, based on classical theory, which supports our approach to feature extraction.

We discuss the problem of correspondence between data features and world features in Chapter 4. We consider different approaches described in the literature and select one

particularly well suited to our application—the model-based multiple hypothesis tracking filter. Models which capture the behavior of target observations are developed to track their evolution along curved, two-dimensional objects. Results follow which demonstrate some of their capabilities and some of their limitations.

Chapter 5 details our experimental results. We begin by describing how we synthesize the final geometry of the scene using a simple representation that allows efficient incremental updates to approximate curved faces. We then present the three experiments performed to validate the approach. The first one demonstrates a similar capability to that found in the terrestrial robotic community for linear geometries. The second experiment, through negative results, identifies the critical issues in tracking curved surfaces. Finally, the third experiment shows the successful shape reconstruction of a curved object.

Also in Chapter 5, we investigate the efficiency of the tracking filter. We then relax some of the earlier assumptions and assess the performance of the algorithm. Specifically, the discussion includes results from three adverse conditions: applying the algorithm to rough surfaces; simulating the effects of uncertain navigation; and reducing the density of data provided.

Chapter 6 summarizes the research contribution in this thesis and discusses future research issues. Recommendations are given and insights shared on extending the research to environments which include a broader range of conditions, as might be found in a natural underwater setting.

The appendix contains a discussion of our experimental techniques. We begin by describing how we calibrate the range measurements produced by the profiling sonar used in our experiments. Next, we describe the sensor mounting apparatus that enables us to locate the sensor arbitrarily in a horizontal plane. Finally, we share some of the difficulties encountered in performing the experiments, and the steps taken to alleviate them.

Chapter 2

Overview and Problem Formulation

In this chapter we describe the specific shape reconstruction problem under investigation. We state the assumptions made on the environment to which our approach applies, and we indicate the criteria that define a successful solution. We then discuss the issues encountered in inverting sonar data. The various techniques found in the literature guide our path in identifying how we wish to approach the problem of shape reconstruction. An overview of our three step approach follows, accompanied by the inspiration from which it springs. We close this chapter by establishing the contribution made by this research.

2.1 The Problem

This dissertation addresses the specific problem of how to reconstruct the geometric shape of two dimensional underwater objects from acoustic measurements. Important enough unto itself, shape reconstruction also shares many issues in common with the problems of autonomous map building and navigation. In particular, all of these applications require the robust interpretation of measurements that often have uncertain origin. We hope that the solution we provide addresses this issue with sufficient generality to find use in both AUV operations and the broader field of robotics in general.

2.1.1 Assumptions

We begin this research by making a set of assumptions that restrict the environment to which the algorithm applies. For this work, we impose the following conditions:

- **two dimensional**, we confine the motion of the sensor to the horizontal plane and we limit the environment to vertically unbounded objects (see figure 2.1);
- **perfect navigation**, we assume accurate knowledge of the location of the sensor;
- **smooth surfaces**, for us to exploit the strict alignment conditions found in specularly reflected returns, object surfaces must have average surface roughness less than one fourth the acoustic wavelength.[1]

In chapter 6, however, we discusses the issues involved in extending the algorithm to handle settings in which these assumptions are relaxed, namely three dimensional scenes that contain objects with rough surfaces.

2.1.2 Criteria for Success

We would eventually like to adapt the approach we develop in this investigation to support the operation of a small AUV, such as the *Odyssey* discussed in Chapter 1. This goal poses unique challenges that must be met by both hardware and software. To meet them we establish the following two criteria:

- **off-the-shelf hardware**, employing a commercially available sonar system saves money and time, relieving us from the onus of developing custom electronics;
- **efficient implementation**, our algorithm must interpret sonar observations in real time in order to provide useful information to the ongoing mission of the vehicle.

With respect to the latter, many existing shape reconstruction implementations only post-process measurements, lacking the capacity to provide interpretation on the fly. We take special interest in *recursive* approaches because they allow for efficient incremental refinement, in contrast with *batch processing* approaches, which recompute everything from scratch upon the arrival of new information.

[1] We typically express roughness of a surface using its *root-mean-square* elevation.

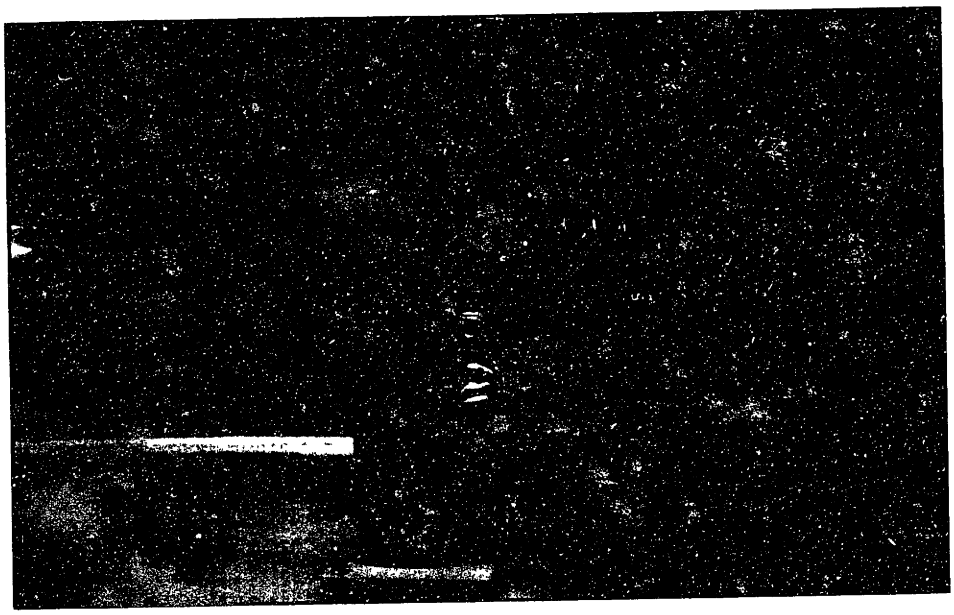


Figure 2.1: A two-dimensional world. By locating the sensor halfway through the water column and *carefully* aligning prismatic objects in a vertical orientation, we are able to make two-dimensional measurements.

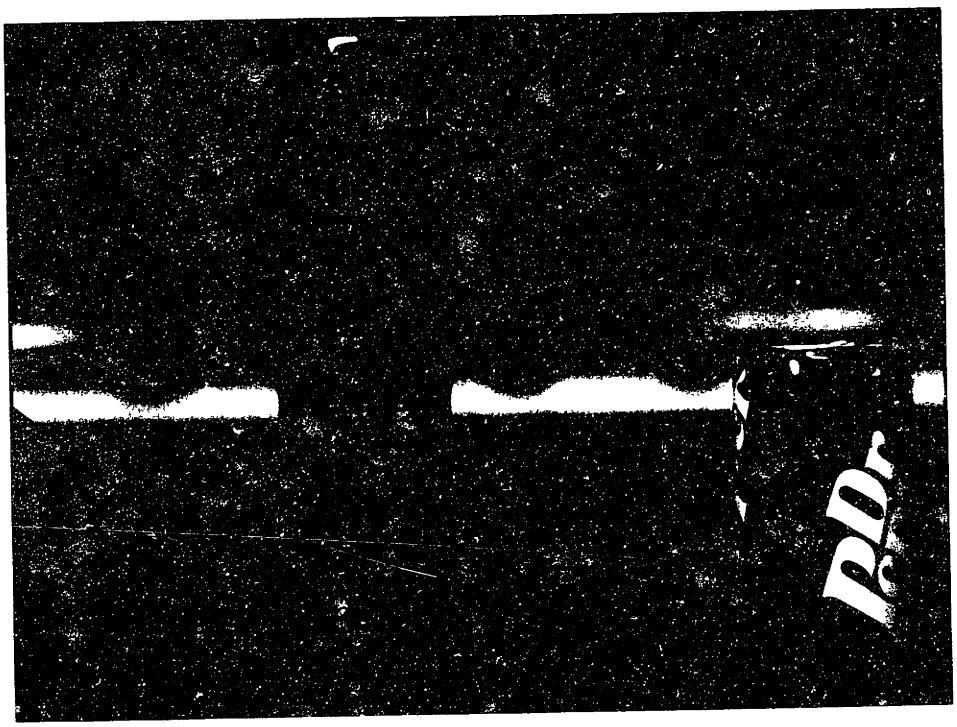


Figure 2.2: Tritech ST1000 1.25 MHz profiling sonar. This commercially available unit has a mechanically scanned head. Its 50 mm diameter circular piston transducer leads to a value of $ka = 133$, with resulting 3 dB beamwidth of 1.7°.

2.2 The Difficult Nature of Sonar Interpretation

For this investigation we purchased a Tritech ST1000 pencil beam profiling sonar, shown in figure 2.2. This commercially available unit is compact, reasonably priced, and has low power requirements. As indicated in Chapter 1, its physical size and energy needs make it a good candidate for integration with a small AUV. This type of system reports the range corresponding to the arrival of the earliest echo (if one exists) for each transmitted pulse—it *profiles* the immediately surrounding environment.

Our decision to purchase the ST1000 also reflects the range accuracy it provides. The 1.25 MHz transmit frequency yields a 1.2 mm wavelength capable of fairly high resolution measurements. The ideal sensor for our application would have absolute range accuracy higher than the ST1000 provides. It would also have a broader beamwidth, although Chapters 3 and 4 will reveal this concern to be of less importance. The economics of designing and building this hypothetical system make it prohibitive, however. The ST1000 balances the conflicting issues of cost versus performance well enough to suit our purposes.

A panoramic scan produced by the ST1000 can be difficult to interpret due to the inherent ambiguity of the individual returns it contains. Kuc and Siegel [41] observe that sonar systems

> produce complex signals that require some interpretation to be used successfully. Problems arise in the straightforward, but naive, interpretation that ... an object must lie along the transducer line-of-sight.

Demonstrating this point, figure 2.3 shows a typical scan from the experiments that we describe in Chapter 5. Displayed using the naive interpretation, the scan gives few clues regarding the underlying geometry of the scene from which it originates.

The lack of agreement between the scan and the actual scene shown in figure 2.4, composed of prismatic objects standing on end in a tank, makes clear that shape reconstruction requires an approach more sophisticated than simple line-of-sight inversion. The concrete far wall shows up well, for example, as do some features of the objects. The glass windows, on the other hand, show up poorly because their smooth surfaces provide an acoustic mirror in which the reflections of other features appear.

The problem results from the very nature of sonar. A given return could originate due

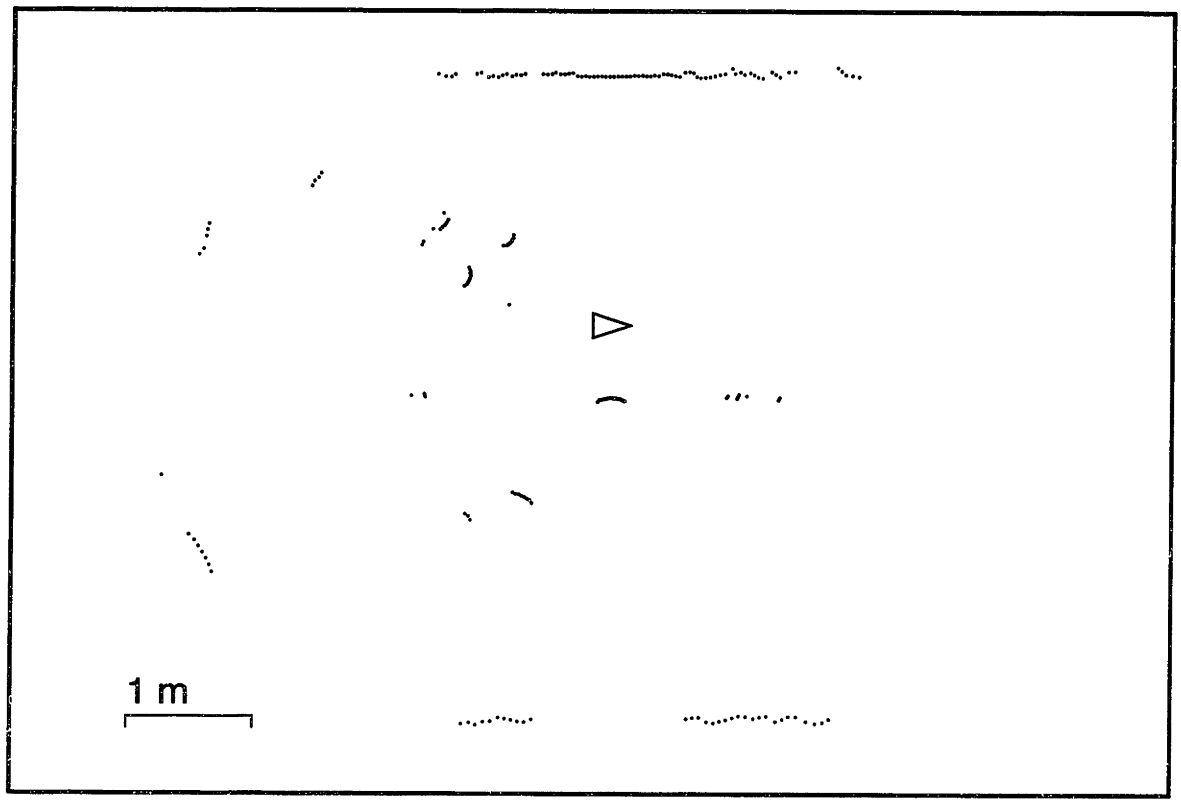


Figure 2.3: Naive line-of-sight interpretation of sonar returns. The unfilled triangle represents the location of the ST1000 profiling sonar and the dots follow from a simple line-of-sight inversion of the measurements. As can be seen, a single scan offers few clues as to the underlying scene.

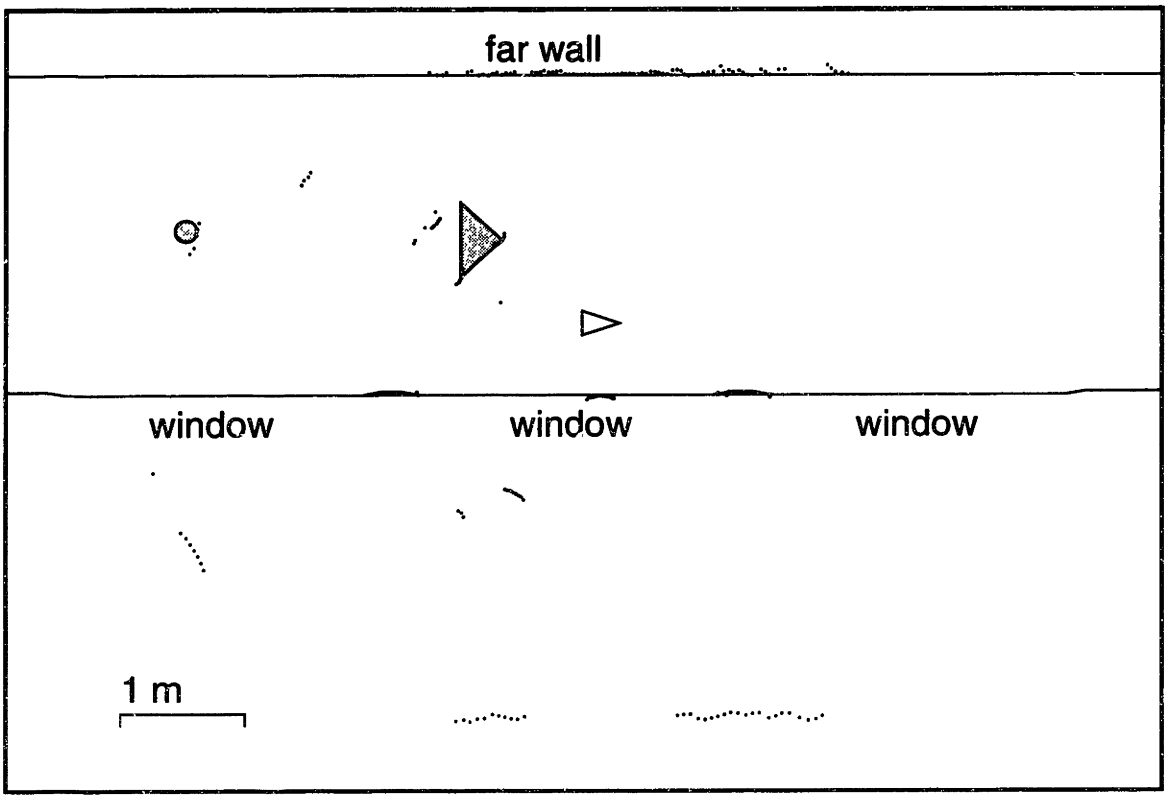


Figure 2.4: The geometry underlying the scan in figure 2.3. Two prismatic objects, an 18 cm diameter aluminum cylinder and a flooded triangle with thin aluminum walls, stand on end in a fresh water tank. The far (concrete) wall of the tank, rough relative to the wavelength of the 1.25 MHz sensor, contrasts with the (glass) windows in the foreground.

to a variety of mechanisms: specular reflection from a smooth surface; diffuse scattering from a rough surface; diffraction from a sharp edge or corner; or even a more complex effect such as a creeping wave [28]. Each mechanism generates a different constraint on the shape of the object responsible for the return. Further compounding the difficulty, sometimes a sensor completely fails to detect the echo from a target. The absence of an expected return confuses the interpretation of a scan as does the presence of an unexpected one. These conditions commonly occur in transient operational settings and in environments characterized by the pervasive problem of multipath.

How then do we interpret acoustic measurements in order to suit the environment modeling needs particular to our application? Many others have asked a similar question and filled a large body of literature with different methods. Their attempts include both *image* and *signal processing*, and *geometric inversion*. The first two of these we discuss briefly in the interest of providing a comparison to the last one, the category in which our own approach belongs.

2.2.1 Image Processing

When underwater conditions render video cameras useless, ROV pilots commonly resort to *imaging* sonar sets to help guide their operations. In their most common use, imaging systems attempt to provide a pictorial view of a scene. They divide elapsed time into a sequence of intervals that correspond to *range bins*, sample the received signal to determine an amplitude level for each bin, and then digitize these amplitudes. The resulting values, often of 4 bit resolution, are used to color the displayed image according to the received signal strength.

A skilled individual, who draws upon considerable experience, often can infer a great deal of information from an imaging sonar display. In an attempt to imitate the techniques used by human operators, some autonomous methods cast sonar interpretation as a problem of image processing [26, 67, 74]. Techniques from computer vision directly apply in these approaches [37]. Edge detection, for example, extracts line segments from the images and correlates them to physical features in the environment.

While a human interpreter may possess the capacity to assimilate the vast amount of

information in a sonar image, and even prefer such a display over other choices, we feel that profiling sonars better lend themselves to autonomous interpretation. We wish to explore the limits of what is possible using only the information provided by this type of sensor.

2.2.2 Signal Processing

Working in the time domain, Freedman [32, 31] applies wave theory to develop a model for echo formulation. This approach suggests looking in the received signal for multiple returns from a single transmitted pulse, under the assumption that each return correlates to a discontinuity in the projected solid angle, or one of its derivatives, of the target. The amplitude of the return indicates the size of the discontinuity.

To our knowledge, no implementation using this approach exists, perhaps because it requires signal amplitude data of immensely high quality. The typical ST1000 waveform shown in figure 2.5, captured using a digital oscilloscope, lacks sufficient detail to enable this technique. Such measurements may be possible in the laboratory using an esoteric sensor that has been painstakingly calibrated, but no commercially available sonar suitable for AUV integration reaches this level of accuracy. Even if one did, amplitude measurements alone lack sufficient information to reconstruct detailed shape.

Some of these issues can be addressed by the algorithms that combine amplitude and time-of-flight measurements, as proposed by theoretical acoustics. To the best of our knowledge, however, none has been implemented using real data. Again, commercially available sonar systems provide only coarse amplitude information.

Audenaert *et al.* [4] describe an approach that transmits a judiciously chosen waveform and then applies a matched filter to detect returns from multiple targets with a single transmitted pulse. In contrast with the above approach, their work uses the amplitude only to establish detection. An accurate estimate of the round trip travel time for each pulse in the received signal determines the range to each target.

Although this technique is implemented in a tri-aural system that triangulates range measurements to localize targets [61], its real contribution lies in the signal processing—simple triangulation has been done before. Because the implementation ignores the data correspondence problem, we agree with the assessment made by the authors that the system

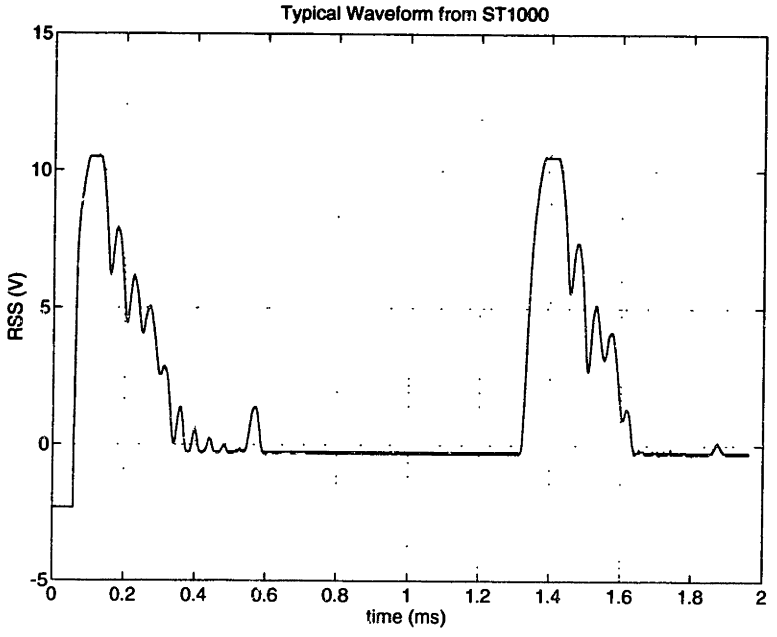


Figure 2.5: A typical waveform from the Tritech ST1000. This plot shows the received signal strength (RSS) as a function of time, captured by a digital oscilloscope, for a single transmitted *ping* incident on a planar target. The short negative region in the beginning corresponds to the duration of the outgoing pulse, followed by considerable ringing of the transducer and detection circuitry, and finally the echo pulse at right.

relies on clutter free data.

Working in the frequency domain, inverse resonance scattering approaches attempt to detect and classify targets by identifying resonant peaks in the received signal that correspond to discrete vibrational modes of targets of interest [25, 33]. (Flax *et al.* [29] survey resonant scattering theory, and Überall [77] edits a recent collection devoted to the same.)

Because these methods rely on *a priori* knowledge of the acoustic signature for each target sought, the need arises to maintain a library of characteristic waveforms. This dependence prohibits the ability to reconstruct free form shapes. Implementations of inverse resonance scattering to date, for example, merely identify instances of known simple shapes such as spheres and cylinders.

Matched field processing (MFP) techniques, of recent interest in ocean acoustics, refer to array processing methods that invert received signals in order to determine unknown parameters in an assumed environment model [5]. Using sophisticated acoustic models, these methods predict parameter values and then compare them to those extracted from data using broadband coherent signal processing. A generalization of plane wave beamforming, the MFP approach has achieved levels of accuracy well beyond the traditional Rayleigh and Fresnel limits for arrays in source localization, for example [6].

Though these methods exploit many aspects of the ocean waveguide, including internal waves in the water column, geo-acoustics of the seabed, and even bathymetry, they have yet to find application in shape reconstruction. An additional concern, MFP predictions rely on the ability to solve the wave equation for the particular environment that applies.

Solving the Wave Equation

For an inviscid, irrotational fluid with spatially invariant density ρ and sound velocity c, the inhomogeneous *standard wave equation*,

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = F(\boldsymbol{r}, t), \tag{2.1}$$

in which ∇^2 signifies the Laplacian, describes the pressure field $p(\boldsymbol{r}, t)$ that results from acoustic excitation $F(\boldsymbol{r}, t)$. If the quasi-stationary environment assumption holds, we can

treat c as a constant and apply the forward Fourier transform to simplify this second order *partial* differential equation. The frequency domain result, known as the inhomogeneous Helmholtz equation,

$$[\nabla^2 + k^2]\, p(\boldsymbol{r}, \omega) = f(\boldsymbol{r}, \omega), \tag{2.2}$$

in which $k = \omega/c$, is a second order *ordinary* differential equation that describes the propagation of monochromatic sound at circular frequency ω through a lossless fluid medium [43, 45, 60]. Together with a set of appropriate boundary conditions, equation (2.2) describes a boundary value problem. Its solution in the frequency domain leads to a time domain solution via an application of the *inverse* Fourier transform.

The boundary integral method (BIM) provides a recipe for the exact solution by application of Green's theorem. Any function which satisfies the Helmholtz equation for a unit point source at arbitrary location $\boldsymbol{r}_0$,

$$[\nabla^2 + k^2]\, G_\omega(\boldsymbol{r}, \boldsymbol{r}_0) = -4\pi\delta(\boldsymbol{r} - \boldsymbol{r}_0). \tag{2.3}$$

is known as a Green's function. For an arbitrary point source $f(\boldsymbol{r}, \omega) = -f_\omega \delta(\boldsymbol{r} - \boldsymbol{r}_0)$ at position $\boldsymbol{r}_0$, the integral representation,

$$p(\boldsymbol{r}) = \underbrace{\int_{V_0} f(\boldsymbol{r}_0)\, G_\omega dV}_{\text{source contribution}} + \frac{1}{4\pi} \underbrace{\int_{S_0} \left[G_\omega \frac{\partial p}{\partial n} - p \frac{\partial G_\omega}{\partial n} \right] dS}_{\text{boundary conditions}} \tag{2.4}$$

in which $\partial/\partial n$ represents the normal derivative, gives the pressure field for all receiver locations $\boldsymbol{r}$. As labeled, the volume integral defines the contribution to the field from the acoustic source, and the surface integral the contribution specified by the boundary conditions.

Except for boundaries with very simple geometry, the integral expression in equation (2.4) must be evaluated numerically. The boundary element method (BEM) performs this task by dividing the fluid boundary S_0 into panels to obtain a matrix equation that, with a finite number of degrees of freedom, represents equation (2.4) in a discrete form [64]. Schmidt [66], for example, offers a numerically stable matrix technique that applies to backscattering from spherically stratified shells.

Additional modern numerical techniques for solving the boundary value problem include the approximate methods of ray tracing, wavenumber integration, normal modes, and matched field processing [65]. The difficulty in implementing one of these analytic or numerical techniques for inverting the wave equation motivates a more intuitive class of approaches to sonar interpretation—ones that exploit the inherent geometric constraints of acoustic measurements.

2.2.3 Geometric Inversion

We begin this discussion by presenting the state of the art for approaches that model the environment using a discrete representation. They decompose the area or volume of interest into cells and assign to each one a set of attributes that describe the environment according to the application.

Moravec and Elfes [52], for example, have implemented a two dimensional approach to map building that uses *stochastic backprojection* to combine multiple wide-angle range measurements into an environment model. By assigning probability profiles to describe the location of the assumed scattering surface for each measurement, they maintain a world map as a grid of cells that contain the probability of being occupied. The technique is applied to build a map of an office environment for a mobile robot.

Extending this approach to three dimensions, Stewart [72] applies it to model the underwater environment. While doing so, he explicitly incorporates the uncertainty in the location of the sensor, providing a more thorough treatment of multisensor fusion because of the inherent difficulty in obtaining navigational information. He similarly models the beam of the primary sensor with a probability distribution, which assumes a diffuse scattering surface, and uses each measurement to refine the model incrementally. Successful results include applications of this method to multibeam bathymetric data, sidescan sonar data, and most interestingly, profile data from the USS *Monitor*.

These implementations offer the benefits of a fixed storage requirement, the size of which depends on the desired resolution of the map and thus can be known *a priori*. If a fixed resolution map proves inadequate, the approach can be modified to use a variable resolution *octree* encoding to represent the occupied volume. Even under a high rate of data arrival,

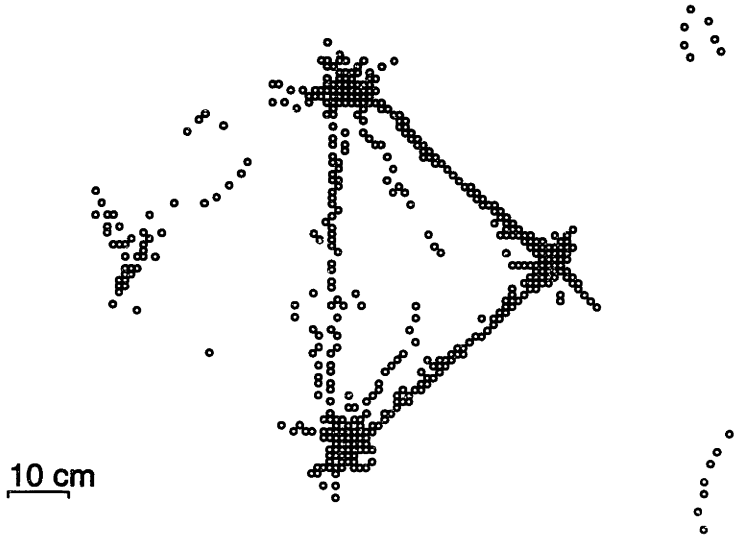


Figure 2.6: Example of discrete backprojection. A high-frequency profiling sonar was placed in discrete locations around the perimeter of a flooded triangular prism. The small circles represent individual returns, inverted using a simple line-of-sight model, *whose locations are digitized into a fixed resolution grid.*

these algorithms allow efficient updates to the map.

The result of applying a simple backprojection to measurements from the ST1000 high-frequency profiling sonar, shown using a discrete representation in figure 2.6, illustrates one of the drawbacks of this approach. A close look at the corners of the triangle reveals a star shaped pattern that always results from reconstructing a point object using a discrete number of projections. To identify such a feature and determine its location requires further processing of the model, an undesirable attribute.

The type of representation used determines what specific information the environment model makes explicit. Stewart [72] reminds us that

> the purposes for which a model can be used and the efficiency with which those purposes can be accomplished follow directly from the choice of representation.

For our application, we need a representation which explicitly indicates the location of the boundary of objects. The task of grasping, implied by object retrieval, requires efficient access to detailed shape information. Because cellular decompositions fail to meet this need,

we now present geometric inversion approaches that attempt to explicitly model continuous boundaries.

In an interesting approach to non-destructive testing and evaluation. Dangelmayr and Wright [24] apply ultrasound to recover the shape of internal cracks in solids. Their technique relies on the geometric constraints in range measurements made from an offset curve, *i.e.*, an arbitrary trajectory along the outer surface of a test piece. The theory of asymptotic inverse scattering forms the basis for this approach [23].

Although such a technique may be possible with data generated in the laboratory from a carefully calibrated ultrasound sensor, it remains to be proven in the field. We expect the measurements made by a commercially available sonar suitable for AUV integration to lack the levels of accuracy required for this approach.

Brown [12, 13] extracts curved surface features with generic range sensors also by solving geometric constraints. This important work reveals that the trajectory of an echo—the point on the surface from which a return originates—follows a geodesic path as the sensor moves. He proposes surface tracking by observing the measurements and estimating the normal vector to the surface. The approach assumes that the echoes all originate from the *same* curved surface, however, which begs the question of how to determine this correspondence.

State of the Art

Reid [63] explicitly solves the correspondence problem with the multiple hypothesis tracking (MHT) filter that, developed for radar tracking of aircraft, uses a Bayesian framework to assign measurements to targets. Cox and Leonard [20] have extended this technique to the *model-based* MHT that allows each target to behave according to a different dynamic model.

They apply the approach to build a map of a two dimensional office environment that contains flat walls and corners. Observations of these features are tracked across multiple scans by comparing them with predictions that are calculated from estimates provided by extended Kalman filters. This work represents an important contribution because it debunks the myth that acoustic sensors fail to provide measurements suitable for reconstructing specular environments that contain smooth, highly reflective targets.

The key to this approach lies in the observation that consecutive range measurements from a scanning range sensor lie on a circular arc tangent to the boundary of the target [41]. This realization enables one to exploit the information present in the critical alignment necessary to generate a return, which we further discuss in Section 3.1.2 [36]. Similar modeling is found in the works of Leonard and Durrant-Whyte [44], and Kleeman and Kuc [39].

The success of these approaches in environment modeling suggests that we follow their lead. Our approach will build upon the model-based MHT to enable it to track the observations of a *curved* two dimensional environment, extending the limitation of flat walls found in earlier work. Before presenting an overview of the three-step approach that we have implemented in this investigation, we present the inspiration that leads to the idea of exploiting geometric constraints.

2.2.4 Inspiration

The impressive echolocation capabilities of mammals suggest that sonar echoes contain detailed shape information. Au [3] and Popper [62] report these findings for dolphins, as does Griffin [35] for bats. After extensively observing dolphin behavior throughout a series of object recognition experiments, Au asks

> what kind of information do the animals obtain ... by swimming almost perpendicular to the line-of-sight direction of the target?

Though we may never know the answer to this question, the observed dolphin behavior triggers an idea.

In figure 2.7 the unfilled triangles represent the positions from which a hypothetical dolphin observes an unknown object while swimming by. Acoustic rays drawn from these locations to a common feature of the object signify the possible measurements that the animal makes. The sketch suggest a geometric approach to shape reconstruction. Combining these hypothetical measurements in the correct manner allows us to infer the existence of a sharp corner in the environment and determine its location explicitly.

We form the intuitive hypothesis that, by applying a geometric approach, we can reconstruct the shape of an unknown object by tracking the point where the sensor beam

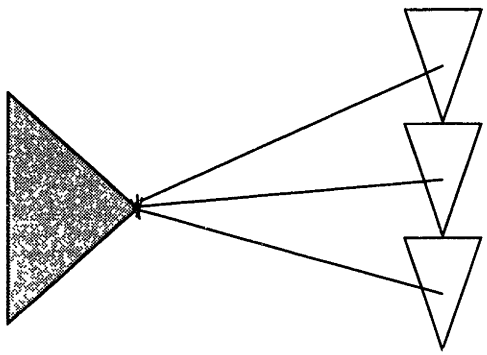


Figure 2.7: The inspiration for our geometric approach. Unfilled triangles represent an autonomous entity swimming by an object, exhibiting a behavior similar to the dolphin observed by Au [3]. Acoustic rays drawn to the corner of the filled triangle indicate observations which, using a geometric reconstruction approach, infer both the presence and location of a sharp corner.

comes into contact with the object. This point traces out the surface of the object in a manner analogous to touch. We will capitalize on the motion of the sensor and exploit the geometric constraints that it provides.

2.3 A 3-Step Approach

The approach we propose synthesizes an environment model using observations made from multiple vantage points. This geometric inversion technique introduces the problem of data correspondence. As is true in any robotic sensing task, the need to solve the correspondence problem raises one of the most difficult questions to answer—*how do measured features in the data set relate to each other and to the underlying geometry of the scene*?

The correspondence question raises both the issue of measuring features and the issue of relating the measurements to the world. The algorithm we develop in this investigation, shown in figure 2.8, addresses these issues with a three step approach: (1) identify in raw data a set of observations that correspond to some geometric feature in the world; (2) track reported observations and classify them according to similar origin in real time, providing

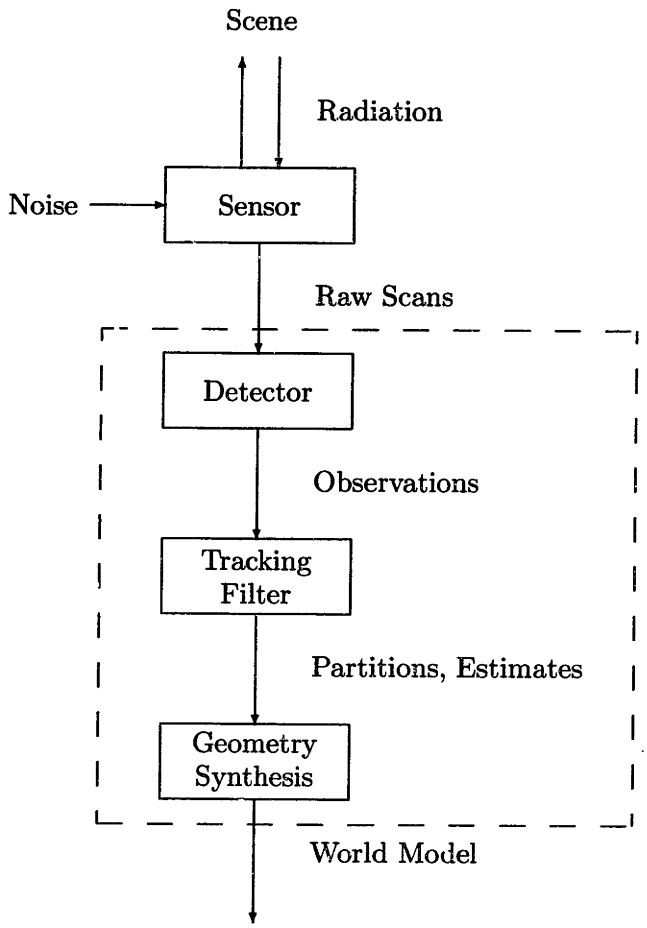


Figure 2.8: Schematic diagram of the approach. An imperfect acoustic sensor interacts with an underwater scene to generate raw scans. These scans pass through a feature detector that extracts a set of observations. A tracking filter partitions the observations into groups which share a common geometric origin. A final model of the world is synthesized from the partitioned observations.

local geometric estimates; and (3) reconstruct global geometry using groups of similarly associated measurements. The remainder of this section summarizes these steps.

2.3.1 Detecting Features in Scans

Determining the nature of the environment responsible for a panoramic scan requires an appropriate framework for interpretation built upon sound physical principles. Lozano-Pérez [46] postulates:

> The performance of any sensing system is largely determined after the feature detectors have finished their work. Models that characterize a sensor or feature detector by a covariance matrix are too weak; they do not capture the relevant physics.

True as stated, however the covariance matrix still plays an important role in providing a measure of the sensor error that we find useful in many instances, measurement validation for example.

Our approach to feature extraction derives motivation from a technique by Hallam [36] which exploits the strict alignment conditions found in specularly reflected returns. His approach leads to a model from Kuc and Siegel [41], developed for use in land robotics, which defines circular arcs as the basic features inherently found in sonar data.

Arcs correspond to sequences of individual returns over which the range value reported effectively remains constant, dubbed *regions of constant depth* or RCDs by Leonard and Durrant-Whyte [44]. The choice of an arc as the basic feature takes root in the underlying physics of echo formation and beam patterns. From an RCD, we extract a bearing angle and a range which combine to determine the location of target. These parameters constitute an *observation* or *report.*

2.3.2 Tracking Target Observations

After processing the raw scans into sets of RCD features, we must provide a technique to combine these observations into estimates for the geometric features in the environment. Many approaches for estimating the geometry of a scene assume a known correspondence between the observations and their origin in the surrounding environment. Sometimes this

assumption is reasonable, when there is very little difference between consecutive observations for example. In general, the unavailability of this information creates a formidable obstacle in interpreting sensor data. The multiple hypothesis approach applies a Bayesian framework to solve for the associations explicitly.

The data association hypotheses formed by the MHT filter enumerate all likely assignments of each reported observation. The filter selects from three possible interpretations for the correspondence of each report:

1. **a prior target**, the filter subsequently applies the report to update the target state estimate;

2. **an unencountered target**, for which the filter generates a new track using the report to calculate the initial state estimate of the target;

3. **a false alarm**, the filter ultimately accepts this option when the alternate interpretations fail to offer an adequately probable explanation.

The latter two treatments apply for all observations. We use validation gates, similar to those described by Singer and Kanyuck [70], to limit the number of hypotheses that associate observations with existing targets.

When the time comes to decide which association hypothesis to use, the MHT filter selects the one with the highest likelihood. This capability requires using the track estimates of the underlying geometry to predict the observations that should occur and comparing them to those that do occur. The likelihood of a particular association includes, among other things, this *statistical distance* because it provides a measure of how well an observation agrees with its expectation.

Our algorithm provides multiple models for the dynamic behavior of measurements associated with environment features. These models include the *extended Kalman filter* and *nonlinear least squares*, used to provide target state estimates. In both of these approaches, we provide an approximate mean-squared error, required for validating measurements and useful for indicating confidence levels in the estimate.

The MHT stands alone in its ability to explain the origins of measurements in the presence of clutter, *i.e.*, measurements that result from random noise. Our version, built

on top of a publicly available framework [50], extends its application from previous work to map building by providing a set of models which add the capacity to track curved surfaces as well as flat.

2.3.3 Synthesizing Global Geometry

Attempting full scene reconstruction in the tracking stage would choke our algorithm, preventing it from keeping pace with the real time arrival of data. The dimensionality of the global solution drives this concern. A complete geometric representation of a complicated scene contains a staggering number of degrees of freedom. No real time application stands a chance of maintaining full-state estimates in this vast a parameter space. For this reason, the tracking filter limits its interest to generating estimates of the local geometry only in regions confined to the proximity of the sensor, using a low number of degrees of freedom.

The algorithm as a whole defers estimating the global geometry of features until after it has determined the correspondence of the measurements. Once assignment decisions have been made, however, we may adopt a representation with as many degrees of freedom as the data can satisfy.

We have adopted a segmented curve for our global geometry. In the interest of simplicity, we use linear spans to connect the junction points that emerge from the tracking stage. This representation makes explicit the boundary of objects, a characteristic which supports efficient interrogation for the purposes of navigation and object retrieval.

2.3.4 The Contribution of Our Research

This work extends the model-based multiple hypothesis approach to nonlinear geometry, *i.e.*, free-form curves. Allowing finite curvature to exist alongside zero and infinite curvature makes this problem a difficult one. A corner or sharp edge, for example, is the extreme case of a cylinder with zero radius (infinite curvature). At the other extreme, one can model a flat wall as a cylinder with zero curvature (infinite radius). Everything between these two limiting cases constitutes finite curvature.

In previous work, the significant difference between the two target behaviors (flat wall and corner) simplified the data classification process [20]. The approach developed for this

research, however, must distinguish between a corner and a surface with high curvature, such as a cylinder with small radius, and between a plane and a surface with low curvature.

An additional contribution, to our knowledge, this work is the first to examine the geometric constraint of specular reflection with real underwater data from a commercially available sonar system.

2.4 Summary

We have defined a research problem of reconstructing geometric shape using an acoustic sensor. After considering the image processing and signal processing approaches found in the literature, we have argued for extension of the geometric processing approach developed by Hallam [36], Kuc and Siegel [41], and Leonard and Durrant-Whyte [44].

The approach we develop for this research divides this task into three parts: feature extraction from raw scans, tracking observed features from scan to scan as the sensor moves, and synthesizing the final geometry from sets of correctly partitioned observations which share a common origin. In the next chapter we investigate the first of these parts.

Chapter 3

Feature Extraction

The first part of our three-step shape reconstruction technique deals with feature extraction. What makes a good feature? All techniques must answer this difficult question. Image processing approaches, for example, assume that line segments in the scans correspond to edges of objects in the world. One approach to extract these segments applies the image convolution operators found in computer vision, of which Horn [37] provides a survey. Alternatively, the type of features sought in inverse resonance scattering are the resonant peaks which make up the acoustic signature of the target, as discussed in Section 2.2. What kind of features exist which suit our purposes?

We answer this question by examining how a simple echoranging sensor makes individual measurements and then combines them to form complete scans. The underlying physical phenomena lead us to a model for the type of features we expect to find. The detected features, in turn, generate bearing and range estimates which constitute the raw observations required by the latter stages of the algorithm. We close this chapter by applying the classical active sonar equation to quantify the range of target attributes over which our feature description holds.

3.1 Scan Formation

Feature extraction begins with an understanding of the physical phenomena that govern the process by which a sensor makes individual measurements. Applying the underlying

theory allows us to synthesize individual measurements into a prediction of what kind of features we expect to find in raw sonar scans. Knowing what to look for gives us a basis to detect and localize instances of the features in raw data.

The output of a mechanically scanned sonar results from the combination of a sequence of individual measurements made while it sweeps its nominal sensor axis through a sector spanning an area of interest. The ST1000, for example, has a profiling mode in which it generates scans that contain 400 individual range measurements spaced at equal intervals of 0.9° over a full 360° panorama. For each individual measurement in a scan, the device follows three steps:

1. transmit an outgoing pulse or *ping* of 60 μs duration;
2. listen to the received signal either until an echo is detected, or until a timeout interval has elapsed—proportional to a user configured maximum range value;
3. reorient the transducer for the next measurement.

Accordingly, the ability to detect the presence of an echo pulse in a received waveform emerges as a critical issue affecting the performance of a profiling sensor. Incidentally, a profiling sensor does not *measure* the range to an object, but rather the round trip travel time of an acoustic pulse; it *calculates* the range, using an assumed value for the speed of sound. Because the two terms are equivalent, differing only by a multiplicative constant, we will use them interchangeably.

3.1.1 Echo Detection and Time Localization

One possible strategy for estimating the range to a target uses a *matched filter* to detect the echo caused by the target and estimate its arrival time [14]. Based on the assumption that the received waveform highly correlates with the outgoing pulse only when in the vicinity of an echo, this approach identifies the local maxima of sufficient amplitude in the convolution between the received signal and a copy of the transmitted pulse. A matched filter classifies any local maximum that exceeds a predetermined threshold as a return. Very accurate travel time estimates, independent of the received signal strength (RSS), are possible using matched filters.

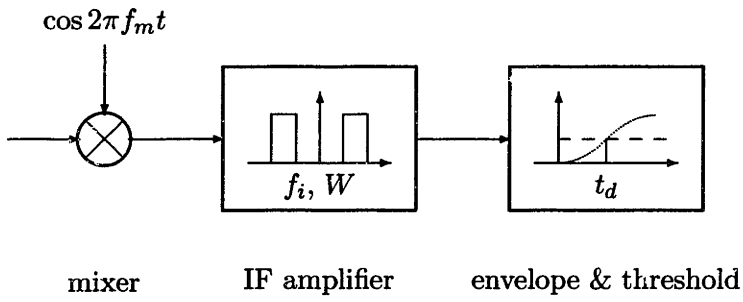


Figure 3.1: Leading edge detector block diagram. The ST1000 uses a superheterodyne receiver to track the envelope of the received signal. It demodulates the receiver signal by mixing it with an oscillator at $f_m = 1.7$ MHz and filtering with a $W = 20$ kHz bandpass filter centered at the difference frequency $f_i = 455$ kHz. The limited bandwidth of the IF amplifier causes a detection delay t_d due to the rise time of the filter output.

The ST1000 employs a less accurate, though simpler method of echo detection. It contains a superheterodyne receiver, whose essential elements are shown as a block diagram in figure 3.1, which tracks the envelope of the received signal. This type of detector demodulates the raw receiver signal by mixing it with a phase locked oscillator at a modulation frequency of $f_m = 1.7$ MHz. The resulting signal has spectral components centered at the respective sum and difference frequencies of 2.95 MHz and 455 kHz.[1] It then passes through a ceramic bandpass filter, centered at an intermediate frequency (IF) of 455 kHz, thus losing its high frequency component.

The detector reports the presence of an echo when the *envelope* of the IF amplifier output exceeds a predetermined threshold. The ideal square wave envelope, in an ideal broad band circuit, exceeds the threshold immediately upon its arrival. Because of the limited 20 kHz bandwidth of the IF stage in the real circuit, however, the envelope has a nonzero rise time which delays its detection in a manner inversely proportional to the received signal strength (RSS).

Recall from linear circuit theory that the reciprocal of the bandwidth approximates the rise time of the impulse response of a filter. For an IF filter with 20 kHz bandwidth, we expect to experience rise times approximately equal to 50 μs. Indeed, the close up view of

[1]The ST1000 has a carrier frequency of 1.25 MHz, as previously indicated.

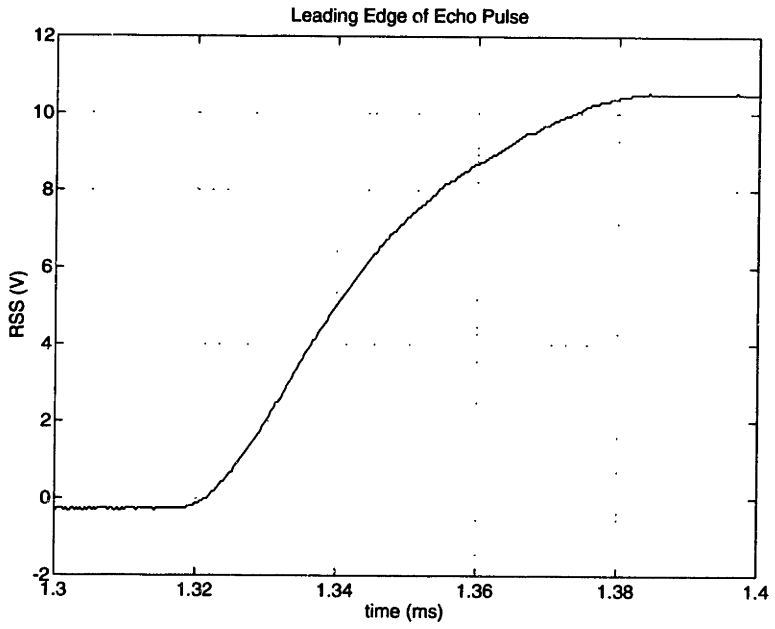


Figure 3.2: Leading edge of ST1000 echo pulse. In this close up view of the leading edge of a typical ST1000 echo pulse (originally shown in figure 2.5 on page 34) the observed rise time corresponds to about 60 μs, nearly equal to the theoretical value of 50 μs.

the leading edge of a typical ST1000 echo pulse shown in figure 3.2 confirms this expectation. The oscilloscope trace reveals a rise time of approximately 60 μs, nearly equal to what linear theory predicts.

Beyond the bandwidth alone, the detection delay depends in general both on the amplitude of the received signal and on the value of the detection threshold. Figure 3.3 shows a sequence of targets with increasing strength to illustrate the effect of the former. With respect to the latter, the delay caused by the time necessary for an envelope to exceed a very high threshold causes the sensor to overestimate the range to the target. Too low a threshold is worse, however, because extraneous noise can trigger false detections. We would rather attempt to correct for a predictable detection delay than face the consequences of dealing with a very large number of spurious returns.

The amplitude of the received signal depends on many factors which involve the orientation of the transducer relative to the target, the strength of the target, and the transmission losses during propagation, to name a few. Postponing the discussion of the issues involving

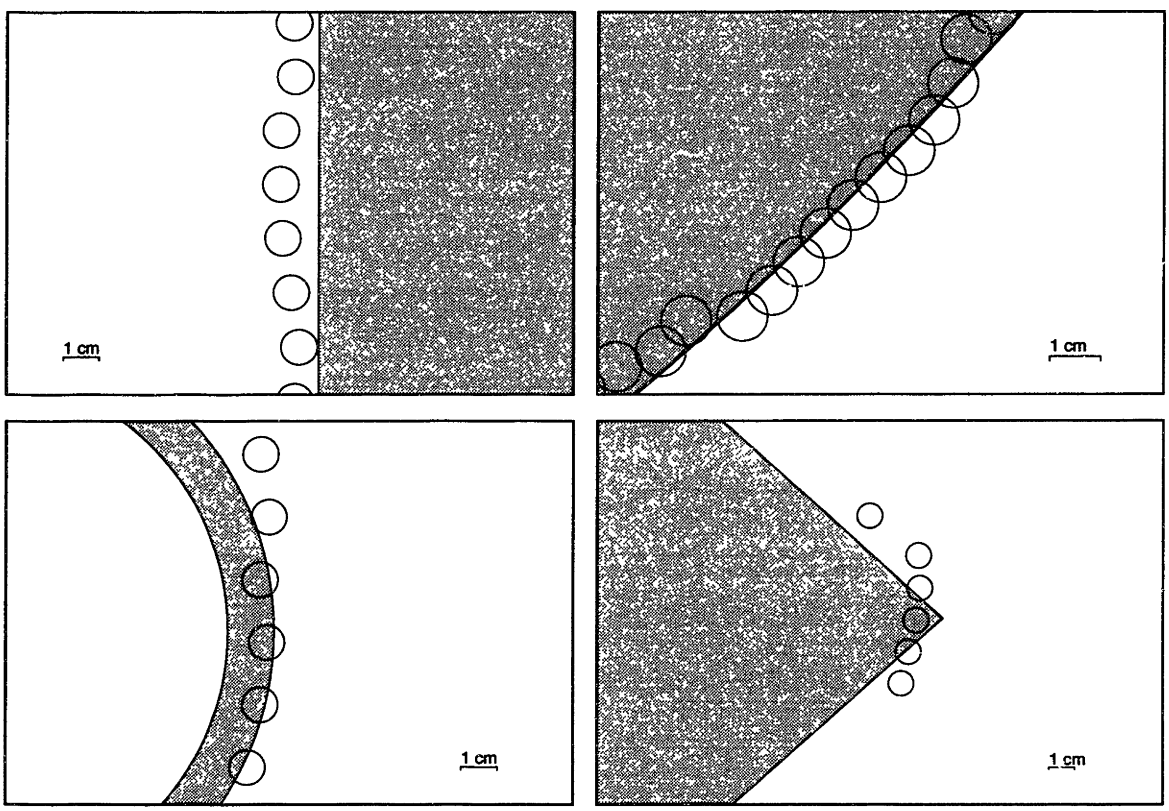


Figure 3.3: Sequence of targets with increasing detection delay. *Top row* shows range dots corresponding to early detection, indicative of strong returns. *Bottom row* shows range dots which penetrate into the objects, indicating weaker returns. (Detection of a strong return appears to precede its arrival due to the calibration procedure which averages the effects of strong and weak returns.)

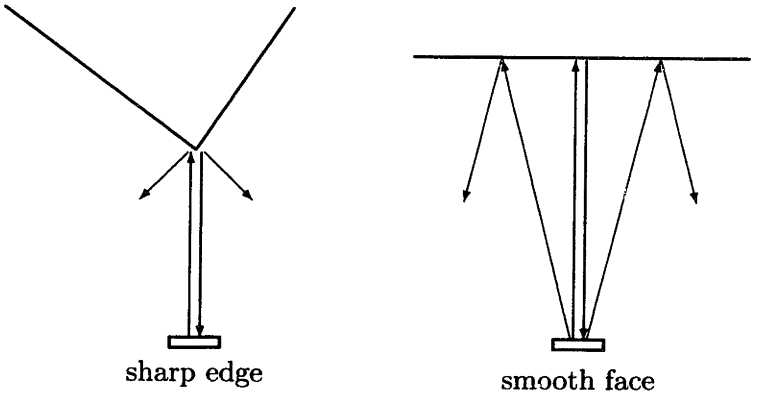


Figure 3.4: Isolated targets. Though governed by different physical phenomena (diffraction and specular reflection) both cases shown here backscatter from a single point on the target. *left*, a sharp edge *actually is* an isolated point. *right*, a sensor detects a smooth face which only when its axis coincides with the surface normal.

target strength and transmission losses until Section 3.2, we now examine how the relative orientation of the transducer influences the nature of the measurements.

3.1.2 Relative Transducer Alignment

Before we develop a model for the features we expect to see in sonar scans, we must consider two geometric configurations which a sensor regularly encounters. The left side of figure 3.4 shows a transducer, in an arbitrary location, oriented directly toward a sharp edge target accompanied by several of the nominal ray paths associated with this configuration. The edge forms its echo through the process of *diffraction*, scattering acoustic rays in all directions. The right side of figure 3.4, on the other hand, shows three different rays incident on a smooth face. Because of *specular reflection*, in which the angle of reflection equals the angle of incidence, only the ray pair that correspond to the transducer shown produce an actual return.

The challenge in defining a feature stems from the very existence of the different physical phenomena which govern the process of echo formation. Different mechanisms cause significant differences in how targets appear. Consider the far wall in figure 2.3 (on page 30), for example. Due to its rough surface, we model it as a collection of point scatterers which radiate acoustic energy similar to a sharp edge, producing a distributed response over a broad range of incidence angles, rather than just the one aligned with the surface normal.

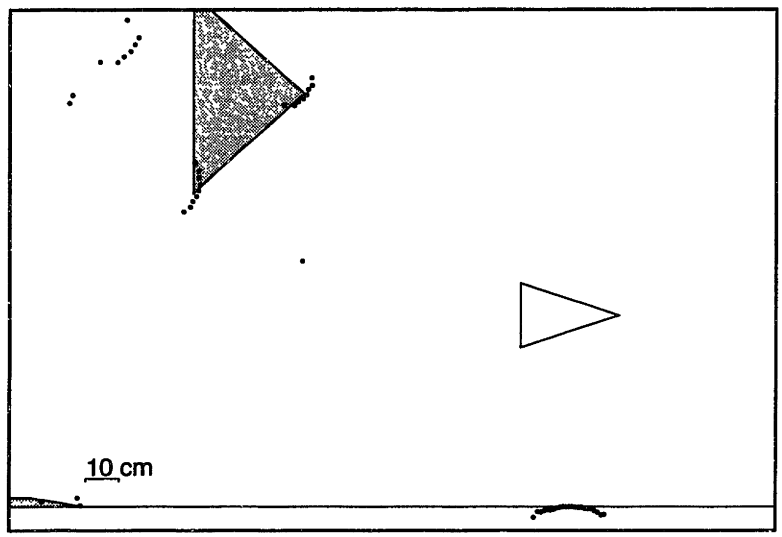


Figure 3.5: Close up view of returns from smooth face and sharp edges. The unfilled triangle indicates the location of the sensor and circles represent individual measurements. The returns from a glass window in the foreground and the from the sharp edges of a triangular prism to the left contradict the expectation that these targets produce single returns.

For both cases shown in figure 3.4, however, a return only exists when the transducer axis meets a strict alignment condition. This argument intuitively suggests that isolated targets will appear as single returns in our range scans. Unfortunately, figure 3.5 directly contradicts this prediction. It shows that a multitude of returns correspond to the plate glass window and the two sharp edges as the ST1000 sweeps past. The intuition fails because it assumes that the transducer has zero beamwidth. In the real world, however, even a highly directional transducer has a finite beamwidth.

3.1.3 Beam Pattern

As previously indicated, the ST1000 contains a circular piston transducer with a diameter of 50 mm. For continuous wave (CW) sonar, the standard model for such a transducer predicts a beam pattern $b(\theta)$ in terms of the first-order Bessel function,

$$b(\theta) = \left(\frac{2J_1(kR_p \sin\theta)}{kR_p \sin\theta} \right)^2, \tag{3.1}$$

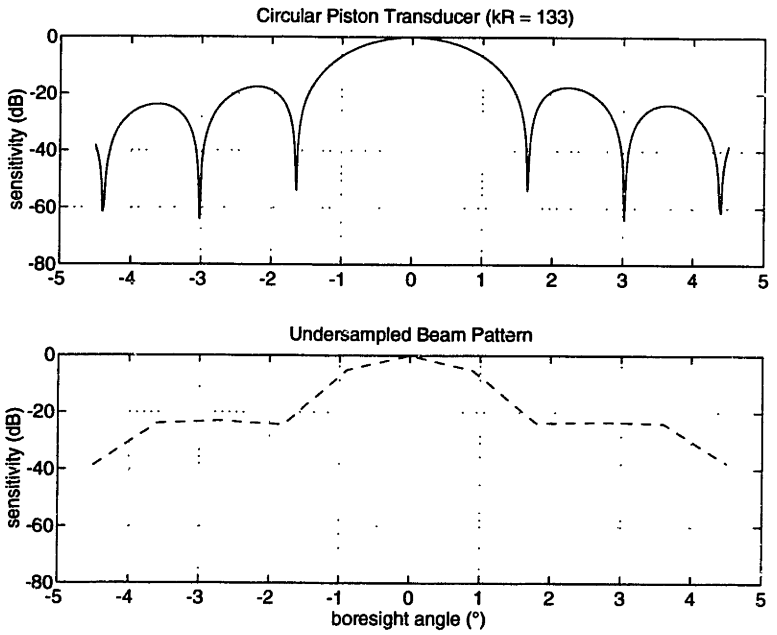


Figure 3.6: Effective beam pattern of the ST1000. *Top*, a circular piston transducer with $kR_p = 133$ has an idealized beam pattern in which the first null occurs at 1.65° and the 3 dB down point occurs at 0.7°. *Bottom*, the 0.9° stepper motor spatially undersamples the ideal pattern to produce an effectively wider beam of about 7°.

where θ is the angle from the sensor axis (the *boresight* angle), k is the wavenumber, and R_p is the radius of the transducer. Though the pattern function in equation (3.1) follows from arguments that exploit the nature of CW transmission, we feel that it models the true beam pattern for a single ping of the ST1000 reasonably well. The sensor has a transmitted pulse duration of 60 μsec, which at a frequency of 1.25 MHz spans 75 full cycles.

Applying equation (3.1) to the parameters of the ST1000 (resulting in $kR_p = 133$) produces the nominal beam pattern shown in the top half of figure 3.6. The first null in the transducer beam pattern occurs approximately 1.65° off axis. The stepper motor in the ST1000 moves the transducer axis in 0.9° increments, however, *spatially undersampling* the ideal beam pattern. The effective beam pattern as shown in the bottom half of figure 3.6 results from spatial aliasing.

Assuming a sufficiently sensitive detector, we should observe a sequence of returns, which approximately spans the effective beamwidth, that have identical range. The top half of

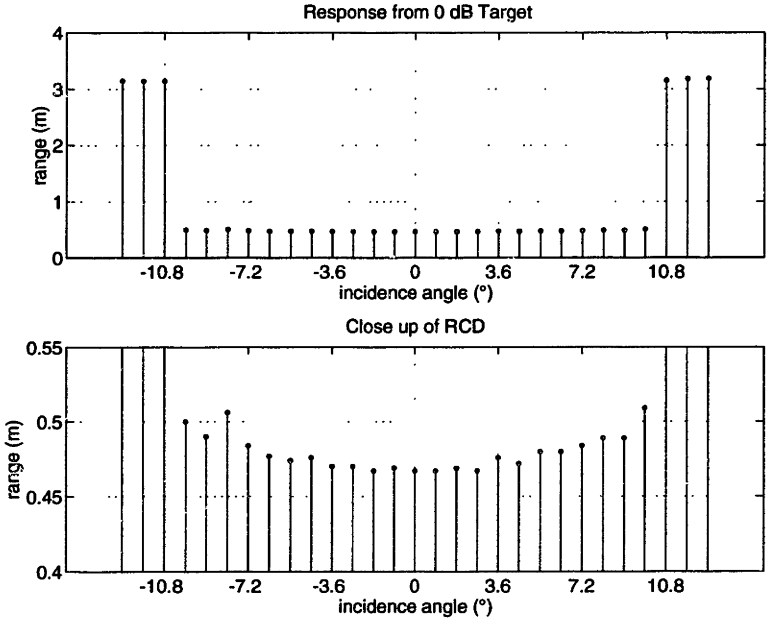


Figure 3.7: Returns from a target with a smooth face. *Top*, at the scale shown, the returns appear to have constant range. *Bottom*, in a close up view, the returns vary in range over the full sequence because the peripheral ones experience more delay due to their reduced signal strength.

figure 3.7 confirms this expectation. Using a 4 m length scale the returns associated with the glass window in figure 3.5 all appear to have identical range. This phenomena closely matches a similar one in terrestrial robotics [39, 41]. Further supporting this result, Leonard and Durrant-Whyte [44] observed sequences of individual returns which follow circular arcs in a polar plot and thus coined the term *region of constant depth* (RCD) to describe this type of feature.

A closer look at the measurements using a much finer scale, shown in the bottom half of figure 3.7, reveals that the returns are not exactly constant, but rather vary in range by a small amount. The detection delay associated with the lower signal strength found in the peripheral returns causes the feature to deviate from a circular arc as the incidence angle increases. Even so, the RCD representation suits our needs. Now that we have defined their model, we must implement a method by which to extract these features from raw sonar scans.

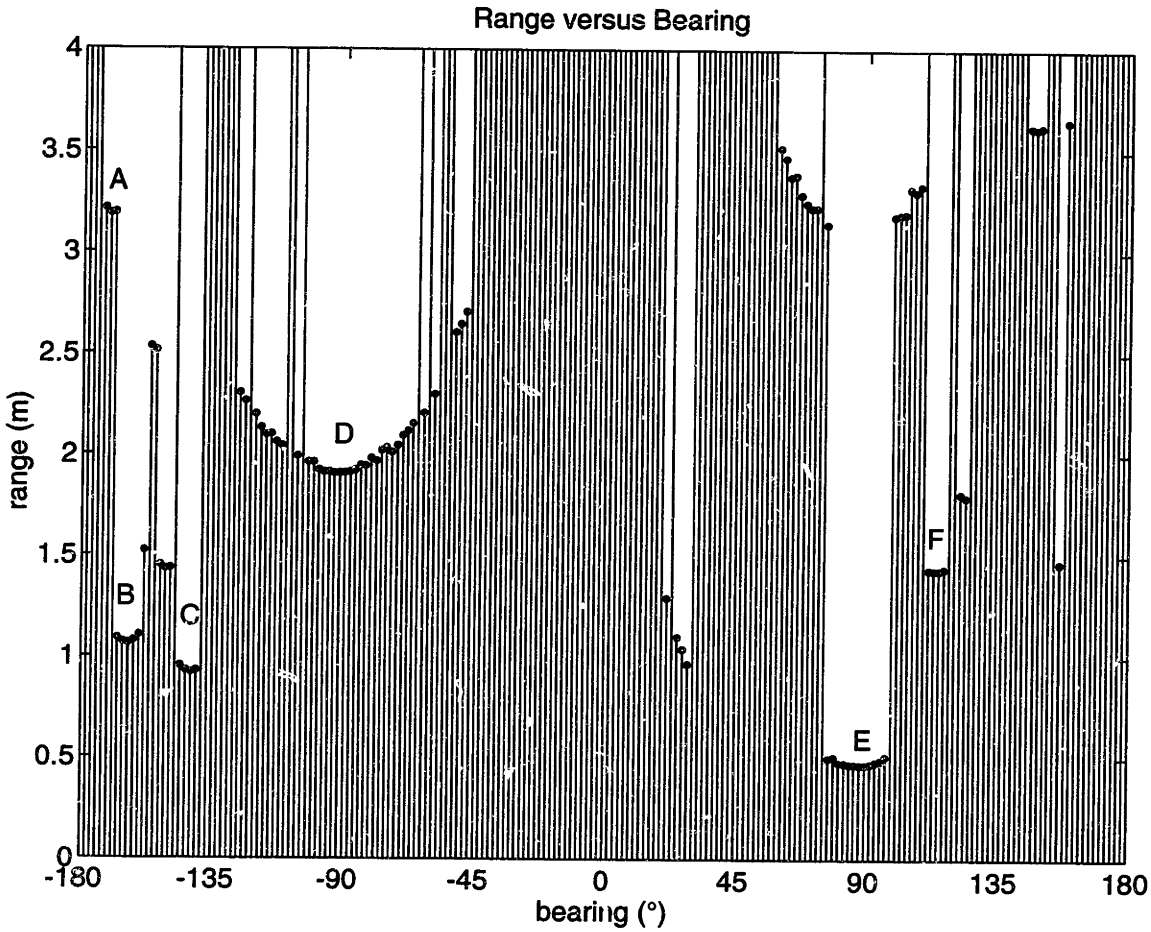


Figure 3.8: A scan shown as a "signal" of range versus bearing. Local minima correspond to features in the underwater scene: A is the cylinder, B and C are triangle vertices, D is the far wall (rough), E is the glass window (smooth), and F is the reflection of a triangle vertex in the glass window (cf. figures 2.3 and 2.4 on pages 30 and 31).

3.1.4 RCD Extraction Filter

A possible strategy for extracting RCDs follows from considering a raw scan as a type of range versus bearing signal. Canny [15] has devised a class of one-dimensional convolution filters which detect and localize ridges in a signal of this type. (A ridge is a sequence of similar returns with range greater than the surrounding ones.) Since each RCD corresponds to a local minimum—a valley or *inverted* ridge—in the plot of figure 3.8, we could extract them using this approach. Unfortunately, the ridge detection operators respond equally well to the isolated edges in the signal. In order to ensure that the particular feature which

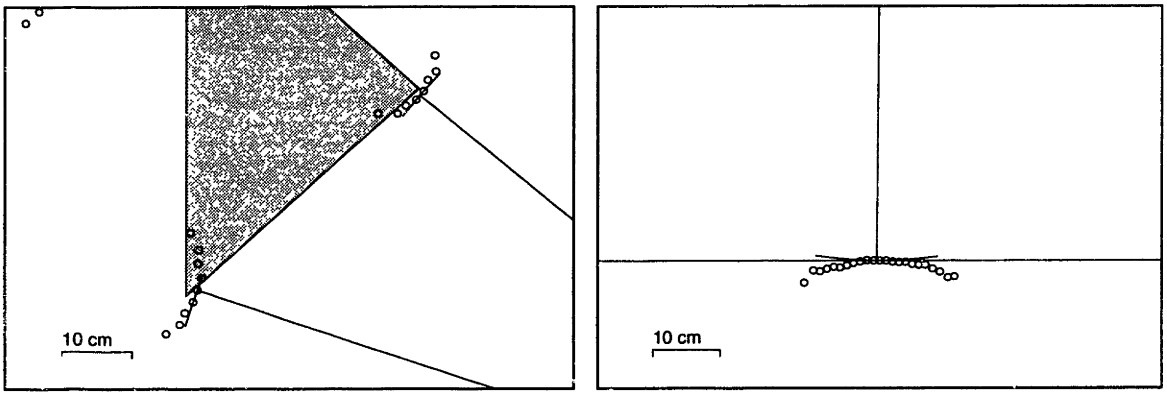


Figure 3.9: Close up views of edge and face returns. *Left*, the sharp edges produce 9 returns each. *Right*, the smooth face produces 23 returns. In general the number of returns which result correlates to the strength of the target. The circular arcs indicate the angular extent of the RCDs for $\tau = 1$ cm and the rays connecting them to the sensor location indicate their estimated bearing angle.

causes a strong response is indeed the desired ridge, one must combine the results with those of an edge detection operator in some heuristic manner. We observe that the shortcoming of this approach results from its failure to account for the relevant acoustics which govern the process of echoranging. Thus we seek a more natural alternative—one which relates the underlying physical principles to the observed phenomena.

Based on the observation that the individual returns from a target vary in range by a small amount, previously shown in the bottom half of figure 3.7, we introduce a tolerance parameter τ to compare consecutive measurements. An RCD, more formally, contains a set of consecutive returns which differ in range from their adjacent neighbors by less than this tolerance,

$$|r_{i\oplus 1} - r_i| < \tau, \tag{3.2}$$

where $\oplus$ indicates cyclical addition. For a scan with m returns, $m \oplus 1 = 1$.

In the close up views shown in figure 3.9, we see a variation in the number of returns that correspond to different features. Because it ignores the number of returns in a resulting group, Equation (3.2) describes a condition used merely to identify which returns belong together. We require an additional condition that acts as a threshold on the minimum number needed to achieve a particular level of confidence that an RCD corresponds to a world feature.

RCD Yield for Figure 2.3 Scan					
	w (°)				
	0.9	1.8	2.7	3.6	4.5
1	26	8	2	0	0
2	31	9	2	0	0
5	34	14	6	4	2
10	31	17	11	8	4
15	30	19	13	8	6

Table 3.1: RCD yield as a function of extraction parameters. As we vary the tolerance and threshold parameters τ and w, the number of RCDs detected changes. A balance exists between *quantity* of RCDs and *quality* of RCDs.

Consequently, we introduce a threshold parameter w, to establish the minimum angular width an RCD must span. If α_1 and α_2 refer to the bearing angles belonging to the two outermost returns, the condition

$$|\alpha_2 \ominus \alpha_1| > w \tag{3.3}$$

(where $\ominus$ indicates principal valued angular subtraction) acts as a noise suppressor. We can choose a threshold small enough to allow an RCD with only two returns ($w = 0.9°$), for example, or we can increase w to establish a more stringent extraction filter.

Varying these parameters vastly affects the quantity of RCDs found, as table 3.1 indicates for the scan shown in figure 2.3 on page 30. Figure 3.10 shows the resulting RCDs extracted from our example scan using the values of $\tau = 1$ cm and $w = 3.6°$, which represent good choices based on both empirical observation and the upcoming arguments in Section 3.2. In choosing these parameters, we make a trade between *quantity* and *quality* of RCDs. As shown in figure 3.10, a few undesirable RCDs meet the established criteria, but we will deal with these in Chapter 4.

The extraction filter itself merely establishes the criteria for identifying which returns constitute an RCD. Values for bearing and range still remain undetermined. We calculate the range by taking the minimum of the individual returns in the RCD, for the obvious reason that these returns result from the earliest detection, implying strong echo (and confidence) levels. The angular centroid of these minimum range individuals supplies the bearing angle of the RCD. The arcs and rays in figure 3.9 indicate these values for the sharp edges and smooth face.

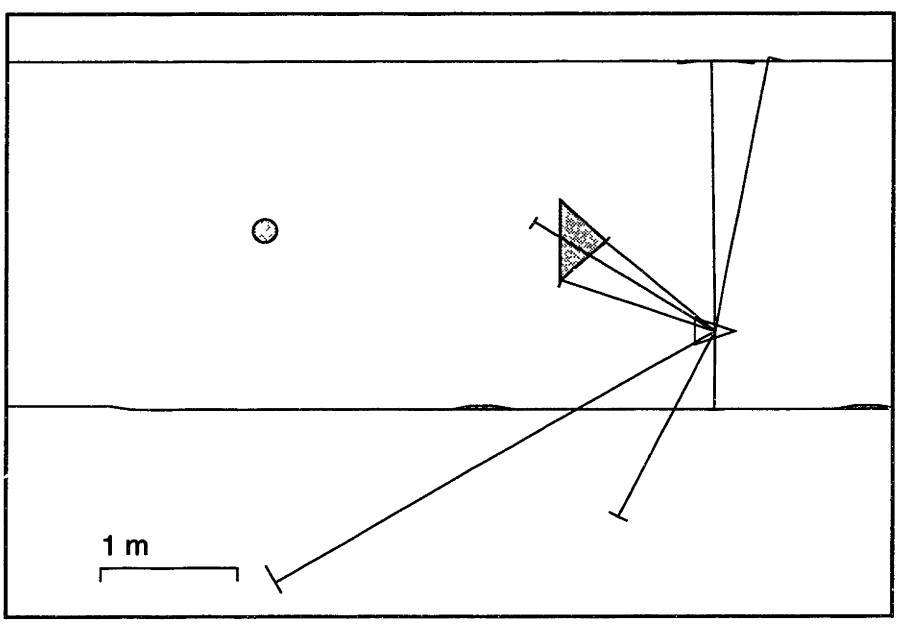


Figure 3.10: Sample RCDs extracted using typical parameters. This figure shows the RCDs extracted from the scan shown in figure 2.3 on page 30 using $\tau = 1$ cm and $w = 3.6°$, which represent good choices based on empirical observation and physical arguments. *Clockwise from top*, two originate from the rough wall, one from the glass window, two from reflections of features in the window, and three from the triangle. The middle one of these three corresponds to a ray path which passes through the lower right face and reflects from the left face, finally terminating on the upper right one (wow).

3.2 Receiver Echo Level using Active Sonar Equations

Detecting the presence of a target in a received waveform depends on the projector and receiver characteristics, the transmission medium, and the target strength. The classic active sonar equation for the echo level EL from Urick [78],

$$\mathrm{EL} = \mathrm{SL} - \mathrm{TL}_1 + \mathrm{TS} - \mathrm{TL}_2 + \mathrm{DI}, \tag{3.4}$$

quantifies the detectability of a target with strength TS for a projector of strength SL, a medium with outbound and inbound transmission losses TL_1 and TL_2, and a receiver with directivity index DI.[2] We deliberately differentiate between the outbound and the inbound transmission losses because the shape of the target influences the latter.

The ST1000 profiling system uses the same transducer to project and receive, thus (for a unit source of 1 W radiated power) we substitute $\mathrm{SL} = 171.5 + \mathrm{DI}$.[3] For circular piston transducers, a simple model for the directivity index gives $\mathrm{DI} = 10\log(kR_p)^2$, where k is the wavenumber and R_p is the radius of the transducer. Using this model, the ST1000 has a resulting DI of 42 dB. In light of these substitutions, equation (3.4) simplifies to

$$\mathrm{EL} = 255.5 - \mathrm{TL}_1 + \mathrm{TS} - \mathrm{TL}_2, \tag{3.5}$$

per watt of projected power. In the simplest sense, EL must exceed some threshold or the target remains undetected. Given detection, EL influences the range value used to calculate the location of the target, as previously discussed in Section 3.1.1. A weaker signal takes longer to exceed the detection threshold than a stronger one.

3.2.1 Target Strength

Two properties of a target, the shape and the material, exhibit the most important influence over how much acoustic energy scatters back toward the receiver. Although the most

[2]The classic sonar equations specify distances in *yards*, due to the convention that when converting to dB, the reference intensity be given at 1 yard. Throughout this discussion, we follow the convention in modern acoustics to adopt SI units, which specifies a 1 m reference distance.

[3]The constant term 171.5 results from the characteristic impedance ρc of sea water. Fresh water yields 171.48, but why split hairs?

general approach to determining the receiver intensity I_r results from solving the integral equation (2.4) to calculate the full pressure field, we can approximate receiver intensity by making some simplifying assumptions about the strength of a target.

In the approach presented here, we decouple the contributions from material reflection characteristics and geometric diffraction, consider them separately, and then combine their individual contributions into a composite result. Towards this end we write the receiver intensity in terms of the incident intensity I_i,

$$I_r(\hat{r}) = I_i(r)RA(\hat{r}, r), \tag{3.6}$$

where r is the range from the projector to the target, $\hat{r}$ is the range from the target to the receiver, R is the complex reflection coefficient which depends on the material characteristics, and $A(\hat{r}, r)$ describes the geometric interaction between the incident field and the target shape.

Although we find it useful to consider the true shape of the incident wavefront when describing the geometric effects of the target shape, we benefit substantially by making a plane wave simplification when studying the material reflection coefficient of the target. We achieve this simplification by making the assumption that targets are far away from the transducer. The wavefront from all projectors resembles that from a point source beyond a certain range, though it varies for each projector. This point marks the transition from the near or Fresnel field to the far or Fraunhofer field. Coates [18] gives the far field for a circular piston of radius R_p at a range of R_p^2/λ, which is approximately 0.5 m for the ST1000.

Because we have located the target beyond the point of Fraunhofer transition in the projected field, we can represent the true *spherical* shape of the outbound wave in the vicinity of target insonification as a *plane* wave,

$$p(\boldsymbol{r}) = \hat{p}_0 e^{i\boldsymbol{k}\cdot\boldsymbol{r}}, \tag{3.7}$$

where $\hat{p}_0 = p_0/r_0$ and r_0 is the nominal range from the source. In this vicinity, the intensity of the field remains essentially constant, $I(\boldsymbol{r}) \approx I_0/r_0^2$. Incidentally, equation (3.7), which

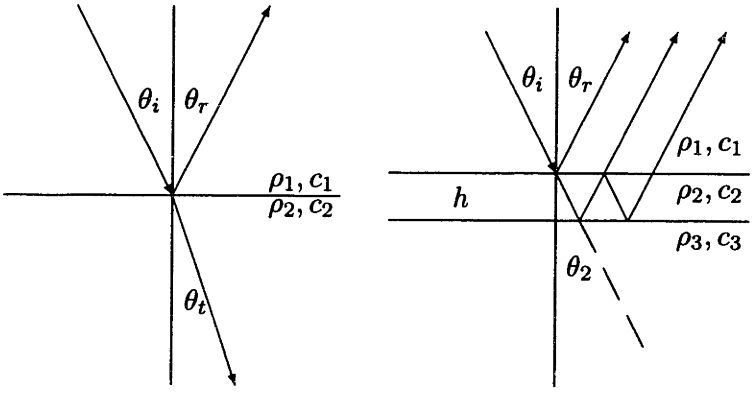


Figure 3.11: Plane wave reflections. *Left*, Snell's law, $c_1 \sin\theta_t = c_2 \sin\theta_i$, governs the reflection and transmission at the boundary of a homogeneous fluid half-space. *Right*, for a homogeneous layer, the reflected field is the infinite sum of components like the ones depicted. (The intermediate angle of transmission, θ_2, is confined to the layer.)

describes a wave propagating in the direction of vector $\boldsymbol{k}$ ($|\boldsymbol{k}| = k$) represents one solution to the Helmholtz equation (2.2). We further simplify the situation by assuming that the target itself is locally planar throughout the insonified region.

At a planar boundary between two homogeneous fluid half-spaces (ρ_1, c_1 and ρ_2, c_2) we write the *Rayleigh* reflection coefficient,

$$R_{12} = \frac{\zeta_2 - \zeta_1}{\zeta_2 + \zeta_1} \tag{3.8}$$

in terms of the *acoustic impedances* of the upper and lower media, $\zeta_1 = \rho_1 c_1 / \cos\theta_i$ and $\zeta_2 = \rho_2 c_2 / \cos\theta_t$, as shown to the left in figure 3.11. For a homogeneous layer of thickness h, the net reflection coefficient R,

$$R = \frac{R_{12} + R_{23} e^{2ik_2 \cos\theta_2 h}}{1 + R_{12} R_{23} e^{2ik_2 \cos\theta_2 h}}, \tag{3.9}$$

comes from the infinite sum of contributions from multiple half-space reflections, a few of which are shown to the right in figure 3.11.

At this point, we must ask if these formulae, which describe longitudinal compression waves in a fluid, remain valid when echoranging solid targets? After all, solids can contain transverse shear waves in conjunction with the waves described here. The excitation of these waves leads to an effective softening of the boundary, which results in a reduction in the magnitude of the reflection coefficient [10].

Indeed this analysis does relate, because of the alignment conditions characteristic to isolated targets. Normal incidence imposes conditions under which solids behave similarly to fluids. We deliberately choose to ignore the softening effect because normally incident insonification does not excite any shear waves. Further, the dependence on the angle of incidence θ_i for the solid layer reflection coefficient is less important than in a fluid of similar impedance. It suffices then, to approximate the reflection of a sonar pulse from a target by considering a plane wave incident on the target boundary.

As an example, we shall now take a closer look at the normally incident reflection coefficient for aluminum. Used in the manufacture of the targets for the experiments described in Chapter 5, rolled aluminum has a density ratio ($m = \rho_a/\rho_w$) of about 2.7 and index of refraction ($n = c_w/c_a$) of about 0.23 compared to *fresh* water. Except when the thickness of the layer corresponds to a value near an integer number of half-wavelengths, most of the energy reflects back from the layer, as shown in figure 3.12. Considering only the energy level, determined solely by the magnitude of the reflection coefficient, leads us to believe that such a layer presents a target nearly ideal for echoranging.

We must include the phase of the reflection coefficient in our analysis, however, for it plays an important role in this situation. We can write the complex reflection coefficient as $R = |R(\omega)|e^{i\phi(\omega)}$ and expand the phase $\phi(\omega)$ in a Taylor series about the carrier frequency ω_0. Neglecting higher order terms, we find

$$\phi(\omega) \approx \phi_0 + \phi_0'(\omega - \omega_0), \tag{3.10}$$

where $\phi_0 = \phi(\omega_0)$ and $\phi_0' = d\phi(\omega_0)/d\omega$. We now isolate the term which is linear in ω to write the reflected spectrum,

$$p_r(\omega) = e^{i\omega\phi_0'}\hat{R}\, p_i(\omega), \tag{3.11}$$

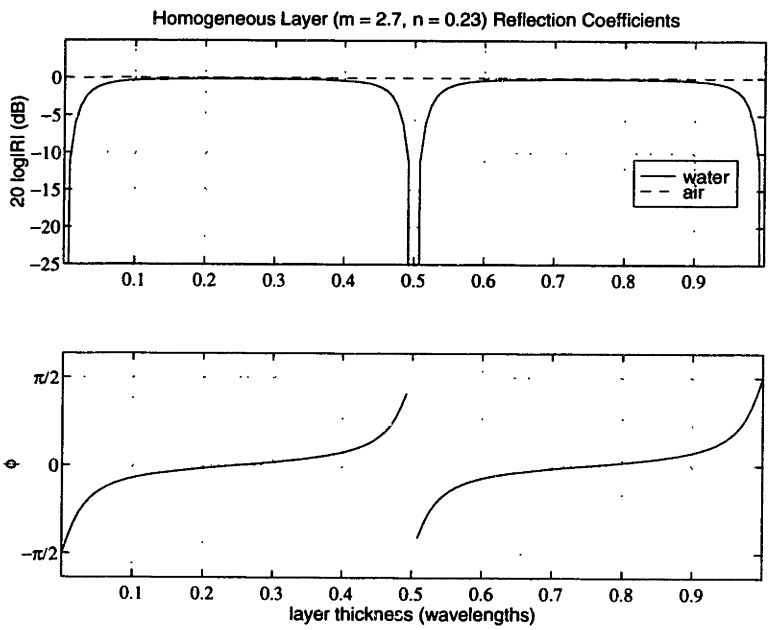


Figure 3.12: Reflection coefficient for aluminum layer. *Top* shows the magnitude and *bottom* shows the phase of the plane wave reflection coefficient at normal angle of incidence for a rolled aluminum layer. Two cases are presented, water-layer-water and water-layer-air. The layer thickness is normalized by the wavelength in the layer ($\lambda_{Al} = 5.13$ mm in aluminum at a frequency of 1.25 MHz).

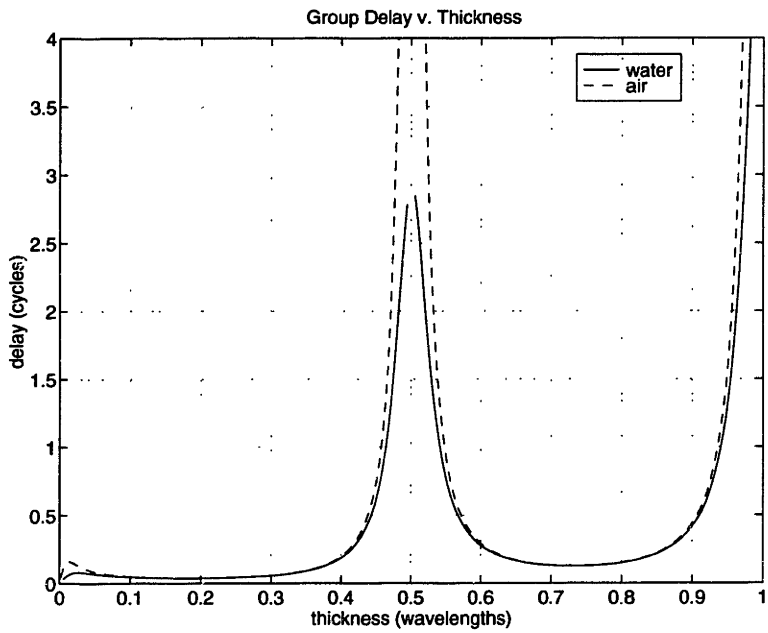


Figure 3.13: Comparison of group delays for an aluminum layer. Although an air-backed layer reflects more energy (cf. figure 3.12), its return is more delayed than a water-backed layer.

in which $\hat{R} = |R|e^{-i\omega_0\phi_0'}$. This result yields a time domain solution via the inverse Fourier transform,[4]

$$p_r(t) = \Re[\hat{R}\, p_i(t + \phi_0')], \tag{3.12}$$

in which $\Re$ denotes the real part of a complex signal. The reflected wave is a scaled version of the incident wave *delayed in time by* $-\phi_0'$. Using the same values for the material properties as before, figure 3.13 shows the time or *group* delay [69] as a function of layer thickness for both air- and water-backed aluminum layers.

Considering just the reflected energy, the air-backed layer appears the more acoustically ideal, with the target strength approximately equal to 0 dB everywhere. We observe through the group delay, however, that the time-of-flight value under this configuration contains larger variability and thus conclude that a water-backed layer better serves our needs, providing that the thickness does not correspond to an integer multiple of $\lambda_{Al}/2$.

[4]Recall from the time delay property of the Fourier transform that $F(t-\tau) \Leftrightarrow e^{-i\omega\tau} f(\omega)$ are transform pairs.

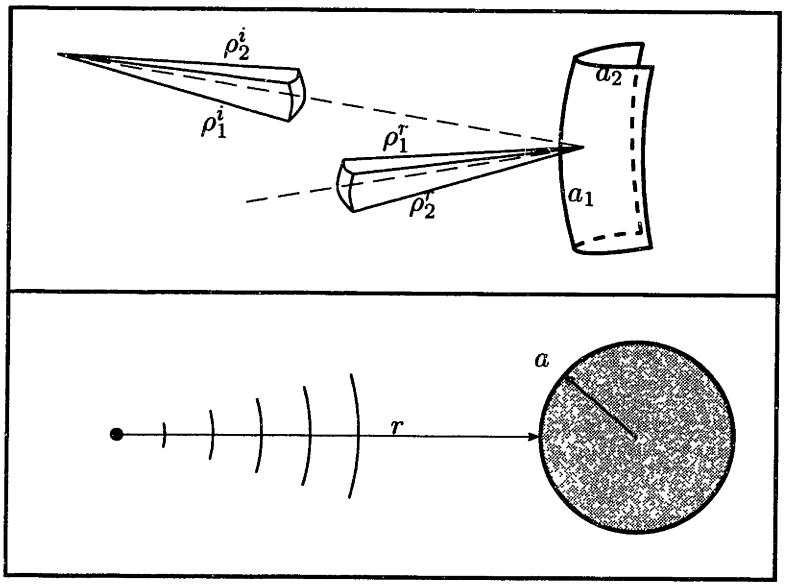


Figure 3.14: Schematic view of specular reflection. *Top*, under certain alignment conditions, the principal radii of the reflected field depend only on the principal radii of the incident field ρ_j^i and the surface a_j. *Bottom*, a cylinder of radius a at a distance r from the projector.

Returning now to our earlier expression for the target strength, equation (3.6), we explain the term $A(\hat{r}, r)$, which describes the geometric spreading of the reflected wave as a function of the distance $\hat{r}$ it propagates away from the target. (Recall that r is the distance from the projector to the target.)

When an incident field reflects from a smooth surface, the geometric theory of diffraction tells us that the resulting astigmatic ray tube geometrically spreads with factor A,

$$A(\hat{r}, r) = \frac{\rho_1^r \rho_2^r}{(\rho_1^r + \hat{r})(\rho_2^r + \hat{r})}, \tag{3.13}$$

where ρ_1^r and ρ_2^r, both of which depend on r, are the principal radii of curvature of the reflected wavefront [2, 49]. We show the geometry of this situation in the top half of figure 3.14. Under the normally incident alignment conditions previously discussed, the principal radii of curvature of the reflected wavefront depend only on the geometry of the incident wave, with principal radii ρ_1^i and ρ_2^i, and on the principal radii of curvature of the target surface, a_1 and a_2.

The general expression for the reflected wavefront radii of curvature is quite complicated,

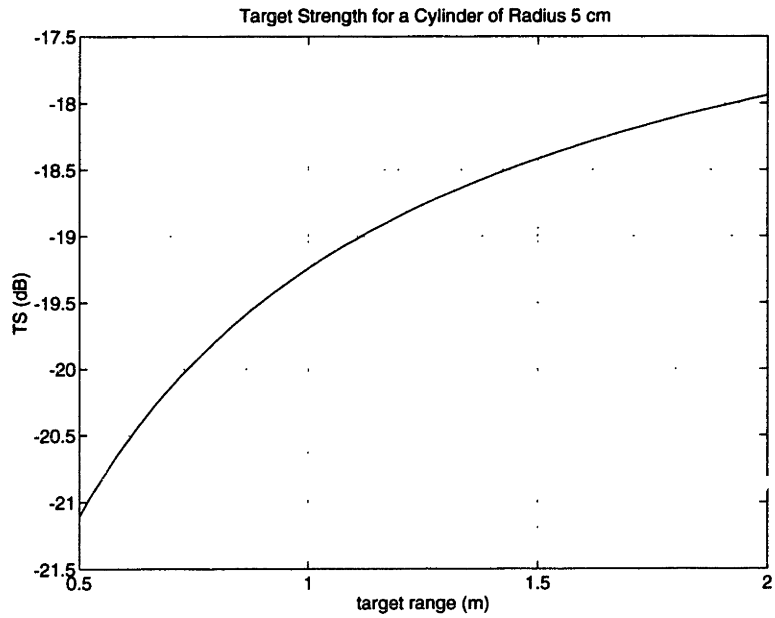


Figure 3.15: Target strength of cylinder with 5 cm radius. The shape of the target and the incident wavefront interact to determine the intensity of the reflected field. This figure shows how a spherical wave interacts with a cylinder of radius 5 cm. We assume perfect reflection from the cylinder and that the incident field is normal to its axis.

but for a spherical wave incident upon and normal to the axis of a cylinder, the reflected radii of curvature can be written

$$\rho_1^r = \frac{ar}{a+2r} \quad \text{and} \quad \rho_2^r = r, \tag{3.14}$$

where r is the range from the projector to the surface of the cylinder of radius a, as shown in the bottom half of figure 3.14.

In our earlier calculation of the complex reflection coefficient R, we approximated interaction between the incident wave and the target as a plane wave reflection. This approximation adequately characterizes a such a localized phenomenon. We feel that the spreading factor results from a more global phenomenon, however, in which a better approximation to the true incident wave geometry more closely applies.

In figure 3.15, we show the target strength of a perfectly reflecting cylinder with 5 cm radius as an example. This particular object represents a weak target in the spectrum of

those used for our experiments. The range dependence follows from the fact that the shape of the incident field influences the degree to which the reflected field defocuses.

3.2.2 Transmission Losses

When an acoustic wave propagates through a real medium, it loses intensity with increasing range traveled. Two principal mechanisms influence how the intensity drops: absorption and geometric spreading. In the former, the intensity loss in dB grows linearly with the distance traveled, with proportionality constant $\alpha_{\log}$. Models exist to calculate $\alpha_{\log}$, known as the logarithmic absorption coefficient, but it suffices to use an approximate value of $\alpha_{\log} = 100$ dB/km for a frequency of 1.25 MHz in fresh water. Using this value, the total round-trip loss due to absorption predicted for our experiments does not exceed 0.3 dB, an insignificant level. As a result, the effect of geometric spreading dominates the transmission loss terms.

We make the same arguments as earlier regarding the far field placement of targets to model the outgoing wave as that from a point source. This behavior results from the close correspondence between the far field approximation and the solution to the Helmholtz equation (2.2) for a point source at the origin,

$$p(\boldsymbol{r}) = p_0 \frac{e^{i\boldsymbol{k}\cdot\boldsymbol{r}}}{|\boldsymbol{r}|}, \tag{3.15}$$

which describes a spherical wave whose intensity decays as $1/|\boldsymbol{r}|^2$ in a lossless medium.

The combination of both absorption and spherical spreading leads to an expression for the net outbound transmission loss,

$$\mathrm{TL}_1 = 20 \log r + \alpha_{\log} r, \tag{3.16}$$

shown in figure 3.16 over the limited range of distances used in our experiments.

The inbound transmission loss contains the same absorption term as the outbound loss. It differs from the outbound only in that the presence of the target has influenced the principal radii of curvature of the reflected wavefront, as previously indicated. Rather than simply model the inbound wave as if spherically spreading from a point source, we use the

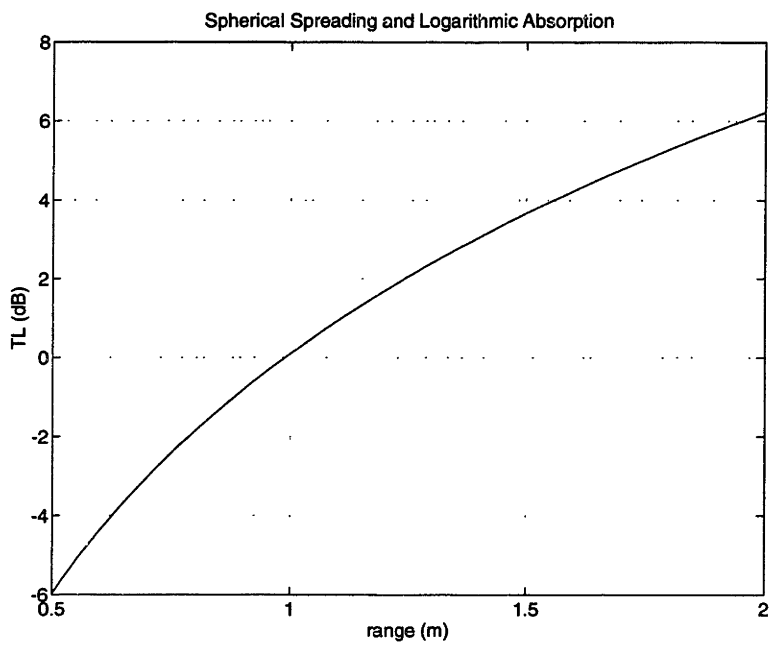


Figure 3.16: Outbound loss for spherical spreading and logarithmic absorption. For the limited ranges shown above, logarithmic absorption with $\alpha_{\log} = 100$ dB/km contributes very little to the overall loss.

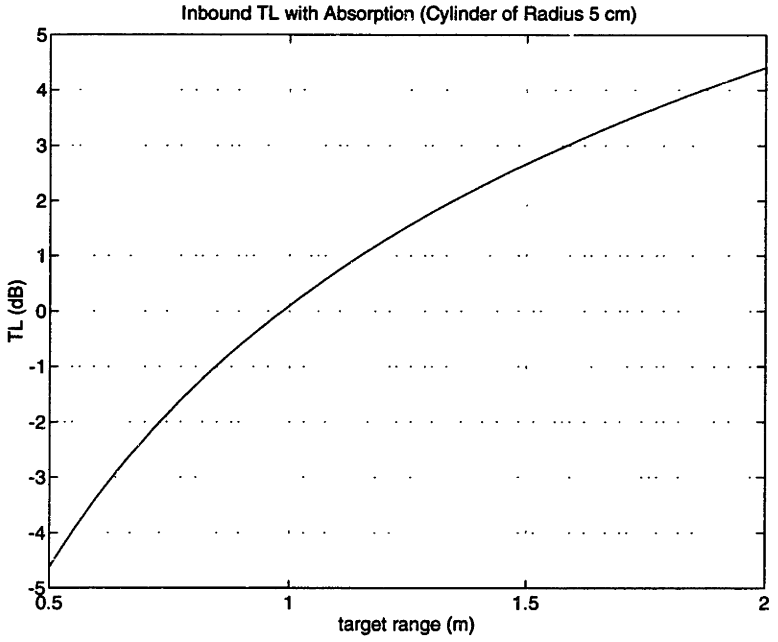


Figure 3.17: Inbound transmission loss due to cylindrical target. An incident field with spherical wavefront insonifies a cylindrical target of radius $a = 5$ cm. The logarithmic absorption coefficient of $\alpha_{\log}$ is 100 (dB/km). This plot only applies to configurations where the receiver distance equals the projector distance, as in our *monostatic* system.

spreading factor $A(r, \hat{r})$ determined in equation (3.13),

$$\mathrm{TL}_2 = 10 \log A(r, \hat{r}) + \alpha_{\log} \hat{r}, \tag{3.17}$$

in terms of the receiver distance $\hat{r}$. For the same 5 cm radius perfectly reflecting cylinder introduced above, figure 3.17 shows the inbound transmission loss for a configuration with equal receiver and projector distances. A *monostatic* system such as the ST1000, in which the receiver and projector *locations* coincide, clearly satisfies this condition.

All of the terms in equation (3.4) for the echo level have now been discussed. The top half of figure 3.18 shows the overall EL for two targets, one strong and one weak. The strong target is a 0.7 mm flat aluminum plate, with a thickness ratio $h/\lambda_{\mathrm{Al}} = 0.13$ (in aluminum, $\lambda_{\mathrm{Al}} = 5.1$ mm). For the weak one we keep the same 5 cm radius cylinder previously presented only adding to it wall thickness 5 mm. The bottom half of figure 3.18 assumes

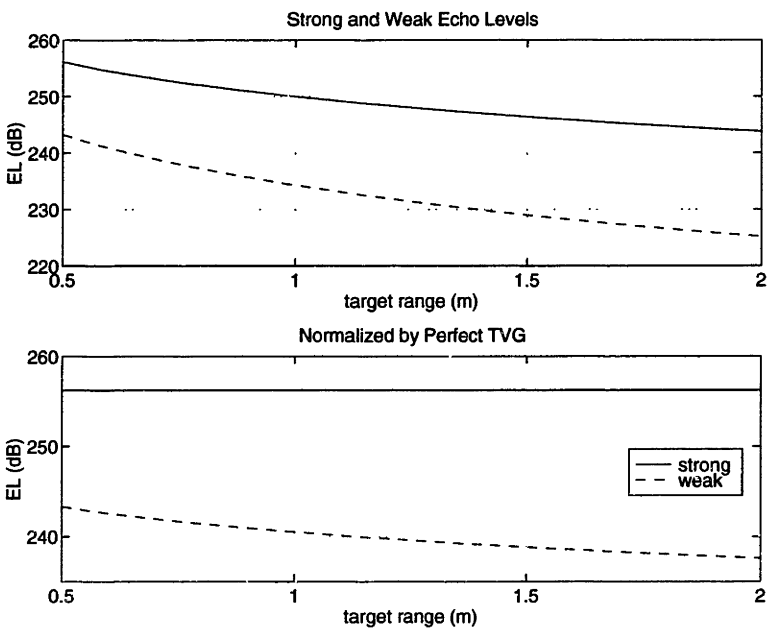


Figure 3.18: Comparison of EL for strong and weak targets. *Top*, an aluminum plate 0.7 mm thick has a higher EL than a 10 cm diameter aluminum cylinder with wall thickness 5 mm. *Bottom*, the same two targets, but with a TVG applied to correct for both logarithmic absorption and spherical spreading.

that the sonar device contains a time-variable gain (TVG), a strategy that increases the gain applied to the received signal as time elapses. The model used to determine the TVG assumes logarithmic absorption and spherical spreading.

We would like a sensor capable of making consistent range measurements independently of the strength of the target. Failing that, we conclude from this analysis that consistent measurements of objects that exhibit a 20 dB range in target strength suits our purposes. The ST1000 shows a small variation throughout this range, but in practice we have found it consistent enough to produce measurements of sufficient accuracy for us to achieve the results shown in Chapter 5.

3.3 Summary

We have just described how a simple echoranging device generates individual returns. Our discussion included the process by which the device detects the presence of a target in the received signal. We have considered how this process can be influenced by the type of target, the relative transducer orientation, and the beam pattern of the sensor.

Consistent with these issues, we have offered the circular arc or RCD as the canonical feature found in the panoramic scans of the ST1000. The RCD model provides the raw observations, through estimates of its bearing and range, for the second part of our algorithm—tracking and data association. The next chapter focuses on this issue.

Chapter 4

Tracking and Data Association[1]

This chapter addresses how our algorithm partitions observations according to their origins in the world. It begins with a brief introduction of multitarget tracking in general, and then discusses several approaches found in the literature. We select the model-based multiple hypothesis tracking filter and discuss the specific extensions that we develop to meet the needs of tracking curved surfaces. At the end of the chapter we apply the filter to example data from our underwater experiments and discuss the results.

4.1 Introduction

The field of tracking and data association was born out of a need to process the vast amount of information generated by the highly sensitive sensors found in modern surveillance systems. In air and naval defense and in air traffic control, for example, radar and sonar arrays generate potentially huge data loads when exposed to multiple targets, decoys, and clutter. In each of these cases the need exists to establish correspondence between known targets and observations, to explain the remaining observations, and to provide target state estimates that evolve in agreement with the observed data.

Establishing the correspondence between an observation and a target requires a means by which to assess the quality of the assignment—a comparison between the *actual* values and the *expected* values. Predicting the expected values in an observation, on the other

[1] Apologies to Bar-Shalom and Fortmann [7].

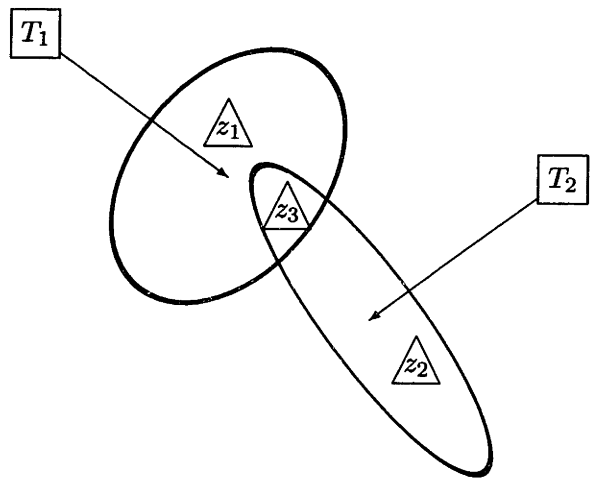


Figure 4.1: Example of multitarget tracking with clutter. State estimates at time k for targets T_1 and T_2 lead to predictions (arrows) for time $k+1$, the uncertainty of which sizes the elliptical validation gates. The existence of z_3 confuses the assignment of observations z_1 and z_2, making the simplest configuration (which associates each prediction with its nearest measurement) possibly erroneous.

hand, requires a target state estimate based on at least one previously assigned measurement. These mutually dependent tasks are difficult to perform because each observation contains both *inaccuracy* and *uncertainty*. Multitarget tracking simultaneously addresses these issues—resolving the inaccuracy in the *values* of measurements concurrently with the uncertainty in the *origins* of measurements.

Surveillance applications typically implement tracking as a discrete time recursive filter, incrementally updated each time the sensor generates a new scan. To illustrate the concept, figure 4.1 shows the time k estimates for two confirmed targets, T_1 and T_2. These estimates reflect all previously available and current information, denoted $Z^k = \{Z(1), \ldots, Z(k)\}$. At time $k+1$ a sensor reports three distinct observations, $Z(k+1) = \{z_1, z_2, z_3\}$, each with unknown origin. Predictions at time $k+1$, based on the previous state estimates, provide a basis from which to compare the observations. The uncertainty in each prediction determines the size and orientation of its corresponding elliptical validation gate. In the absence of z_3, the most natural partitioning assigns measurements z_1 and z_2 to targets T_1 and T_2. Enter z_3 and the obvious solution vanishes. Three equally likely explanations now

exist for *each* measurement: new target, confirmed target, false alarm.

Data association uncertainty transforms straightforward applications of trigonometry and linear algebra into harrowing exponentially complex algorithms. Best summarized by Bar-Shalom and Fortmann [7],

> the tracking effort for n targets can be substantially more costly than n times the effort for a single target, because establishing the correspondence between targets and observations is not a trivial matter.

Though it sounds discouraging, we adopt a multitarget tracking strategy with good reason. It offers probabilistic treatments of uncertainty that outperform earlier attempts based on implicit ad hoc models.

Surveillance applications parallel our own, We also have a sensor that scans an area and generates target observations that evolve from one scan to the next. Our situation differs by having a fixed world and a moving sensor, yet the analogy still holds. We define a target to be the points on an object that satisfy the critical alignment conditions discussed in Section 3.1.2. Sharp corners satisfy these conditions, for example, as do those points on a smooth face where the normal vector coincides with the sensor axis.

4.2 Data Association Techniques

Fortmann *et al.* [30] offer a method for tracking a known number of targets in the presence of extraneous observations. Known as the joint probabilistic data association filter (JPDAF), this approach recursively updates the estimate for each target state using the *combined innovation* of all validated measurements. To update the state estimates in figure 4.1, for example, the JPDAF calculates combined innovations for T_1 using z_1 and z_3, and for T_2 using z_2 and z_3.

The combined innovation for a set of measurements averages the individual innovations of its members, weighing each according to the association probability for the particular target being estimated. The JPDAF calculates association probabilities jointly across all known targets, not just the validated ones, hence its name. The individual innovations for each measurement result from comparing them to the hypothetical measurements whose values coincide with the predicted states (the end of the arrows).

Because it considers only the latest set of measurements $Z(k+1)$, the JPDAF lends itself particularly well to efficient implementations relative to *backscan* filters, which calculate probabilities across multiple time steps. The latter offer potentially more accurate tracking, however.

The principal drawback with this approach is that it averages the contributions from multiple measurements when updating state estimates. It never explicitly attempts to explain the origin of the observations, partly because the JPDAF lacks a method for track initiation. It is a target-oriented filter which seeks to maintain the tracks of a known number of targets, assumed to have been initialized elsewhere. The joint likelihood track formation method, from Morefield [53], offers a potential solution to this issue. This approach relaxes the condition that the number of targets is known. It partitions the data set Z^k into trajectories by maximizing the joint likelihood function of all the measurements over all of the feasible partitions.

The alternative approach that combines a Bayesian framework for calculating the probability of associations with an ability to initiate new tracks is the multiple hypothesis filter (MHT). Reid [63] pioneered multiple hypothesis testing for tracking point targets in the far field, *e.g.*, radar tracking of aircraft. This promising approach has applications in surveillance, air traffic control, collision avoidance and navigation.

The MHT algorithm has received little attention in the literature. Mori *et al.* [54] thoroughly reviewed the mathematics of the MHT, at the same time removing from it the requirement that *a priori* target distribution densities are known. Kurien [42] has addressed practical aspects of its efficiency for purposes of developing real time implementations. Finally, Cox and Leonard [20] have extended the MHT to handle the multiple target models necessary to interpret air sonar measurements in land robotics.

The MHT distinguishes itself from the previous examples by being a measurement-based approach. Rather than traverse the list of targets, considering how to update each state estimate in turn, it attempts to explain the origin of each measurement. To do so it chooses from three possible explanations: an already known target for which a state estimate exists, a previously undetected target, or a false alarm.

Perhaps its most important feature is that it defers making assignment decisions until

later iterations of the filter. Rather than rely solely on current observations to make decisions, it also looks into the *future*. In essence, it selects the most likely assignment based on how subsequent measurements reinforce certain data associations and refute others.

This approach offers the most promise for our application. Towards this end we have adapted an extended version of the MHT filter, the *model-based* MHT and applied it using our own target models and estimation techniques.

4.3 The Model-Based MHT Filter

Reid [63] originally implemented his MHT filter under the assumption that all targets share the same dynamic model, *i.e.*, each target represents a unique instance of a specific dynamic behavior. The method need not limit its application to this class of problems, however, hence its extension by Cox and Leonard [20] to the *model-based* MHT. This more general tracking filter allows for multiple behavior models of similar targets as well as completely different types of targets.

Our actual implementation results from a combined effort. Cox and Miller [50] have developed, and made available to the public, a body of source code which handles the *multiple hypothesis* portion of the MHT technique. The framework that they provide, which we will refer to as the *MHT kernel*, assumes responsibility for the generic items which are common to all applications of the model-based MHT: generating continuation hypotheses, propagating likelihoods from parent to child hypotheses, and pruning the hypothesis and track trees.

To build a specific application using the MHT, one develops the models for the behavior of targets and supplies them to the kernel. The methods developed locally for our implementation include:

- **track initiation**, which determines how to calculate the initial state estimate used to start a track from an observation.
- **measurement validation**, the technique which decides that a particular measurement either *might* or *does not* belong to a specific track;
- **track updating**, the techniques which detail how a model updates the state estimates

of a target within a single behavior, and how it generates estimates across differing behaviors;

- **likelihood calculation**, which assesses the quality of the hypothesis that a measurement belongs to a track;

Much of the complexity of the algorithm lies in the third point, which we will clarify later in this chapter.

4.3.1 Target Model Representations

Objects which meet the conditions put forth in Section 2.1.1 contain two types of geometric features: *vertices* and *faces*.[2] Vertices correspond to the sharp edges of our objects and faces correspond to the smooth surfaces between them, whether straight or curved. Some objects used in the experiments described in Chapter 5, cylinders for example, lack vertices.

Having two types of features implies we need *at least* two target behavior models. We have instead developed *three* different behaviors: one for vertices and two for faces. (Our ability to handle the latter improves with the inclusion of two distinct behaviors because a face can be planar or curved.) Though each one has a unique model h_m to predict observations $\boldsymbol{z}$, they all share the same form,

$$\boldsymbol{z}(k) = \left\{ \begin{array}{c} \alpha \\ r \end{array} \right\} = h_m[k, \boldsymbol{x}(k)] + \boldsymbol{w}(k) \qquad \boldsymbol{H}_m(k) = \frac{\partial h_m}{\partial \boldsymbol{x}}, \tag{4.1}$$

which says that the observation $\boldsymbol{z}(k)$, which consists of a bearing angle α and range r, depends on both time k and the true state of the feature $\boldsymbol{x}(k)$. (Strictly speaking, h_m depends explicitly on the *location* of the sensor $(x(k), y(k))$, which depends on time.) The vector $\boldsymbol{w}$ represents random noise which corrupts the measurements, and $\boldsymbol{H}_m$ represents the Jacobian matrix of the measurement model, evaluated anew each time the sensor moves to a new location.

[2] We borrow from geometric modeling the terms face and vertex to describe parts of an object [55]. Even though we have implemented these algorithms in 2D, we wish to establish a generic approach which readily extends to a full 3D treatment of the measurements.

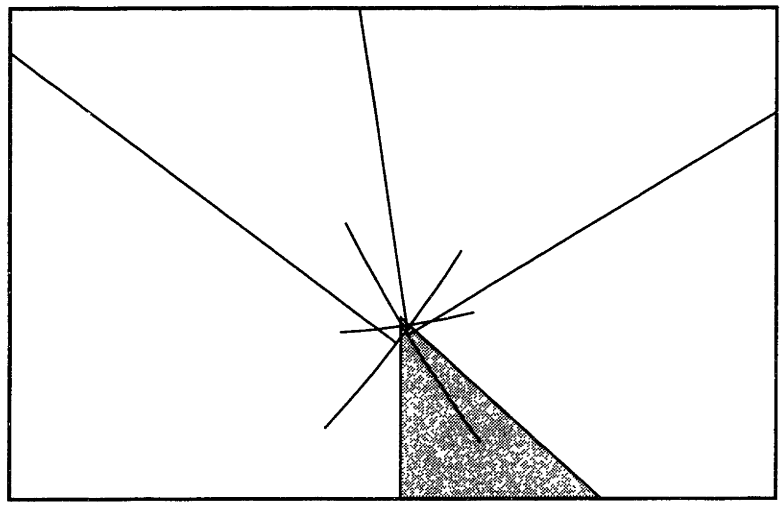


Figure 4.2: Observations of a sharp corner. As the sensor moves from one location to another, the RCDs it produces evolve such that their arcs mutually intersect at the corner. This behavior suggests that we track the observations using a vertex model that explicitly represents its location.

Vertex

As our sensor moves past the sharp corner of an object, we observe that the corresponding RCDs evolve such that their circular arcs remain in contact with the corner, as shown in figure 4.2. This behavior suggests a target model that exploits the geometric constraints of mutual intersection. The appropriate representation consistent with this constraint models the geometry of a vertex target with its Cartesian coordinates, $\boldsymbol{x} = (\xi, \eta)$, illustrated in figure 4.3. The model chosen *globally* represents the geometry of a sharp corner because it describes the state of the feature *in toto*.

To predict the bearing and range values in an observation made from the point (x, y), we apply the measurement function h_m, given by

$$h_m = \left\{ \begin{array}{c} \arctan \frac{\eta - y}{\xi - x} \\ \sqrt{(\eta - y)^2 + (\xi - x)^2} \end{array} \right\}, \tag{4.2}$$

which instantiates the general form given in equation (4.1). Note that we must exercise particular care when establishing the quadrant of the bearing angle α when predicting an

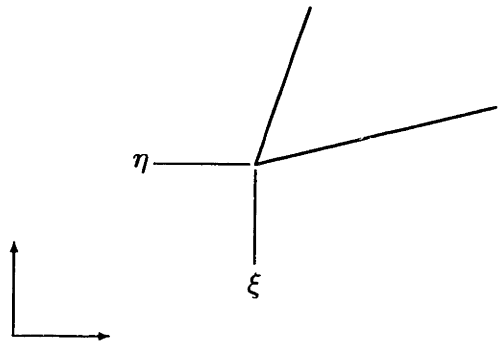


Figure 4.3: Vertex target model. A vertex representation contains its Cartesian coordinates (ξ, η).

observation.[3]

Face

Unlike the rather straightforward vertex, whose contact point with an RCD remains stationary, the contact or *ground* point of a face measurement moves as the sensor location changes. The ground point shadows the sensor, with the RCD arc remaining tangent to the face, as shown for both planar and curved faces in figure 4.4. The arguments made in Chapter 3 regarding isolated targets and sensor beamwidth indicate that we should expect this kind of behavior when observing a face. It suggests we use estimates of the *local* geometry of the face to track its observations, because the tangent to a smooth face is a *differential* property, independent of the global shape of a curved face.

The technique parallels the practice found in computational geometry of reconstructing a space curve from its derivatives using integration. We can trace out the shape of a curve exactly from a given starting point, providing we continuously know its differential properties [27] (tangent, curvature, etc.) Our approach applies this idea using approximations. We use a starting point that depends on the first observation and step along the shape of the face using the local estimates that depend on subsequent observations.

We show two target models, named *plane* and *sculpt*, that locally estimate the shape of a face in figure 4.5. The plane model, so named because it estimates the 2D tangent plane,

[3]Most computer languages offer a quadrant sensitive method. C and C++, for example, have the math library function `atan2(y,x)`.

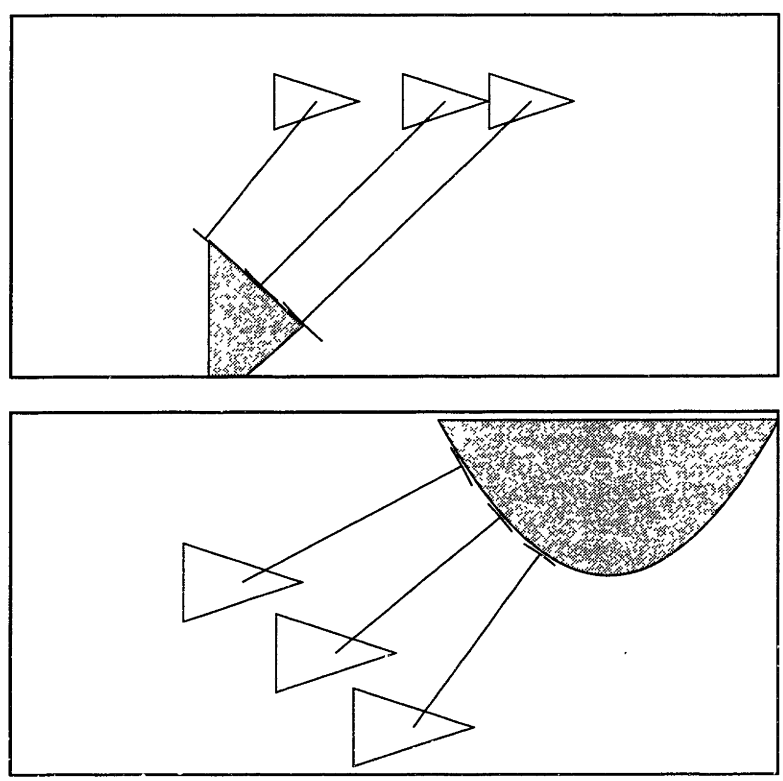
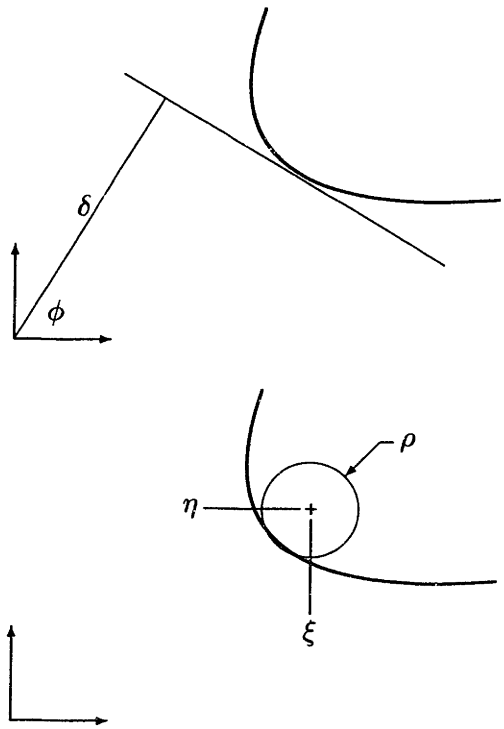


Figure 4.4: Observations of smooth faces. As the location of the sensor changes, the point of contact between an RCD and a face moves as well. It traces the boundary of the object, remaining tangent to both planar and curved faces, *top* and *bottom*.



$$h_m = \left\{ \begin{array}{c} \phi\ (+\pi) \\ |\delta - x\cos\phi - y\sin\phi| \end{array} \right\}$$

$$h_m = \left\{ \begin{array}{c} \arctan\frac{\eta-y}{\xi-x}\ (+\pi) \\ |\sqrt{(\eta-y)^2+(\xi-x)^2} - \rho| \end{array} \right\}$$

Figure 4.5: Primitive states of the face model. For a curved face, shown in bold, we provide two options: the plane and sculpt primitive states, *top* and *bottom*. An orientation angle ϕ and distance from the origin δ parametrize a plane primitive. The coordinates (ξ, η) of the center and the radius ρ parametrize a sculpt primitive. The measurement model h_m that maps a state estimate into an observation (α, r) assumes a sensor located at (x, y). We add π to the bearing angles when taking the negative branch of the absolute values to resolve the ambiguity derived from the sensor location relative to the primitive (near or far side of a plane, inside or outside of a sculpt).

contains two parameters: an angle of orientation ϕ and a distance from the origin δ. As the top half of figure 4.5 demonstrates, the tangent plane quickly deviates from the true shape of a curved face—the higher the curvature, the less accurately the plane estimate can predict the next observation for a given sensor location. As a result of this shortcoming, we also implement the sculpt model, named for its capacity to handle free-form sculpted shape. This higher order model uses a circular arc, parametrized by its center (ξ, η) and radius ρ of curvature, to estimate the local shape of a face. It predicts observations more accurately than a plane because it more closely resembles the underlying shape of a highly curved face, as demonstrated by the bottom half of figure 4.5.

4.3.2 Measurement Validation

Though not strictly necessary, measurement validation makes the tracking filter more efficient by limiting the number of association hypotheses generated. It only attempts those associations that meet a minimum condition of acceptability. In this sense, measurement validation acts as a threshold criterion that prevents the filter from wasting computational resources by considering highly unlikely data-to-target assignments.

Upon the arrival of each new scan, the filter uses the available state estimates from the previous time cycle to predict the observations that should have occurred. To perform this task, it applies the appropriate measurement function h_m for each target estimate in the existing tracks. Since an exact match occurs approximately never, the filter builds a tolerance region around the prediction to allow for disagreement. The shape and orientation of this region depend on the uncertainty in the state estimate and the average measurement uncertainty. An observation that lies within the region is considered sufficiently close to its expected value to be feasible, resulting in the generation of an association hypothesis for the target being considered.

Using the terminology and equations of a standard Kalman filter as it is described by Bar-Shalom and Fortmann [7], we now detail how to validate a particular observation $\boldsymbol{z}(k+1)$ using the target state estimate given by $\hat{\boldsymbol{x}}(k|k)$. The filter first makes a state prediction $\hat{\boldsymbol{x}}(k+1|k)$ to describe any changes that may have occurred between time k, when the estimate was calculated, and the current time $k+1$. Given the known location

of the sensor, it predicts the expected observation $\hat{\boldsymbol{z}}(k+1|k)$ by applying the measurement model to the state prediction,

$$\hat{\boldsymbol{z}}(k+1|k) = h_m[k, \hat{\boldsymbol{x}}(k+1|k)], \tag{4.3}$$

similar to equation (4.1).

The filter next needs to construct a tolerance region around the predicted value of the measurement. The shape and orientation of this region depend on the innovation covariance $\boldsymbol{S}(k+1)$, which depends in turn on three pieces of information: (1) the Jacobian matrix $\boldsymbol{H}_m$ of the measurement function; (2) the estimated covariance of the state prediction $\boldsymbol{P}(k+1|k)$ which reflects changes in the state uncertainty $\boldsymbol{P}(k|k)$ since the last measurement; and (3) the measurement covariance $\boldsymbol{R}$. Combining these terms, we find

$$\boldsymbol{S}(k+1) = \boldsymbol{H}_m \boldsymbol{P}(k+1|k) \boldsymbol{H}_m^T + \boldsymbol{R}. \tag{4.4}$$

The first term in this equation maps the uncertainty in the predicted state estimate into measurement space where it is then combined with the average measurement uncertainty.

The innovation, $\boldsymbol{\nu}(k+1)$, indicates the level of agreement between the *actual* observation and its prediction,

$$\boldsymbol{\nu}(k+1) = \boldsymbol{z}(k+1) - \hat{\boldsymbol{z}}(k+1|k), \tag{4.5}$$

specifying the *absolute* difference between the two. Normalizing the difference to establish a *statistical distance*, the inequality

$$\boldsymbol{\nu}^T \boldsymbol{S}^{-1} \boldsymbol{\nu} \leq \gamma^2 \tag{4.6}$$

establishes a *validation gate*, whose size follows from the value of the threshold parameter γ, akin to the number of standard deviations. Observations that satisfy the inequality all fall within an ellipsoid in measurement space centered on the predicted value of the measurement.

We choose a value of γ based on the desired probability that a measurement falls within its validation gate. Table 4.1, derived from the chi-squared distribution with $n = 2$, gives

γ	mass
1	0.393
2	0.865
3	0.989
4	0.9997
5	1.0
3.03	0.99

Table 4.1: Probability mass inside validation gates. The chi-squared distribution for $n = 2$ degrees of freedom gives the probability mass of a measurement falling inside a γ-sigma sized validation gate [7].

the probability mass contained inside validation gates of different sizes. This result follows from the fact that the innovation covariance, $\boldsymbol{S}$, has a chi-square distribution with number of degrees of freedom equal to the dimensionality of the measurements.

4.3.3 Generating Hypotheses

Consistent with our target behavior models, the MHT kernel generates hypotheses that enumerate all possible explanations for each of the observations in every new scan [50]. Faced with the task of performing *track continuation*, the filter offers the following types of hypotheses:

- **assignment**, if an observation validates within the track, the filter generates the association hypothesis and calculates updated state and covariance estimates. If the track corresponds to a face, we generate *two* continuation hypotheses, one using the plane behavior model and one using the sculpt behavior model.

- **skip**, because all sensors occasionally or even frequently fail to detect an existing target, the skip hypothesis provides for track continuation in the absence of new information.

- **end**, the kernel allows for the possibility of track deletion in case a target ceases to exist. We have disabled this capability for our experiments, however, by setting the likelihood of deletion to zero.

In general, the provisions for deleting tracks and for tracks that skip measurements make the filter more robust to real data, for they allow it to respond uniformly when the sensing system fails to detect an existing target. Figure 4.6 illustrates the structure in which this enumeration occurs with a block diagram that indicates the location at which track continuation hypotheses are generated with a dashed box.

An important point merits attention. Though the face model generates two track continuation hypotheses for each valid observation, our implementation of the MHT filter applies a feasibility check after calculating the state estimates. Conditions occasionally arise in the data which make the estimate calculations numerically ill-conditioned, an infinite radius sculpt for example. Should this occur, the filter removes the track continuation hypothesis from the newly generated set. We devote more attention to the specifics of this test in our discussion of the state estimate calculations in Section 4.3.5.

From the point of view of *measurement explanation*, the MHT filter generates the following hypotheses:

- **track initiation**, the filter gives each observation an opportunity to generate a *start* hypothesis;
- **track continuation**, for each track in which the measurement validates, as described above;
- **false alarm**, spurious observations sometimes occur as a result of random clutter in the environment.

The last of these becomes the default explanation when no other association emerges with sufficiently high likelihood. We choose this strategy with the philosophy that we would rather ignore a measurement than assign it to a target incorrectly.

The Start Hypothesis

As previously indicated, each observation has the opportunity to initiate a new track for both classes of geometric features: sharp corner and smooth face. We would like to have a method which initializes *all* of the necessary parameters in the appropriate target state vector and its error estimate from a *single* observation. A problem arises, however, when

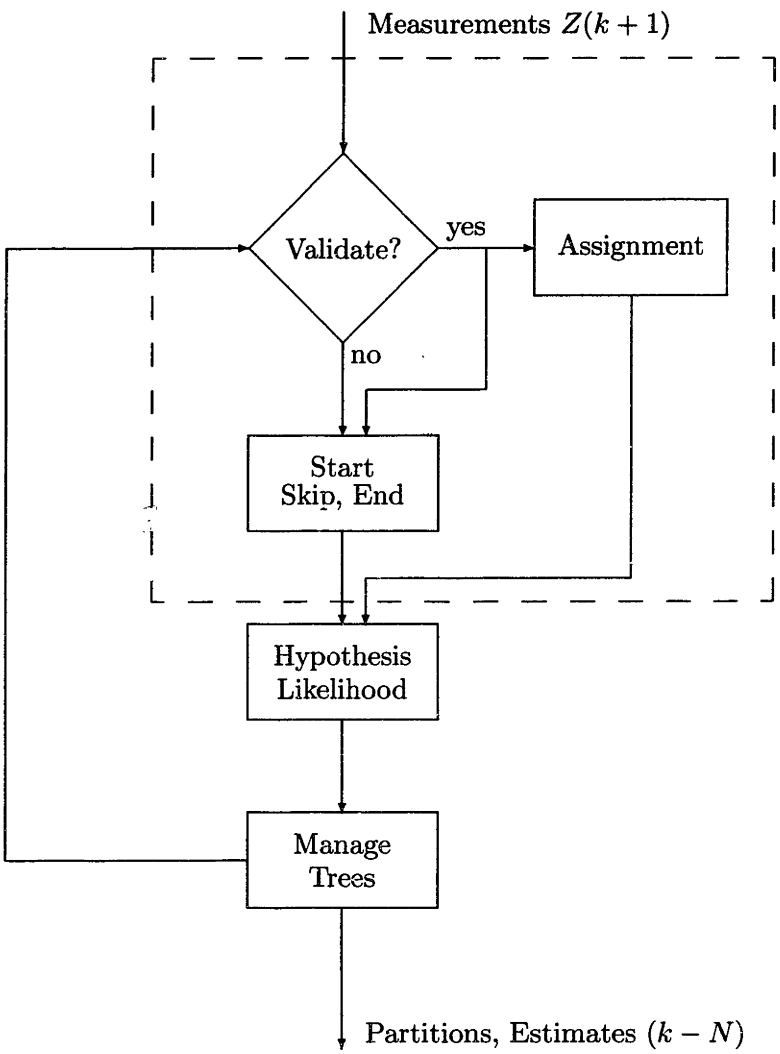


Figure 4.6: The model-based MHT filter. Each track has an opportunity to validate the observations in the current scan $Z(k+1)$. The filter generates a continue hypothesis for each match. In addition, the filter generates both false alarm and track initiation hypotheses for all observations and skip hypotheses for all existing tracks. (All child hypotheses are generated inside of the dashed box.) After calculating the likelihoods for the new set of hypotheses, it finally makes a firm decision regarding the observations N-scans back in time.

the state vector contains more degrees of freedom than can be satisfied by one measurement. Consider tracking a point target, for example, using a sensor that measures only the bearing angle. A single observation constrains the location of the point to lie along the sensor axis, but it provides no unique solution for the full state vector.

One approach addresses this issue through deferred initialization, waiting until sufficiently many measurements exist. To determine the location of a vertex, for example, requires *two* measurements from an angle-only sensor (extremely similar to determining the location of a boat using only a compass and a chart). We feel strongly against this approach because of its fundamental assumption that the subsequent measurements share the common origin of the initial one. The experimental data presented in Chapter 5 invalidates this assumption at every conceivable opportunity, hence the importance of solving the correspondence problem represents a significant focus this research.

The second common strategy assigns values to all of the state parameters using a combination of the initial observation, best guesses, and default conditions [7]. This results in very large uncertainty around the initial estimate, modeled as a large initial covariance estimate. Because this approach provides a uniform treatment of subsequent observations, it shows promise for our application.

We have implemented a hybrid technique which combines the best aspects of both of these strategies. We limit the track initiation hypotheses to only those behavior models which allow full state initialization from the available information in a single observation. The vertex and plane target models both meet this criterion, but the sculpt model does not. We only introduce the sculpt behavior as one which emerges from a previously initialized plane behavior.

One could argue that this approach repeats the second of the commonly encountered approaches discussed above, assigning a default value of zero curvature to the higher order sculpt model. The difference, however, stems from the manner in which we allow the lower order model to continue to exist. As a result, we avoid a problem typically found in using a high order model to estimate a low order behaviors—noisy and erratic behavior of the estimated parameter values.

For the vertex and plane target models, we create an initial state estimate by applying

an inverse measurement function to the initial observation $\boldsymbol{z}(1)$. Denote this inverse model by

$$\boldsymbol{x}(k) = g_m[k, \boldsymbol{z}(k) - \boldsymbol{w}(k)] \qquad \boldsymbol{G}_m = \frac{\partial g_m}{\partial \boldsymbol{z}}, \tag{4.7}$$

written to resemble the inverse h_m^{-1} of the forward measurement function given by equation (4.1). (Realize, of course, that $\boldsymbol{w}(k)$ is unknown and immeasurable.) For a sensor located at (x, y) and an observation (α, r), we have,

$$\begin{Bmatrix} \xi \\ \eta \end{Bmatrix} = \begin{Bmatrix} x + r\cos\alpha \\ y + r\sin\alpha \end{Bmatrix}, \tag{4.8}$$

$$\begin{Bmatrix} \phi \\ \delta \end{Bmatrix} = \begin{Bmatrix} \alpha \\ r - x\cos\alpha - y\sin\alpha \end{Bmatrix}, \tag{4.9}$$

for the vertex and plane. Unlike its corresponding measurement prediction h_m that sometimes needs to add π to the calculated bearing angle due to the restriction that $r > 0$, the plane model initializes unambiguously because we allow both positive and negative values for δ.

The initial target state estimate, given by,

$$\hat{\boldsymbol{x}}(1|1) = g_m[1, \boldsymbol{z}(1)], \tag{4.10}$$

relies on the zero-mean assumption for the measurement noise, $E[\boldsymbol{z} - \boldsymbol{w}] = \boldsymbol{z}$. To calculate the uncertainty $\boldsymbol{P}(1|1)$ in the initial target state estimate, we develop an idea from Smith *et al.* [71]. Expanding the inverse measurement model in a truncated Taylor series,

$$\boldsymbol{x} = g_m|_{\hat{\boldsymbol{z}}} + \boldsymbol{G}_m(\boldsymbol{z} - \hat{\boldsymbol{z}}) + \ldots. \tag{4.11}$$

The uncertainty of the estimate is its *mean squared error*, $E[\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}^T]$. The error in the estimate takes the value $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \hat{\boldsymbol{x}} = \boldsymbol{G}_m(\boldsymbol{z} - \hat{\boldsymbol{z}})$. Relying again on the assumption that the measurements are normally distributed about their expected value, the uncertainty becomes

$$\boldsymbol{P}(1|1) = \boldsymbol{G}_m \boldsymbol{R} \boldsymbol{G}_m^T, \tag{4.12}$$

in which the Jacobian G_m depends explicitly on the sensor location for the first observation.

Beyond initial state and uncertainty estimates, a track start hypothesis requires an initial likelihood of being true. One approach, quite common in the limited body of MHT literature, models new features and false alarms as either poisson or uniform distributions. It then combines these distributions with the detection probability for the sensor to arrive at the probability of the assignment event [20, 63]. The specifics of the particular distribution chosen, in general, depend on both environmental factors and the performance characteristics of the sensor system. In theory, we can measure these parameters and arrive at a reasonable value for the initial probability.

Faced with the absence of a reasonable estimate for such distributions, we instead strive for a very simple treatment of initial likelihood. We assign constant and uniform values to the likelihood of new features and false alarms such that the possible classifications are collectively exhaustive. Since we have two different types of features, we assign to each new feature hypothesis an initial likelihood of 0.33 and to the false alarm hypothesis a likelihood of 0.34. Subsequent observations of actual features increase the likelihood of the new feature tracks, whereas the likelihood of a false alarm assignment remains static. If none of the subsequent observations validate in any of the feature tracks, the false alarm assignment dominates. A false alarm should be of *slightly* higher likelihood than a new feature because we would rather discard an observation than classify it incorrectly.

4.3.4 Recursive Likelihood Calculation

In the previous section, we described how the algorithm assigns initial likelihoods, based on simple assumptions, to the hypotheses that initiate new tracks. This section describes how the algorithm recursively calculates the likelihood of child hypotheses which result from assignments of subsequent observations to those existing tracks.

The model-based MHT considers a vast multitude of possible assignment hypotheses. It does this by growing a hierarchical tree of assignments to enumerate the possible explanations for the observations made. The relentless branching in the assignment tree which occurs at each time step, however, causes exponential complexity in the algorithm. The need arises to prune the tree, in order to ensure the computational tractability of the filter.

The pruning strategies we use rely on having an estimate for likelihood of each enumerated hypothesis. Two of these strategies, K-best and minimum ratio, limit the number of associations as they are made. The third, N-scan back pruning, makes firm decisions regarding the tentative associations generated N steps earlier in time. We discuss these strategies in more detail in Section 4.3.6.

Let the notation $\theta_m(k+1)$ indicate an arbitrary assignment event for the current scan $Z(k+1)$. It associates some new measurements with established tracks, some with new targets, and some with false alarms. We construct a new association history at time $k+1$ for this particular assignment as the joint event

$$\Theta_m^{k+1} = \{\theta_m(k+1), \Theta_{l(m)}^k\}, \tag{4.13}$$

recursively defined in terms of a time k association history (or parent hypothesis), $\Theta_{l(m)}^k$. Our goal is to approximate the likelihood $P\{\Theta_m^{k+1}|Z^{k+1}\}$, conditioned on the cumulative measurement set, of this event.

In his original development of the MHT, Reid [63] gives the derivation for the *actual* event probability. Applying Bayes rule, he writes

$$\begin{aligned} P\{\Theta_m^{k+1}|Z^{k+1}\} &= \frac{1}{c} p[Z(k+1)|\theta_m(k+1), \Theta_{l(m)}^k, Z^k] P\{\theta_m(k+1)|\Theta_{l(m)}^k, Z^k\} \times \\ &\quad P\{\Theta_{l(m)}^k|Z^k\}, \end{aligned} \tag{4.14}$$

which has normalization constant c. The three terms in this calculation reflect the following: (1) the probability that the particular observations will occur conditioned on the previous assignments and earlier data; (2) the probability that the assignment event correctly associates the current observations, also conditioned on the previous assignment hypothesis and earlier data; and (3) the probability that the parent hypothesis contained the correct assignments.

Our approach parallels this derivation. For the first term we provide the statistical distance of the measurement, which assesses how well it agrees with the estimates. For the second term we provide the likelihood of what type of event has occurred, *i.e.*, skip or continue. The previous iteration of the filter makes the last term available, and the

constant c averages out, so we ignore it. If $\lambda_h(\Theta^k_{l(m)})$ represents the log-likelihood of the parent hypothesis, then for the current hypothesis Θ^{k+1}_m we find

$$\lambda_h(\Theta^{k+1}_m) = \lambda_d(k+1) + \lambda_e(k+1) + \lambda_h(\Theta^k_{l(m)}), \tag{4.15}$$

with λ_d and λ_e referring to the *distance* and *event* log-likelihoods.

The association event describes the specific means by which a track evolves, *i.e.*, with or without a skipped measurement. For its event log-likelihood, we use

$$\lambda_e = \begin{cases} \ln p_d & \text{if assignment hypothesis} \\ \ln(1-p_d) & \text{if skip} \end{cases}, \tag{4.16}$$

with p_d equal to the probability that the sensor will detect the particular feature under consideration.

The detection probability of a world feature follows from the specific characteristics of the sensing and extracting systems and the nature of the target. In theory, we can measure these parameters directly, through careful observations of sensor performance under realistic operating conditions. The interrelationships between features and the motion of the sensor make this task extremely difficult, however. In practice, we opt to treat the probability of detection similarly to the way we establish the likelihood of a start track hypothesis in Section 4.3.3. Namely, each feature has the same probability of detection, equal to 2/3. We deliberately choose a value higher than 1/2 so that the algorithm favors tracks with consecutive observations, the upper branch in equation (4.16), and disfavors ones with skipped observations.

The distance likelihood λ_d borrows from the validation gate, previously defined in equation (4.6). Recall that the gate acts as a tolerance region, centered on the expected observation, inside of which the actual measurement must occur in order for the filter to even generate an association hypothesis. The test for this condition requires calculating the statistical distance between the measurement and its expectation. Similarly for the first term in the hypothesis log-likelihood, equation (4.15), we assume that the observation has a normal distribution around its predicted value. Equivalently stated, the innovation ($\boldsymbol{\nu} = \hat{\boldsymbol{z}} - \boldsymbol{z}$)

has zero-mean Gaussian probability density function,

$$p[\boldsymbol{\nu}] = |2\pi \boldsymbol{S}|^{-\frac{1}{2}} e^{-\frac{1}{2}\boldsymbol{\nu}^T \boldsymbol{S}^{-1}\boldsymbol{\nu}}, \tag{4.17}$$

with covariance $\boldsymbol{S}$, earlier defined in equation (4.4). Reflecting such a distribution, we find for the distance log-likelihood λ_d,

$$\lambda_d = \begin{cases} -\frac{1}{2}(\ln |2\pi \boldsymbol{S}| + \boldsymbol{\nu}^T \boldsymbol{S}^{-1}\boldsymbol{\nu}) & \text{if assignment hypothesis} \\ 0 & \text{if skip} \end{cases} . \tag{4.18}$$

The zero indicates that the track continuation hypothesis that skips an observation definitely follows from its parent in the absence of any new information, *i.e.*, with a probability of one.

4.3.5 State and Covariance Estimates

A real time tracking application such as ours needs to update its parameter estimates at regular intervals because it uses them to make repeated data association decisions. A tracking filter bases these decisions on both the value of the estimate and on its degree of uncertainty. In light of these considerations, an estimator should meet the following criteria:

- **constant time**, the estimate requires the same amount of computation time for every iteration of the tracking filter. This property enables the filter to keep pace with incoming data indefinitely, continuing to produce timely results even if no upper bound exists on the total number of observations.

- **self diagnostic**, it provides a measure of its error, establishing a degree of confidence with which we accept its value. We need this information to facilitate the process of sensor fusion, in which estimates from various sources are combined such that each one is weighted in inverse proportion to its uncertainty.

The first criteria causes us to prefer recursive over batch estimators, the latter of which derive their name from the characteristic processing of the entire data set with the arrival of each new measurement. The ever increasing calculation time inherent to this approach

eventually chokes any real time application.

The second criteria proves equally important. Having a measure of error accompanying each estimate enables the sensible combination of multiple estimates. We expect these estimates, nominally from different sources, to contradict each other. It will prove useful, however, to calculate a composite estimate in which each of the individual components makes a contribution weighted according to its degree of confidence.

Vertex Target Model

We successfully estimated vertex target states using a first order extended Kalman filter (EKF). This method is fully recursive, with the current estimates of state and mean-squared-error providing *sufficient statistics* on the past history of the state.

One normally applies an EKF when estimating parameters of a dynamic system, usually written as $\boldsymbol{x}(k+1) = f[k, \boldsymbol{x}(k)] + \boldsymbol{v}(k)$, in which $\boldsymbol{v}$ represents *plant noise.* A vertex feature belonging to a fixed target has static behavior in the normal sense. As we have argued, however, dynamic behavior emerges when we move the sensor from one location to another (recall figure 4.2 on page 79). Though the true state $\boldsymbol{x}$ of the vertex remains constant, each observation we make of it reflects the current location of the sensor.

Given estimates at time k of both the state $\hat{\boldsymbol{x}}(k|k)$ and the error $\boldsymbol{P}(k|k)$ we make a prediction of the new state and its covariance,

$$\hat{\boldsymbol{x}}(k+1|k) = \hat{\boldsymbol{x}}(k|k), \tag{4.19}$$

$$\boldsymbol{P}(k+1|k) = \boldsymbol{P}(k|k). \tag{4.20}$$

In a properly dynamic system, these predictions reflect changes undergone by the true state $\boldsymbol{x}$ since our last measurement. The static vertex assumption, however, renders this prediction trivially simple. The filter uses the state prediction in conjunction with the known sensor location to predict the value of the expected measurement $\hat{\boldsymbol{z}}$. We have already shown how it performs this step with equation (4.2).

Recall the innovation $\boldsymbol{\nu}$ and its covariance $\boldsymbol{S}$, the same as those used in calculating the validation gate, tell us how well the actual observation agrees with its expectation. The

EKF updates the state estimate by averaging the predicted state with the new measurement. The filter gain tells us how to weigh the contributions from these different sources,

$$\boldsymbol{W}(k+1) = \boldsymbol{P}(k+1|k)\boldsymbol{H}^T\boldsymbol{S}^{-1}(k+1), \tag{4.21}$$

by using the Jacobian of the measurement model $\boldsymbol{H}$ to map the uncertainty uncertainty $\boldsymbol{P}$ (in the predicted state estimate) into measurement space and then normalizing it with the innovation covariance $\boldsymbol{S}$. The update step applies the filter gain in the following manner:

$$\hat{\boldsymbol{x}}(k+1|k+1) = \hat{\boldsymbol{x}}(k+1|k) + \boldsymbol{W}\boldsymbol{\nu}, \tag{4.22}$$

$$\boldsymbol{P}(k+1|k+1) = \boldsymbol{P}(k+1|k) - \boldsymbol{W}\boldsymbol{S}\boldsymbol{W}^T. \tag{4.23}$$

Note that the estimated state uncertainty $\boldsymbol{P}$ actually *decreases* because new information has arrived. Also illustrating this trend, figure 4.7 presents a sequence of estimates for a vertex target that corresponds to the corner of a triangle. The state uncertainty estimate, shown as an error ellipse scaled by a factor of 8, not only shrinks as more information arrives but also becomes more circular. The initial oblateness results from the large uncertainty in the observed bearing angle of an RCD feature.

Face

Recall in Section 4.3.1 that the vertex model provides a *global* representation for the geometry of a sharp corner. The EKF works well for tracking these targets because the recursively calculated estimates it provides reflect information found in *all* of the measurements. Anticipate the failure of the EKF in tracking the plane and sculpt target models that correspond to smooth faces, however, because both of these models provide only *local* estimates of the underlying geometry.

The alternative approach we have implemented uses *nonlinear least-squares* estimation. Since we are unaware of any recursive formulations for this approach, we only apply the estimator to a moving data window that comprises the n most recent observations. Implemented in this manner, the normally batch processing approach meets the criterion established earlier in this section that its estimates be calculated in constant time. This

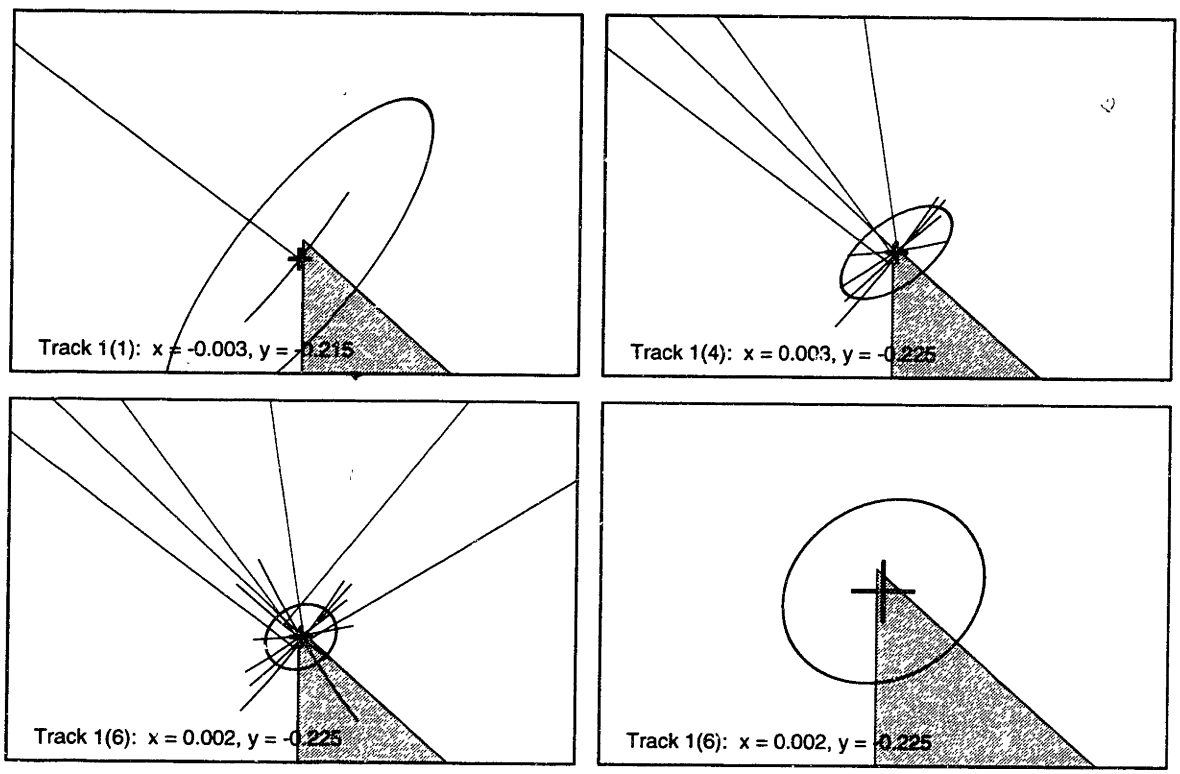


Figure 4.7: Evolution of a vertex estimate using the EKF. The 2 cm "+" in each frame defines the scale. Estimated states $\hat{\boldsymbol{x}}$ are drawn with an 8σ error ellipse (determined by the covariance matrix $\boldsymbol{P}$). *Top left*, we initialize the EKF using the time-of-flight inversion method. *Top right*, after four observations have arrived, the EKF gives a much better estimate of the vertex location than the initial estimate. *Bottom left*, after six observations, the error ellipse shrinks and becomes less oblate. *Bottom right*, a closer look reveals the accuracy of the EKF estimate.

particular method appeals to us for the following reasons:

- target state vectors, with insufficient capacity to approximate the global shape of the face, respond only to the most recent observations;
- in the absence of a true recursive formulation, operating on a small batch of recent data represents the best chance at achieving real time performance.

The least-squares estimate $\hat{\boldsymbol{x}}^{\mathrm{LS}}(k)$, given by

$$\hat{\boldsymbol{x}}^{\mathrm{LS}}(k) = \arg\min_{\boldsymbol{x}} \sum_{j=0}^{n} \{\boldsymbol{z}(k-j) - h_m[k-j, \boldsymbol{x}]\}^T \boldsymbol{R}^{-1} \{\boldsymbol{z}(k-j) - h_m[k-j, \boldsymbol{x}]\}, \qquad (4.24)$$

individually weighs the contribution from the bearing and range components of an RCD, according to the inverse measurement covariance $\boldsymbol{R}$. Using the same n measurements, we calculate $\boldsymbol{P}(k|k)$,

$$\boldsymbol{P}(k|k) = \frac{1}{n} \sum_{j=0}^{n} \boldsymbol{H}_m(k-j) \boldsymbol{R} \boldsymbol{H}_m^T(k-j), \qquad (4.25)$$

the approximate mean-squared-error of this estimate.

The optimal size of the batch depends on both the quality and quantity of the measurements. The number we choose for n represents a compromise between two conflicting ideals. On the one hand, the state estimate will respond very quickly, keeping up with the sensor, when we process very few measurements. On the other hand, it will respond equally well to the noise in the measurements, thus a larger batch enhances the stability of the state estimate. The difference between the orientation angle ϕ of the plane target model and the bearing angle of the most recent RCD is an excellent indicator of this effect, as is evident in figure 4.8. For a size $n = 4$ batch, the orientation angle of the plane lags the most recent RCD by 4.3°, but by only 2.2° for a batch of size $n = 2$. We have compromised by using a batch size of $n = 3$ for the results shown in Chapter 5.

The final concern in calculating the state estimates is their feasibility. Situations arise when the data contrive to ill-condition the least-squares calculation in equation (4.24). Consecutive RCDs which share the same bearing, for example, generate a sculpt state estimate with infinite radius. Aside from the usual problems that arise when one tries to calculate an infinite value, an infinite radius sculpt is indistinguishable from a plane. At the

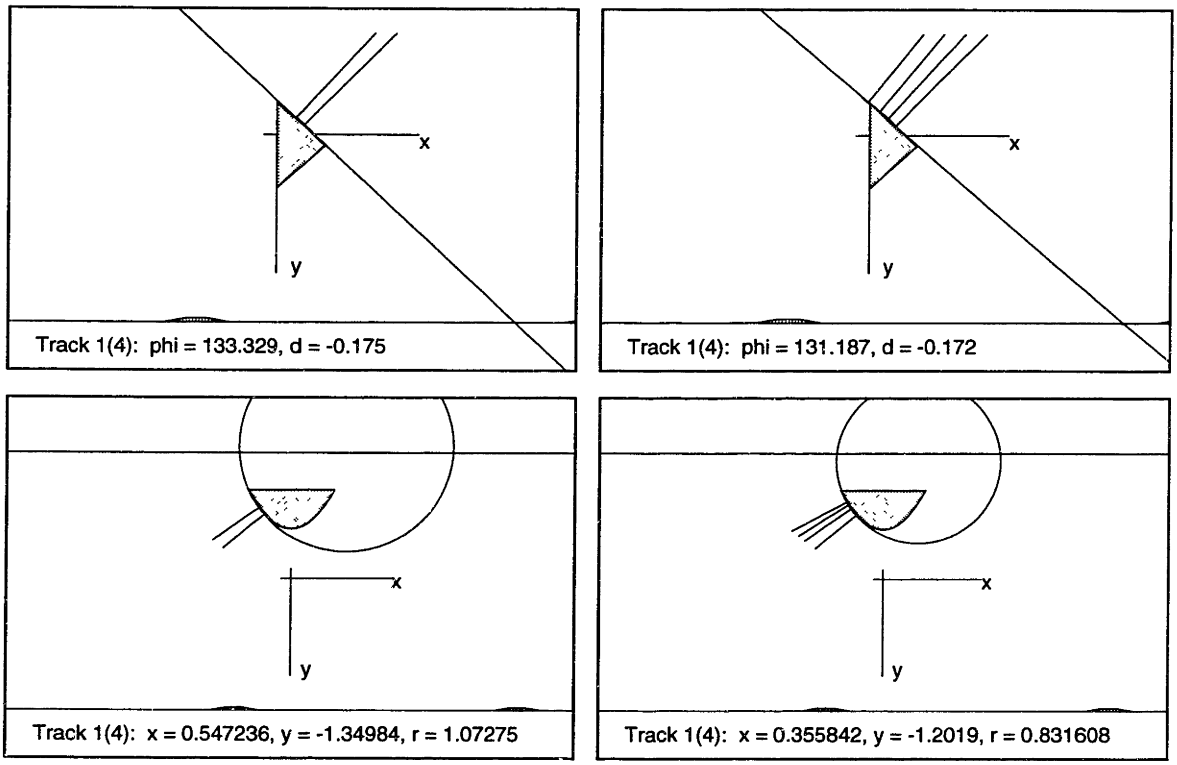


Figure 4.8: Face estimates using nonlinear least-squares. The bylines in these figures indicate that 4 observations in total have validated in track 1, although only the ones used to calculate the actual estimate are shown, followed by its target state estimate, (ϕ, δ) for the plane and (ξ, η, ρ) for the sculpt. *Top row*, in both frames the most recent RCD to the right has a bearing angle of 135.5°. The smaller batch with only 2 RCDs, *left*, yields an orientation angle for the plane of $\phi = 133.3°$, whereas the larger batch of 4 RCDs, *right*, yields $\phi = 131.2°$. The angle of the former estimate is closer to the most recent measurement than the angle of the latter estimate because the larger batch size implies greater inertia. *Bottom row*, though more difficult to ascertain, the estimates of sculpt parameters experience a similar delay. The most recent measurement is to the right.

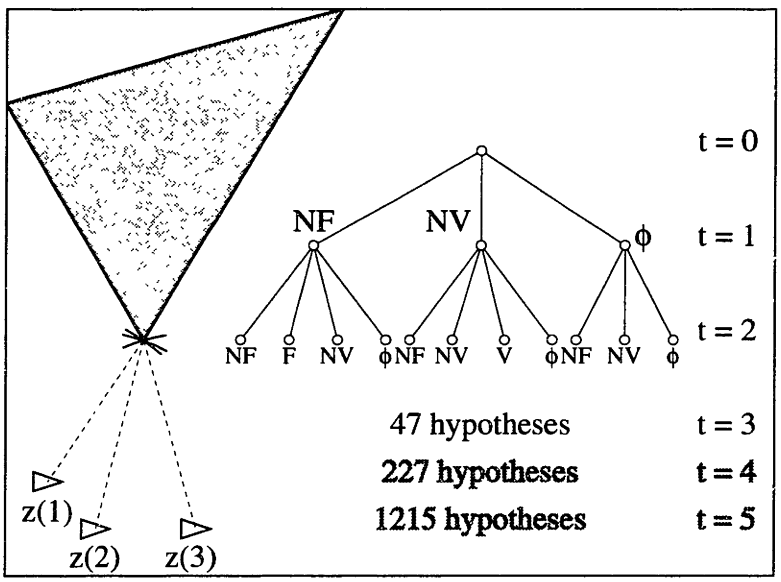


Figure 4.9: An example of very simple branching. Each scan in this sequence contains *only one* RCD, and the face model only allows the plane behavior. Imagine the complexity when the typical scan contains 5–8 RCDs and we allow both plane and sculpt face model behaviors. Even worse still, the structure shown here repeats itself for *all* existing tracks as well.

other end of the spectrum, neither the algorithm nor its author can differentiate between a sculpt with zero radius and a vertex. The vertex target model with the EKF works much better than the small batch under these circumstances. To address the feasibility issue, we set upper and lower bounds on the radius parameter ρ of the sculpt target estimate. Though the filter performs the actual calculation of equation (4.24) using an unconstrained method [34, 59], the results must pass through the established bounds. Those estimates that fail this simple test never generate hypotheses.

4.3.6 Tree Management

The manner in which the MHT filter generates hypotheses causes its computational complexity to grow exponentially, a result of branching. Figure 4.9 shows an example of the branching that results from a very simple case in which each scan contains only one observation and we only allow two different target models, vertex and plane. A realistic example generates far more hypotheses. Our implementation and the MHT kernel together combine

several strategies to limit and manage the growth of the hypothesis and track trees. We take advantage of two different approaches to hypothesis management: screening and pruning. The former, applied prior to the actual generation of hypotheses, includes validation gates and feasibility checks. We have already discussed the implementation of both of these.

Pruning, on the other hand, acts after the hypotheses have been generated. The MHT kernel employs three pruning strategies:

- **K-best**, which limits the assignment hypotheses actually placed on the tree to the K best;
- **minimum ratio**, which limits hypotheses to the ones whose likelihood ratio vis-a-vis the most likely hypothesis exceeds a minimum threshold r_m;
- **N-scan back**, which removes all branches from the tree, save the most likely one, at the node that corresponds to N steps back in time (see figure 4.10).

A superficial glance suggests that the first two of these strategies offer little benefit at high cost. It would appear that they require enumerating all possible assignment hypotheses and ranking them.

Techniques exist to rank assignments without exhaustive enumeration [11, 58]. Cox and Miller [21] incorporate Murty's algorithm [57] in the MHT kernel upon which our implementation is built. This approach, from combinatorial optimization, ranks assignments by cost in polynomial time. By limiting the number of hypotheses generated at the current time step, K-best and minimum ratio pruning limit the *branching* as well. As a result, their effectiveness increases drastically at subsequent time steps.

The minimum likelihood ratio r_m and the maximum number of hypotheses K can potentially trivialize each other. In other words, too small a value for K prevents the kernel from enumerating hypotheses of likelihood anywhere near the minimum ratio. Conversely, too large a value for r_m limits the number of hypotheses so effectively that K is rendered redundant. When balanced, however, they mutually support each other. Experience tells us that setting K to $1/r_m$ achieves balanced results. These values affect the computational throughput of the MHT filter, an issue we address in Section 5.3. The results in Chapter 5 reflect values of $K = 100$ and $r_m = 0.01$.

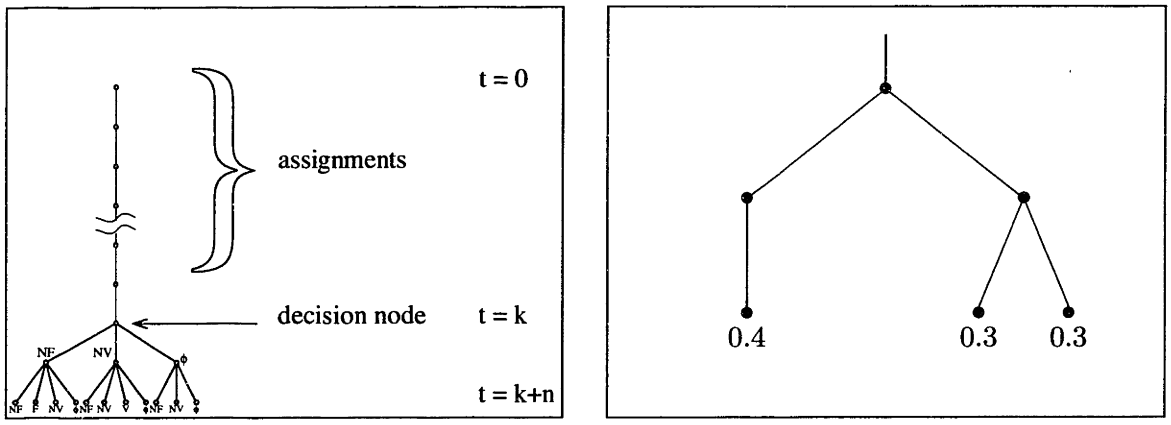


Figure 4.10: An application of N-scan back pruning. *Left*, the MHT prunes the hypothesis tree N-scans prior to the current time step, making a firm decision for the assignments above the decision node. *Right*, a pathological tree for N-scan back pruning in which the filter chooses the right branch despite the best individual hypothesis being on the left one.

Only the hypotheses which escape the first two strategies actually survive to be placed on the tree. There they remain, spawning progeny of their own until N-scans later, when we apply the final strategy. In N-scan back pruning, shown in figure 4.10, the MHT kernel looks at the decision node that corresponds to N time steps prior to the current time. To determine the likelihood of each branch at this node, the MHT adds the individual likelihood contributions of its leaves into a composite likelihood for the entire branch. The most likely branch survives, with the filter pruning away the others. Above the decision node remains a simple list of assignments.

We would like to champion the infallibility of N-scan back pruning, yet reality prohibits us from doing so. The method described for determining the most likely *branch* offers no guarantee of selecting the branch which contains the most likely *leaf*. Figure 4.10 shows a contrived example, selected from a class of pathological trees, for which N-scan back pruning fails.

4.4 Examples

In this section, we present two time sequences which contain snapshots of the decisions made by the MHT filter as data arrives. The first example contains a sequence of scans which correspond to a sensor moving around a triangular prism. These scans are a subset

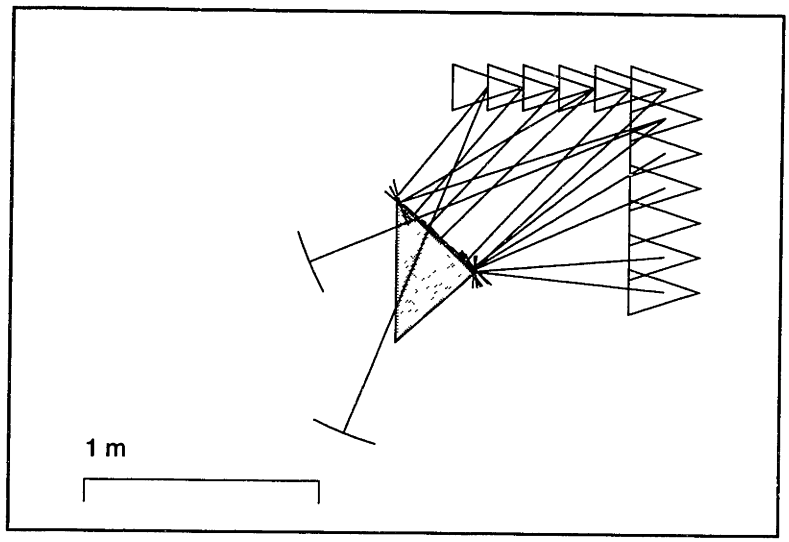


Figure 4.11: Data set for first example. The 12 scan time sequence indicated by the unfilled triangles begins at the top left and proceeds clockwise around the triangular object. Most of the 15 RCDs in these scans correspond to the two sharp corners of the triangular object and the face between them. Two RCDs clutter the data, having no obvious explanation.

of the cumulative data from the experiment described in Section 5.2.1. Figure 4.11 presents an overview of the setup, with the sensor locations and the extracted RCDs overlaid upon the actual geometry of the triangle. For sake of discussion, let the top left sensor location correspond to discrete time step $k = 1$; the sensor moves clockwise with each scan incrementing the clock by one. The data set for this example contains twelve scans with a total of 15 RCDs.

Figure 4.12 shows how the MHT filter assigns each of the 15 RCDs in the cumulative data set. For this example, the filter pruning parameters were set to $K = 100$-best hypotheses, minimum ratio $r_m = 0.01$, and $N = 5$-scan back pruning. The least-squares estimates for the face were calculated with a maximum batch size of three, as shown in the fifth, sixth, seventh, and ninth frames.

One point of note in figure 4.12 is the unexpected presence of a sculpt target state in both the third and the fifth frames. Even though the underlying face of the triangle is perfectly flat, the sculpt target behavior model offers a better explanation for the observations in these two frames, though it certainly gives a very low estimate for the curvature. The

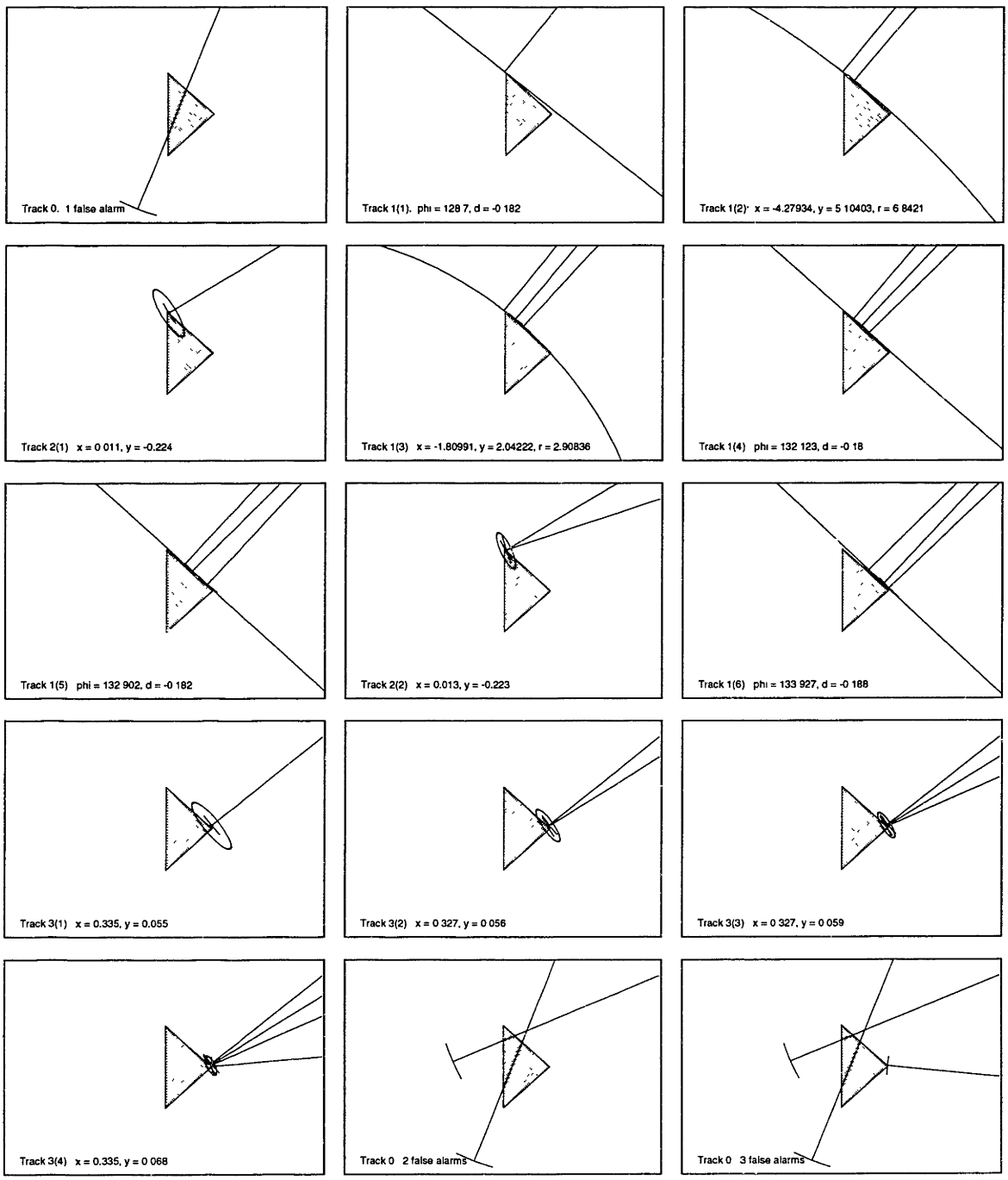


Figure 4.12: MHT time sequence for first example. Recall the bylines signify the track ID number with its number of assigned RCDs, followed by the current state estimate where appropriate. (Apologies for the size of the text, but the partitioning is more important than the numbers). *Left to right, top to bottom*, this figure shows each decision made by the MHT filter with respect to the origin of observations of figure 4.11. The third and fifth frames show that the sculpt primitive state better models the behavior of the observations based on the likelihoods, even though the underlying geometry is flat. In the last frame, the observation is classified as a false alarm because no subsequent measurements exist to support alternate hypotheses.

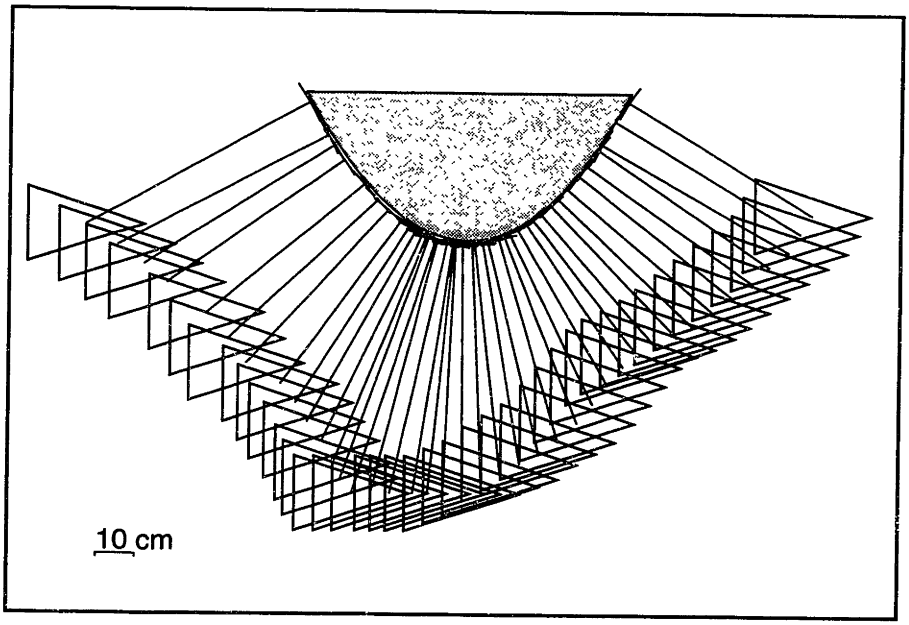


Figure 4.13: Data set for second example. The 40 scan time sequence begins at the left and proceeds to the right. Although each scan contains a single RCD, the MHT filter still branches significantly because it tries to start a plane target and a vertex target with each new observation.

likelihoods that result from the actual data drive these decisions. In the sixth and seventh frames, the plane target model resumes being the most likely alternative.

Another observation we make regarding this example, the filter assigns the very last observation that occurs to the false alarm track (fifteenth frame) even though it quite clearly corresponds to one of the sharp corners. The false alarms in the first and fourteenth frames come as no surprise because they have no physical correspondence in the environment, as indicated in figure 4.11. The MHT filter classifies the last observation as a false alarm because it lacks subsequent measurements that support any alternate data association hypotheses of higher likelihood.

For the second example, we apply the MHT filter to a set of observations corresponding to the smoothly curved face of a parabolic cylinder. (The data set comes from the third experiment that we describe more thoroughly in Section 5.2.3.) The 40 scans in this example contain exactly 40 RCDs, one per scan, as shown in figure 4.13. The MHT algorithm nevertheless continues to experience high complexity because it still branches at each step. It *always* attempts to initiate new tracks using both the plane and vertex target models and to split face tracks into plane and sculpt targets.

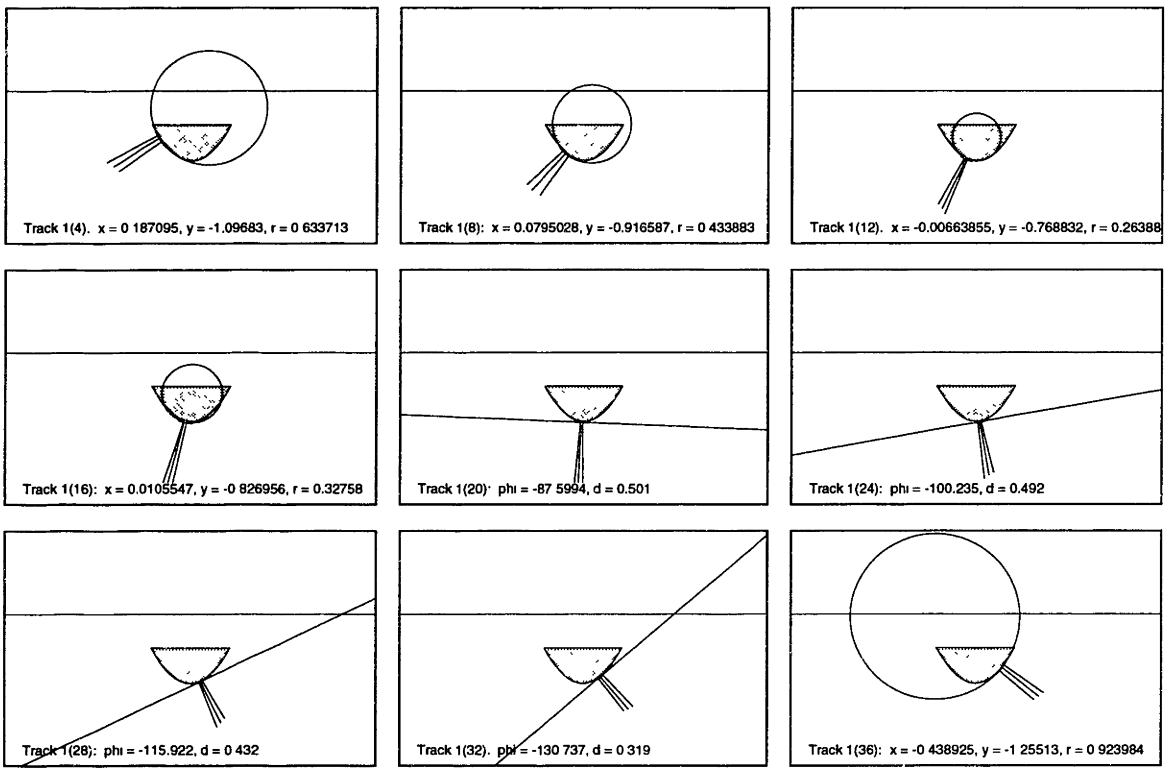


Figure 4.14: Time history of curved face track. Every fourth estimate in the sequence shows that the algorithm switches between the sculpt target model (shown as a circle whose parameters (ξ, η, ρ) are indicated in the bylines) and the plane target model (straight lines with ϕ and δ) when tracking a curved face.

This example leads to the important result that the MHT filter successfully explains the entire set of observations using a single face track, a subset of which is shown in figure 4.14. Thought there are 40 time steps in all for this track, the nine frames (every fourth step in the range 4–36) in figure 4.14 sufficiently demonstrate the ability of the MHT filter to track the face without interruption by any incorrectly interpreted vertex targets.

The first three frames behave exactly as expected—a sculpt target model whose curvature estimate increases as it evolves to match the increasing curvature of the underlying geometry. We see an unexpected behavior in the fifth through eighth frames, when a plane primitive state best explains the observations. The opposite has already occurred once before, when a sculpt primitive best explained some of the observations of a flat face in figure 4.12 of the previous example, but it comes as a surprise nevertheless. We attribute this unexpected result to measurement error in the RCDs.

4.5 Summary

This chapter has described the model-based multiple hypothesis algorithm for performing the difficult task of segmenting sonar observations. The two applications of the approach to real data illustrate the success of the technique. The parabolic face example in particular demonstrates the novelty of our contribution to environment modeling—extending the MHT approach to include the ability to track smoothly curved faces for the first time.

Our next step is to synthesize the global geometry of the faces from fully partitioned observations and display the overall result using complete data sets from the sequence of experiments performed throughout this research.

Chapter 5

Results

In this chapter, we describe the *global* representation for the geometry of a face, in contrast with the *local* geometry estimates used by the tracking filter to explain the observations. The global representation synthesizes the correctly partitioned measurements into an approximation of the shape of an entire smooth face.

Following geometry synthesis, we describe the experiments performed to validate the algorithm on real underwater data generated by the commercially available Tritech ST1000 profiling sonar. The first experiment, which contains a triangular prism, demonstrates a capability underwater that is similar to the one shown on land by an earlier implementation of the model-based MHT filter. The incorrect results from our second experiment using an elliptical object leads to an analysis of the relative curvature of the shape being tracked. We apply this understanding to a parabolic prism in the third experiment, which demonstrates a successful contribution to the field of model-based multiple hypothesis tracking.

5.1 Global Face Geometry Synthesis

The tracking filter described in the previous chapter relies on local estimates of the geometry of a smooth face to partition the measurements correctly. As a result, the algorithm solves the correspondence problem more efficiently than would be possible using global geometry estimates because of their increased number of degrees of freedom. At the tracking stage, maintaining estimates of high order models would be wasteful and the filter would suffer

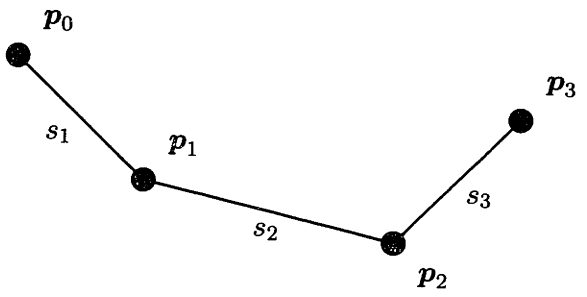


Figure 5.1: Linked rod representation for a curved face. Borrowing the terminology used by Koenderink [40], we represent a smooth face with a piecewise linear model—a set of linked rods. Unlike the original, however, the length of each rod varies individually. We parametrize the curve with chord length, as shown.

from reduced throughput. Given correct partitioning, however, we can now synthesize the global geometry of a face using all of its corresponding measurements rather than a small window of the most recent ones.

Borrowing terminology from Koenderink [40], we model the global geometry of a smooth face using a *linked rod* representation. In the original, each rod has identical length. We deviate from this scheme by allowing their lengths to vary individually. The computer implementation of this model simply maintains a list of points, $\{\boldsymbol{p}_i\}$, $i = 0, \ldots, n$, and a corresponding list of rod lengths $\{s_j\}$, $j = 1, \ldots, n$, as shown in figure 5.1. We parameterize the curve using chord length because it approximates the ideal arc length parametrization as the rods get shorter.

To find any point $\boldsymbol{p}(s)$ on the curve, we must first determine on which span it falls,

$$i_s = \arg\max_i \left(s > \sum_{j=1}^{i} s_j \right), \tag{5.1}$$

and how far down that span to travel,

$$\hat{s} = \frac{s - s_{i_s}}{s_{i_s+1} - s_{i_s}}. \tag{5.2}$$

All that remains is a simple linear interpolation between the appropriate endpoints,

$$\boldsymbol{p}(s) = \begin{cases} (1-\hat{s})\,\boldsymbol{p}_{i_s} + \hat{s}\,\boldsymbol{p}_{i_s+1} & \text{if } 0 \le s \le \sum_j^n s_j \\ 0 & \text{otherwise} \end{cases} \tag{5.3}$$

which provides completeness by allowing any real value of s.

Initialization of this model begins with its first whole link, and thus two measurements are required. This condition never causes difficulty for us, because any track which contains only a single measurement ultimately receives a false alarm label from the multiple hypothesis filter. All actual face tracks contain *at least* two observations.

The zeroth endpoint $\boldsymbol{p}_0$ comes directly from the initial observation associated with the track, using the simple line-of-sight interpretation of the observation. If (x, y, α, r) represents the sensor location, bearing, and range values, we have

$$\boldsymbol{p}_0 = \begin{Bmatrix} x + r\cos\alpha \\ y + r\sin\alpha \end{Bmatrix}. \tag{5.4}$$

To incrementally refine the model, we calculate the location of subsequent points using the same equation but we use an *modified bearing angle.* The tracking filter helps in making this adjustment.

Recall from Chapter 3 that we suggested that the observed bearing angle of an RCD can be inaccurate in contrast with the observed range. We know the range fairly well because the ST1000 measures the time of flight with microsecond resolution and the target strength only affects the range by a few millimeters. We do not know the bearing angle nearly as well because of the approximately 8° effective beamwidth of the sensor. Some flexibility exists with respect to its value.

When the MHT filter proposes an assignment hypothesis for a particular observation it updates the corresponding target state estimate. We can use this local approximation to improve the bearing angle of the most recent measurement. A *modified* RCD contains the ideal bearing angle that would result if that local estimate were indeed the true underlying geometry, as shown in figure 5.2. To determine the improved angle, we simply apply the appropriate measurement function h_m for either sculpt or plane target models to predict a

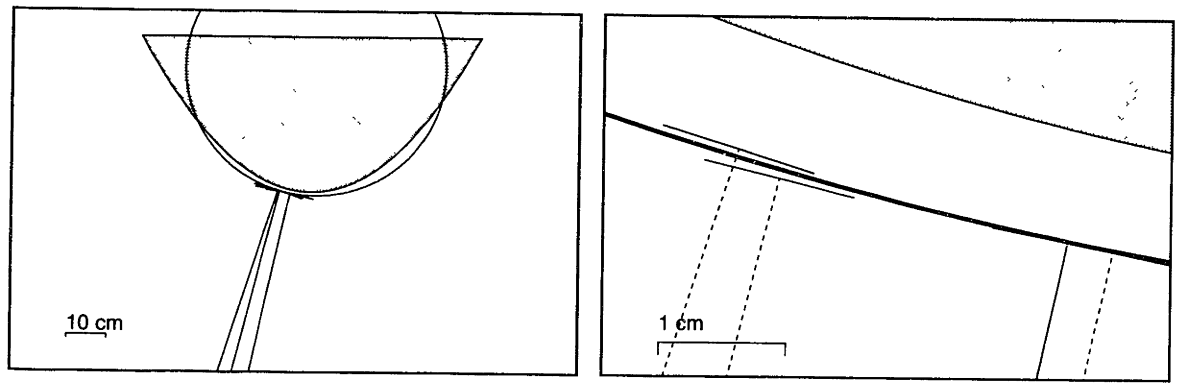


Figure 5.2: The improved bearing angle of a *modified* RCD. *Left*, these three RCDs are used in the least-squares calculation of the sculpt estimate. *Right*, an extreme closeup view shows the original RCDs bearing angles with dotted lines and the improved bearing angle of the *most recent* RCD with a solid line. The latter is calculated by predicting an expected observation based on the current state estimate.

bearing based on the current state estimate. The MHT filter reports this modified bearing angle to the geometry synthesis stage in place of the original RCD bearing angle determined by the feature detector.

As it is described, this method of updating the model allows loops to occur. Macro-scale loops, those which correspond to a large circular target for example, raise no objections. The model *should* allow loops of this size to occur. We would like to prevent micro-scale loops from forming as new points are added, however. At what scale do we make the cutoff? Alas we leave this issue open because no simple answer exists for this question.

5.2 The Experiments

Just accusations often fly that academic research churns out overly theoretical algorithms with little application in realistic situations. We ultimately intend for this research and its offshoots to find use in AUV mapping and navigation, as discussed in Chapter 1. The engineering philosophy we advocate to achieve this end requires experimental validation of the algorithm. Simulation has perhaps played an important role in its development. Before applying the stamp of approval, however, proper validation of the approach mandates that we apply it to real data. Experience has shown that too many things can go wrong in the real world, or even the laboratory, for us to accept a method without proper challenge.

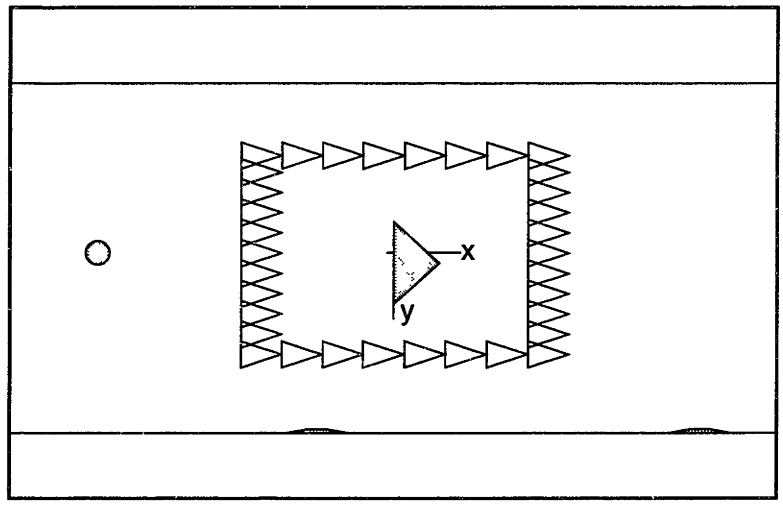


Figure 5.3: Overview of the triangular prism experiment. A circular cylinder located at (-220, 0) cm with radius 8.9 cm and a triangular prism with vertices at (0, 38.6), (0, -23.2), and (34, 7.8) cm from the origin, both made from aluminum, stand on end in the tank. Some of the sensor locations have been omitted for clarity. The axes are 0.55 m long.

We have conducted a series of experiments in a fresh water tank using a commercially available Tritech ST1000 high frequency (1.25 MHz) profiling sonar [75], a photograph of which is shown in figure 2.2 on page 27. The tank is 2.64 m wide and approximately 30 m long, with water 1.2 m deep. It has one concrete that is rough relative to the 1.2 mm wavelength of the ST1000. The water temperature for all three of the experiments ranged from 16–24 °C, corresponding to sound speeds in the range 1469–1493 m/s.

5.2.1 Triangular Prism Experiment

The first experiment seeks to answer the question of whether the algorithm can, using the ST1000 sensor underwater, achieve similar results as reported earlier by Cox and Leonard [20] for a two dimensional office environment. The experiment consisted of moving the sensor along a closed trajectory around the perimeter of the object with step size of approximately 15 cm, as figure 5.3 demonstrates. The triangular prism (shown in a photograph on page 143) contains geometric features equivalent to those of the earlier work: flat faces and sharp corners.

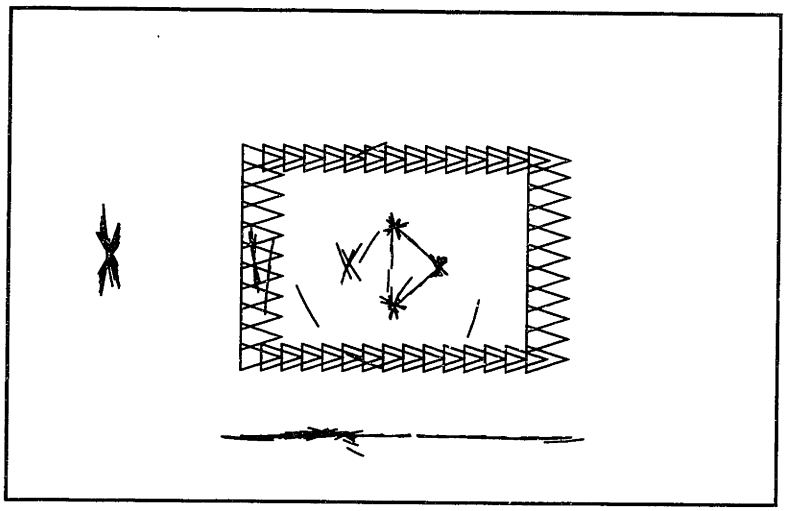


Figure 5.4: Cumulative data set for triangular prism experiment. The raw scans include returns which originate from the rear wall of the tank. Since this boundary is rough, thus not consistent with the assumptions made in Chapter 2, we omit its corresponding RCDs from the data seen by the MHT filter.

Constructed by folding 3.17 mm thick aluminum sheets ($\lambda = 5.3$ mm in aluminum), the triangle has open ends and thus floods with water. A thin layer of aluminum fully immersed in water has a high reflection coefficient, thus the flat faces strongly reflect acoustic energy back at the sensor, consistent with the discussion of target strength in Section 3.2. The sharp corners, in contrast, are acoustically weak targets.

The feature detector parameters of threshold $\tau = 1$ cm and minimum RCD width $w = 3.6°$ result in extraction of 221 RCDs from a total of fifty panoramic scans, yielding an average of 4.4 RCDs per scan. The RCDs used as input to the MHT filter are shown in figure 5.4. Since the rear wall is rough, violating the assumptions made in Chapter 2, we deliberately omit its corresponding observations from the cumulative data set processed by the tracking filter. In Section 5.4.1, however, we investigate the effects of including these RCDs in the data set.

The MHT filter (with minimum ratio $r_m = 0.01$, $K = 100$-best hypotheses, $N = 5$-scan back pruning, and face batch size of $n = 3$) correctly partitions the measurements of this data set into three vertex targets and three smooth faces. Using the linked rod representation for smooth face geometry introduced in Section 5.1, we present the final

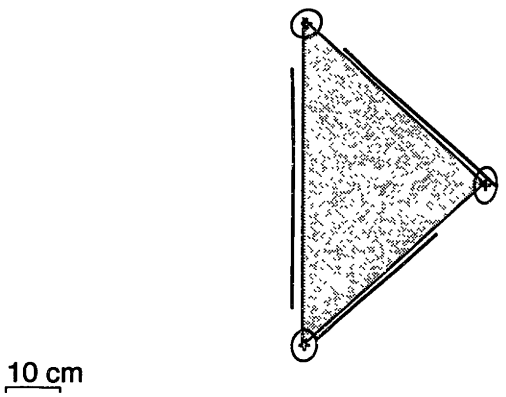


Figure 5.5: Results from the triangular prism experiment. There are three faces and three vertices, the latter drawn with their corresponding 8σ error ellipses.

results for the triangle in figure 5.5. Because the faces contain the strongest sonar returns, the ST1000 detects their arrival earlier than those from the vertices. As a result, the faces are all slightly *outside* of the object while the vertices are slightly *inside*. The three faces appear to be single segments, yet each contains at least two spans. Vertices are shown accompanied by their corresponding uncertainty ellipses, whose size and shape are determined by $\boldsymbol{P}$, scaled by a factor of eight.

The largest deviation between the location of a link and the underlying face is 2 cm, corresponding to the lower end of the vertical face in figure 5.5. Table 5.1 compares the location of the vertices found by the algorithm to the actual corners of the underlying object. As indicated, the maximum error for the three vertices is 10.8 mm.

Vertex Error Summary			
	$\hat{\boldsymbol{x}}$	$\boldsymbol{x}$	error
1	(0.004, 0.385)	(0 , 0.386)	4.12 mm
2	(0.006, -0.223)	(0 , -0.232)	10.8 mm
3	(0.342, 0.082)	(0.34, 0.078)	4.47 mm

Table 5.1: Summary of results for first experiment

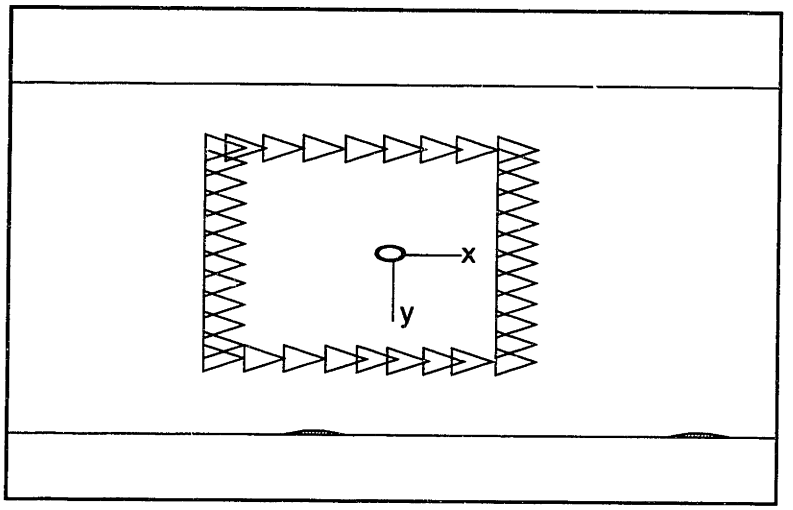


Figure 5.6: Overview of the ellipse experiment. We placed a section of sailboat mast at a position of $(-2.5, -1.2)$ cm from the origin. The mast is aluminum with varying thickness of 4.5–5.4 mm. It has an elliptical cross section with major and minor radii of 10.2 and 5.7 cm. The unfilled triangles represent sensor locations, some of which have been omitted for clarity. The axes are 0.55 m long.

5.2.2 Ellipse Experiment

Having established that the algorithm works well for polygonal geometry, we wish to investigate its performance with curved sections. In the second experiment, we placed a section of mast from a sailboat on end in the tank. As in the first experiment with the triangle, a closed rectangular path was made around the object, shown in figure 5.6. The spacing between sensor locations was again approximately 15 cm. The mast itself is made of aluminum with wall thickness that varies between 4.5–5.4 mm. It has an elliptical cross section, with major radius of 10.2 cm and minor radius of 5.7 cm. In 53 panoramic scans, the feature detector ($\tau = 1$ cm and $w = 3.6°$) found a total of 271 RCDs, for an average of 5.1 RCDs per scan. As in the previous experiment, however, we omit the ones from the rough wall in the rear of the tank.

The MHT filter divides these RCDs into groups which correspond to individual vertices distributed around the perimeter of the mast, as shown in figure 5.7. Why has the filter failed to identify any faces in this data set?

To answer the question, we must realize that the performance of the MHT filter in

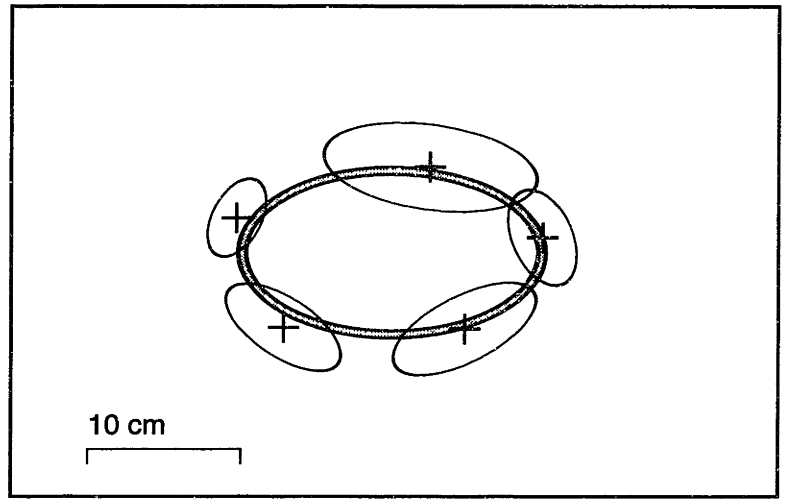


Figure 5.7: Results from the ellipse experiment: the MHT filter chooses the most likely explanation that the RCDs correspond to a sequence of disjoint vertices rather than a smooth face; why did it fail?

identifying and tracking a face depends on how distant the sensor is relative to the curvature of the face. As powerful as the MHT is at assigning measurements to features in the world, its ability to resolve details depends directly on the proximity of the sensor to the objects being scanned. The parameter that expresses this relationship we call the *characteristic range*, ρ_c, shown in figure 5.8. Defined as the ratio of the actual range from the sensor to the smooth face to the radius of curvature (evaluated at the contact point), this value indicates the distance between sensor and target in the appropriate length scale of the target. The characteristic range inversely correlates to the degree the MHT filter can discern details.

The filter easily handles the two limiting cases, $\rho_c = 0$ and $\rho_c = \infty$. The former arises when a nominally curved face locally resembles a flat one. The MHT filter tracks these well, as it should. The latter case arises when the tracking a vertex, and we have already shown that the MHT filter tracks these as well. It also arises when the sensor is very distant from the object. In this situation, we should expect a reduction in performance because of the difficulty in resolving fine details from far away. Do we really care if the algorithm fails to correctly classify very distant features?

The difficulty occurs "somewhere in the middle," which precisely coincides with the

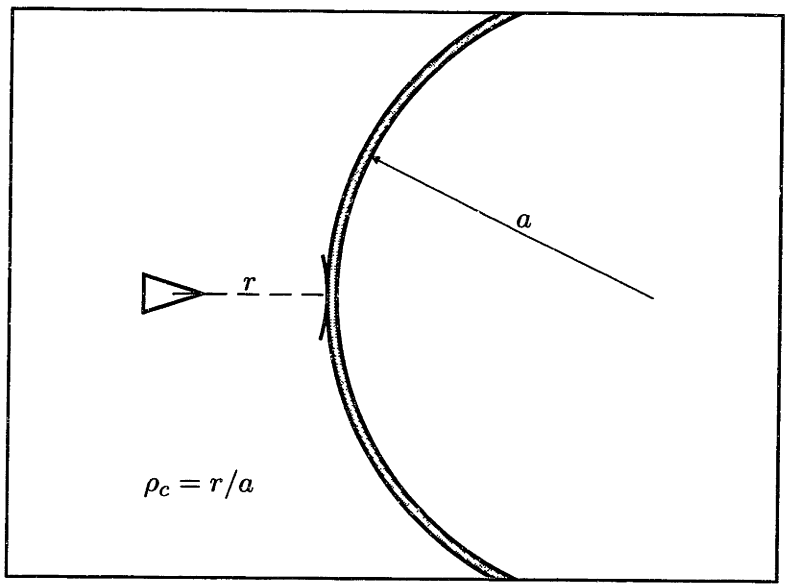


Figure 5.8: Definition of *characteristic range*. The quality of the assignments made by the MHT filter degrades when the sensor distance is large relative to the local radius of curvature of the face. The characteristic range quantifies this distance in a length scale appropriate to the shape of the object.

elliptical mast. The broad side of the mast has a 17.9 cm radius of curvature, which implies that the ST1000 sensor, with its minimum useful range of approximately 0.6 m, will never approach closer than $\rho_c = 3.3$ to the *broad* side of the mast. During this experiment, however, we placed the sensor at least than 0.75 m from the object, giving an actual value of $\rho_c = 4.2$ in the best case. The characteristic range exceeds 20 when looking at the narrow end from a distance of 0.75 m.

We should expect the quality of the assignment decisions made by the filter to degrade as the sensor is placed further away. By simulating the data that would result from generating scans along closed trajectories around elliptical objects of various curvatures, we verified this expectation using similar parameters as the actual experiment. The top half of figure 5.9 shows the number of individual tracks that the model-based MHT filter produces as the value of ρ_c increases—the correct answer is always one. The bottom half of figure 5.9 shows the percentage of tracks that are of type vertex. This distinction is worth making because sometimes the filter breaks off a face track in a region of high curvature only to initiate another face track. Though strictly speaking this behavior is wrong, we award partial credit

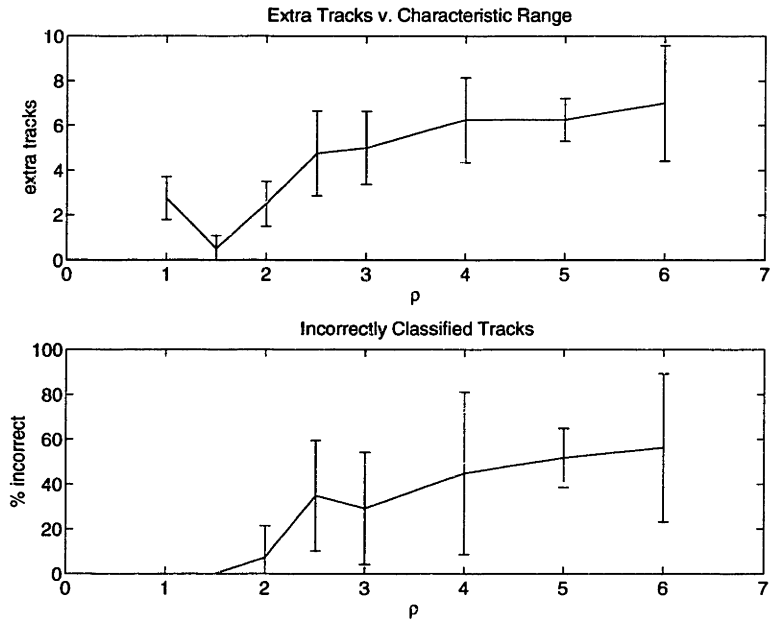


Figure 5.9: Adverse effects due to high relative curvature: the effect that sensing from different characteristic ranges has on the algorithm.

to the algorithm for for identifying faces only. Summarizing this experiment, we find that the algorithm breaks down gradually as the relative sensor distance increases.

5.2.3 Parabola Experiment

The second experiment taught us the importance of the characteristic range and how it influences the ability of the algorithm to reconstruct the shape of curved faces correctly. The third experiment incorporates the lesson learned with respect to this parameter. We maintain the effort of reconstructing curved geometry, but this time we select an object which allows a lower value for the characteristic range than was possible using the elliptical mast in the previous experiment.

The object consists of a thin sheet of aluminum (0.7 mm thick) bent around a wooden frame and glued to a pair of carefully shaped profile supports. The resulting parabolic shape has a focal length of 114 mm. The minimum radius of curvature of the object, located at the turning point of the parabola, is 22.8 cm. This value allows the ST1000 to close in

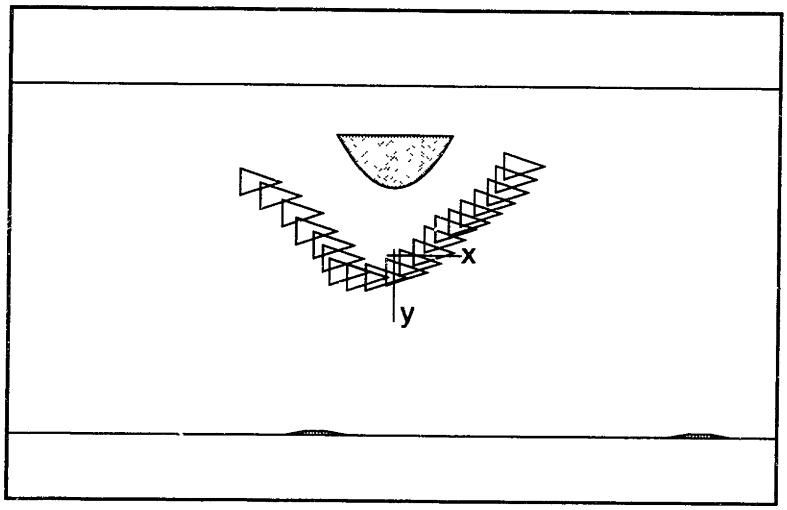


Figure 5.10: Overview of the parabola experiment. We placed a parabolic cylinder, made of 0.7 mm thick rolled aluminum, on end in the tank. The parabola has a focal length of 114 mm and symmetric endpoints located at $(\mp 42.6, -90.6)$ cm. (Cf. figure 4.13 on page 104 for the RCDs that correspond to just the parabolic face). The unfilled triangles represent sensor locations, some of which have been omitted for clarity and the axes are 0.55 m long.

to a characteristic range of $\rho_c = 3.25$, as determined using a sensor distance of 0.75 m. Section 3.2 predicts a high acoustic target strength for this target.

For this experiment, only the curved face of the object interests us. We therefore choose a sensor trajectory that approximately parallels the bent aluminum sheet, as figure 5.10 demonstrates. The RCDs which correspond to this feature alone are shown in figure 4.13 on page 104. They represent a subset of the cumulative data fed into the MHT filter, which contains observations that correspond to the internal framework of the object as well.

Some of the RCDs from the interior of the object result from its internal corners. We have already demonstrated successful results that highlight the ability the MHT filter has in explaining such features. Other RCDs, however, result from the rough surfaces of the support members. These observations violate the assumptions made in Chapter 2. Our only available option, unfortunately, is to manually cull them from the cumulative data set. Reluctant to intervene in this manner, we accept whatever results the MHT generates, offering an explanation where necessary.

With all other parameters the same as the previous two experiments, figure 5.11 presents

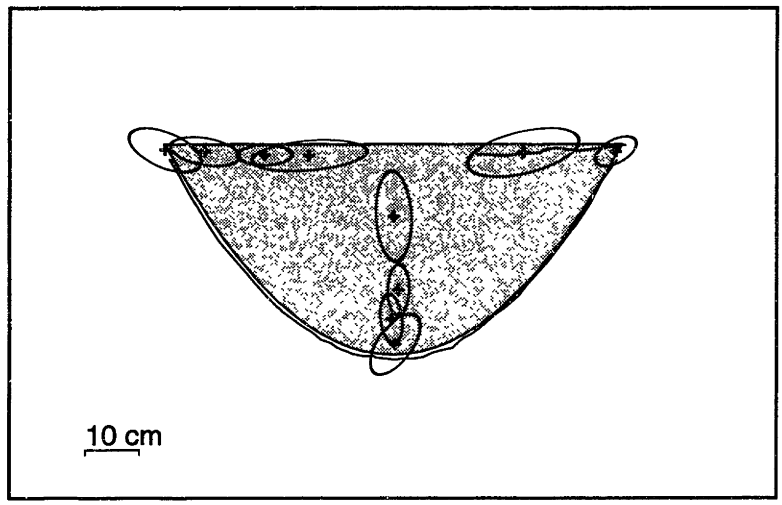


Figure 5.11: Results from the parabola experiment. The MHT filter successfully identifies the two corners and the single uninterrupted face that joins them. The algorithm partially reconstructs the rear face of the object, using measurements that pass through the front face. Some of the vertices correspond correctly to internal corners in the support frame. The remaining ones result from small bubbles in the epoxy coating that protects the internal plywood from water damage.

the final geometry of this experiment. As shown, the MHT filter reconstructs a portion of the rear face of the object, using measurements that pass through the front face. The vertices in the interior of the object belong to the internal structure, a framework made of epoxy coated plywood. Small bubbles in the epoxy act as point resonators, hence the many vertex targets identified by the algorithm.

The most important result of this experiment is the continuous track of the curved face of the parabola. We show the final geometry of this track accompanied by the observations that generate in figure 5.12. The successful partitioning of these RCDs demonstrates our contribution to environment modeling using the model-based MHT approach.

5.3 Efficiency of the MHT Filter

Experimental observations have shown that the MHT filter makes better quality assignments when allowed to generate more hypotheses. This trend exists only up to a point, however, past which the assignments do not change. The number of available options it ponders

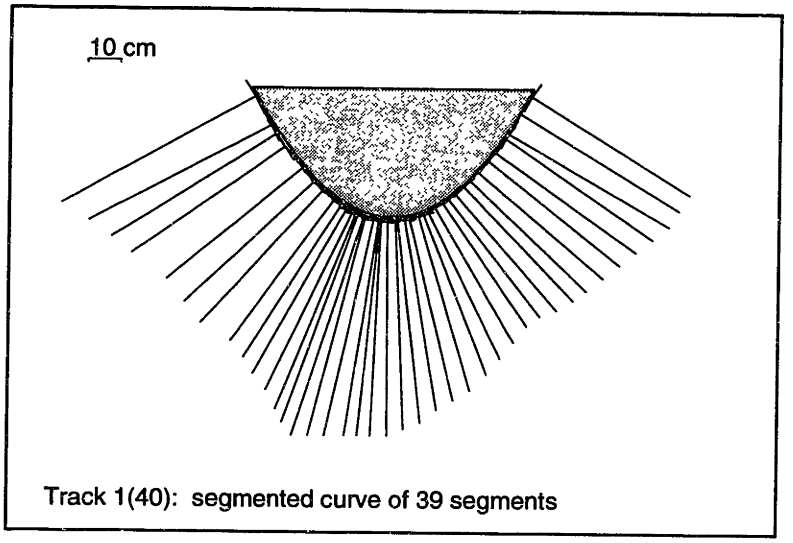


Figure 5.12: The smooth face track of the parabola. This figure gives a close up view of the successful partitioning of observations corresponding to a curved face. It demonstrates our extension of the model-based MHT approach to environment modeling.

affects the speed at which it makes the assignments, severely limiting its overall efficiency when too many are allowed.

For any real time tracking application, the speed at which decisions are made strongly influences the ultimate utility of the algorithm. In this section, our goal is to understand how the filter pruning parameters of minimum ratio r_m, K-best hypotheses, and N-scan back influence the *time* it takes for a 50 MHz RISC workstation to run the MHT filter on complete sets of real sonar observations.

The data for this investigation comes from the first and third experiments, discussed in Sections 5.2.1 and 5.2.3. For the first example, the cumulative data set has 168 RCDs, distributed over 50 scans. The average of 3.4 RCDs per scan represents a moderate level of branching. The second example contains 166 RCDs in 42 complete scans, yielding an average branching factor of 3.9 RCDs per scan, also moderate.

The results shown in table 5.2 summarize the performance of the kernel as the pruning parameters vary. The *user* time gives the important figure of merit, directly indicating the computational efficiency of the algorithm. The low times in the middle of the parameter

Summary of CPU Time for Algorithm						
N-scan back	minimum likelihood ratio r_m (K-best $= 1/r_m$)					
	0.05	0.02	0.01	0.005	0.002	0.001
First experiment (Section 5.2.1)						
2	4.3u 0.2s	4.3u 0.1s	4.7u 0.1s	4.9u 0.1s	5.3u 0.1s	6.5u 0.2s
3	11.4u 0.2s	15.6u 0.2s	16.6u 0.2s	17.7u 0.1s	19.4u 0.2s	15.2u 0.3s
5	7.3u 0.1s	9.4u 0.1s	10.5u 0.1s	11.7u 0.1s	15.7u 0.2s	19.5u 0.2s
7	8.8u 0.1s	12.7u 0.1s	16.0u 0.1s	20.8u 0.2s	29.7u 0.3s	35.7u 0.5s
10	9.7u 0.2s	14.2u 0.1s	18.8u 0.2s	25.0u 0.2s	37.2u 0.4s	52.5u 0.7s
13	10.9u 0.1s	19.7u 0.1s	27.4u 0.2s	37.9u 0.2s	57.0u 0.4s	79.7u 0.9s
16	12.8u 0.1s	19.9u 0.1s	30.9u 0.2s	44.3u 0.4s	61.1u 0.8s	84.9u 1.2s
Third experiment (Section 5.2.3)						
5	11.7u 0.2s	14.3u 0.1s	17.0u 0.1s	19.7u 0.1s	23.7u 0.1s	27.1u 0.2s

Table 5.2: CPU time to run the MHT filter. The `time` utility of the UNIX™ C-shell documents the efficiency of the algorithm when run on a 50 MHz RISC workstation. The *user* time (u) indicates seconds spent in calculation and the *system* time (s) indicates seconds spent performing system overhead tasks such as memory allocation and i/o. The boxes indicate the values that represent a good compromise between accuracy and efficiency, as empirical observations have shown.

space are quite encouraging. For example, the first experiment has 50 discrete time steps, yet it only requires from 10–20 seconds to process the entire data set. Applying these figures leads to an average cyclical rate of 2.5–5 Hz.

In the first two lines of table 5.2, with $N = 2$ and $N = 3$-scan back pruning, the filter makes some erroneous assignments of observations. The limited depth of the decision tree prevents the filter from finding the correct solution to the data correspondence problem. Practical experience has shown that the values indicated by the boxed regions in table 5.2 represent a good compromise between accuracy in assignments and efficiency in calculation.

As the depth of the tree increases, and the resulting number of hypotheses, so does the *system* or overhead time required to process all the additional options. Painstaking attention to the details of careful memory management, taken by the author of the kernel [50], have paid off, for the system time remains essentially constant except for the unusually complex cases, almost never encountered in practice.

5.4 Relaxing the Assumptions

Giving credit where due, Reid [63] has pioneered a very powerful approach to solving the correspondence problem. Of the available techniques, it offers the most robust performance in interpreting data of questionable quality. Inevitably, however, its ultimate ability rests upon how well the target behavior models that we provide adequately capture the nature of the observations.

The conditions restricting the nature of the environment put forth in Chapter 2 form the basis for the tracking models developed in Chapter 4. How well do these behavior models fare when the assumptions fail to adequately characterize the environment? Are they robust to adversity? By exploring deliberately beyond the limits within which the algorithm works well, we gain deeper insights into understanding all aspects of data interpretation than we would by remaining in the realm of safety. Our goal in this section is to consider how the algorithm handles adversity of three types: rough surfaces, uncertain sensor location, and increased spacing between scans.

5.4.1 Rough Surfaces

We must face the realization that the smooth surface condition effectively limits our theater of operations to rigidly structured environments. Natural settings, in contrast, contain many rough boundaries under interrogation by high frequency sonar. However, we must emphasize that the degree of roughness follows from a direct comparison of the average variation in surface elevation relative to the acoustic wavelength. Increasing the wavelength, as a result of reducing the ability to discern fine detail, allows us to share an observation offered by Koenderink [40]

> **1st convenience:** *real-life tolerances make everything smooth.*

Our approach remains perfectly valid under such consideration, although we make it partially with tongue in cheek.

The phenomenon of specular reflection forms the basis for the tracking models developed for use in the data association filter. Strictly speaking, insonified vertices reradiate echoes formed by *diffraction*, rather than reflection. The specular model adequately captures this

behavior, however. Recall our discussion of isolated targets in Section 3.1.1, which produce returns only when a strict alignment condition is met. The sensor axis must nearly coincide with the surface normal vector.

A rough surface, of course, behaves differently—as a *distributed* target. One treatment, which results from the application of Huygens principle, models a rough surface as a collection of point scatterers [73]. Under this assumption, we expect the outcome to be a disjoint set of vertices distributed along the surface boundary.

As is evident in figure 5.13, the expected outcome does indeed result when we apply the algorithm to the rear wall of the tank. We do witness the existence of a face track, coincident with the wall, which contains all of the RCDs whose bearing angle is of near normal alignment. More importantly however, we find that the remaining tracks are all vertices, uniformly distributed along the length of the target. In Chapter 6, we propose several different possibilities for dealing with these issues.

5.4.2 Uncertain Sensor Motion

When presenting the overview of our approach, we assumed precise knowledge of the location of the viewpoint in Section 2.1.1. Our critics may characterize this assumption as completely naive, but they do so unfairly. We realize the difficulties met in obtaining position information, still an unsolved problem in robotics research. The observed trend in the literature, however, indicates that navigational precision will certainly improve in the future [38].

Modern acoustic navigation systems, equipped with transponder nets, deliver remarkable accuracy and repeatability [51].[1] Such a system requires an initial survey, however, during which the transponders are deployed in the region of interest. We envision that acoustic navigation technology will eventually exist in a self calibrating package suitable for an AUV to deploy on the fly.

Until an accurate navigation system of this kind exists (ideally in an inexpensive commercially available package), we must cope with the effects of imprecise knowledge with

[1] From a nominal range of 100 m, the EXACT navigation system provides position reports at a rate of 2 Hz, accurate to within a 1 cm cube with repeatability of 3 mm [48].

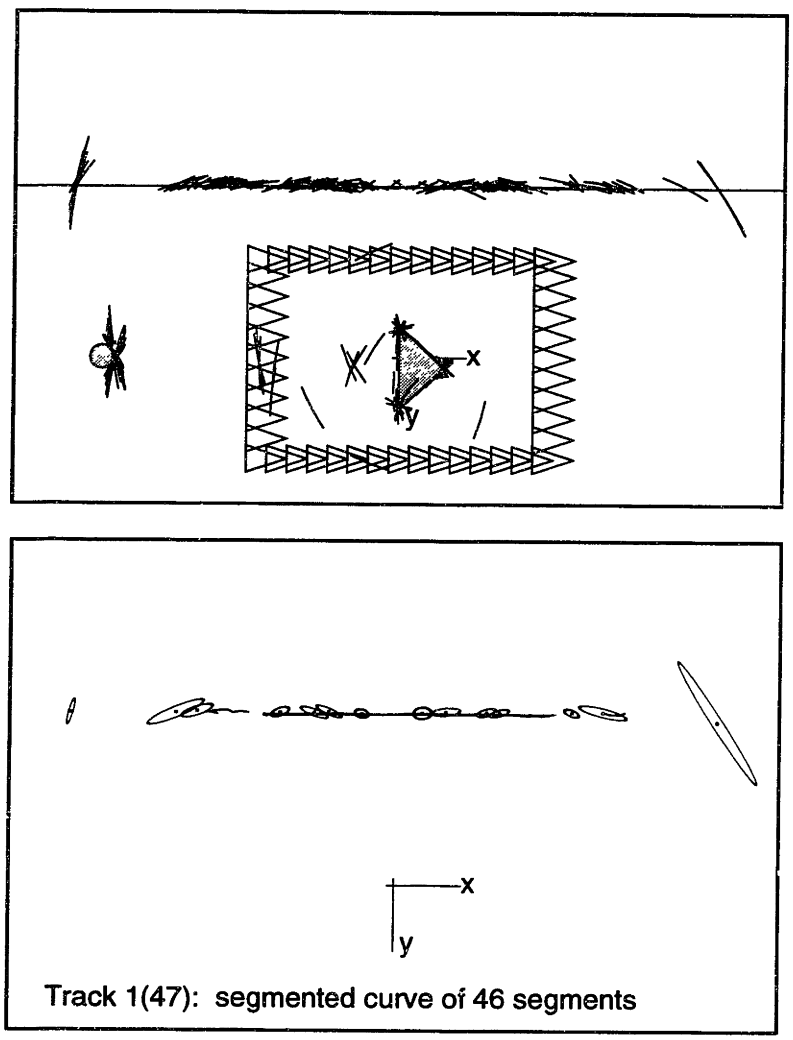


Figure 5.13: Algorithm applied to rough wall. *Top*, the cumulative observation set consists only of the RCDs belonging to the rear wall of the tank. *Bottom*, the algorithm generates a face with 46 segments and many vertices, as predicted by Huygens principle [73].

respect to the sensor location. An investigation of the effects of position uncertainty on our algorithm addresses this issue. We define a time varying position error $\epsilon(t)$ to be a zero-mean Gaussian random variable with distribution

$$p[\epsilon(t)] = N[0, \sigma_0 + t\sigma_1]. \tag{5.5}$$

Given judicious choices of σ_0 and σ_1, this model offers two different ways in which the position uncertainty can behave:

- **unbounded**, ($\sigma_1 > 0$) found in systems whose error grows with time due to a nonzero drift rate, e.g., inertial navigation, dead reckoning;
- **bounded**, ($\sigma_1 = 0$) characteristic of systems which determine position relative to known external references, e.g., LORAN, GPS, acoustic nets, geophysical [1, 76].

For the remainder of this discussion, we chose $\sigma_0 = 1$ cm and $\sigma_1 = 0.02$, with the belief that they adequately characterize position error with a reasonable value.[2]

To demonstrate how imprecise knowledge of the sensor location affects the state estimates alone, we have applied random noise to the sensor positions for the measurements used in figure 4.7 on page 96. The imprecise results shown in figure 5.14 result from inaccurate *values* of the measurements, but they still have certain *origins*. Because of the consistent partitioning of the observations, the target state estimates that result deviate little from the earlier ones, posing an insignificant effect. If this were indeed the only consequence of position uncertainty, we would still be able to build a model of the underwater environment with the same level of accuracy.

Real difficulty arises when the position uncertainty causes erroneous partitioning of the measurements, which is highly likely. For a full run on the data in the third experiment, with noisy measurements shown in figure 5.15, the MHT fails to keep the longer tracks alive. It ends up launching new ones due to the very low likelihood that the later measurements belong to the track. For this case, the MHT identified a total of 19 tracks, with the ones

[2]Under normal circumstances in navigation, σ_1 has units of velocity and is known as a *drift rate*. Since our "stop and go" method of operation lacks a direct measure of time, we specify a dimensionless σ_1, which produces an error proportional to cumulative distance traveled by the sensor.

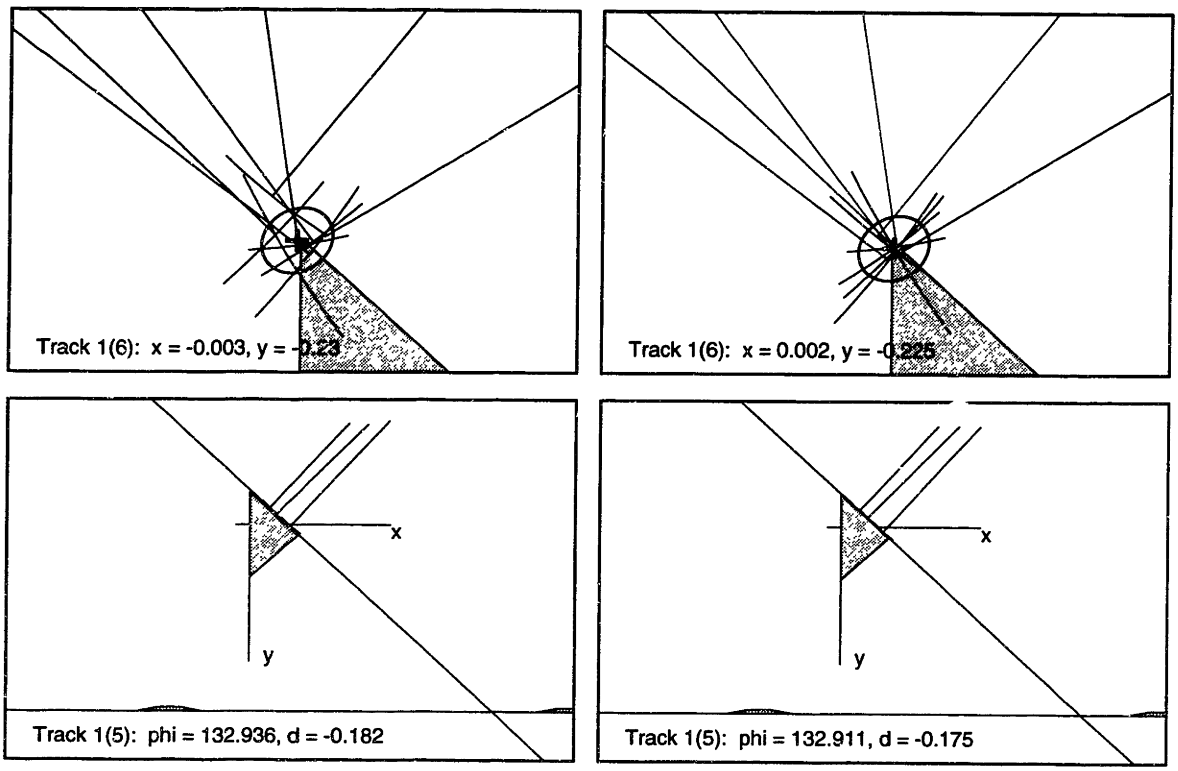


Figure 5.14: State estimates due to uncertain motion. Even with a growing position uncertainty ($\sigma_0 = 1$ cm and $\sigma_1 = 0.02$) a correctly partitioned set of observations produces inconsequential error in the final result for this example. The algorithm encounters real difficulty when the error grows large enough to influence the assignments, however. *Top row*, the vertex state estimate calculated from noisy data, *left*, despite the larger noise deviation, differs by only 5.1 mm from the ideal, *right* (repeated from figure 4.7 on page 96). *Bottom row*, similarly for a plane, the 7 mm and 0.02° error differs little from the ideal, shown at *right*.

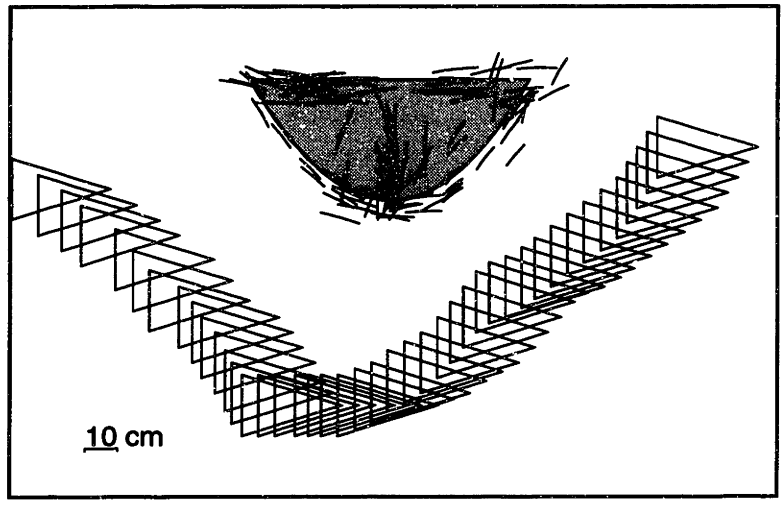


Figure 5.15: Noisy measurements for the parabola experiment. We set the sensor motion uncertainty model to reflect growing error, $\sigma_0 = 1$ cm and $\sigma_1 = 0.02$.

belonging to the parabola shown in figure 5.16. The lack of a consistent treatment of the position error contributes to this shortcoming.

One strategy for addressing this issue is to treat the location of the sensor in the same manner as the observations it produces—a measured quantity that has an error associated with it explicitly. We expand the sensor model to include the covariances of x and y. Since the error in our hand measurements of the location are much smaller than that of the sensor values, we have actually been ignoring their contribution. Whether or not correct partitioning results remains unresolved, however.

5.4.3 Sparse Data

The final item we investigate in this chapter deals with the rate at which data arrives, *i.e.*, how far the hypothetical vehicle has traveled between each scan. In Chapter 4, we made the analogy that tracking represented a sequence of estimates of the local *differential* properties intrinsic to a contour. This realization certainly should indicate that the best performance results from closely spaced scans. The possibility to sample too densely indeed exists, but the reduction in performance should be confined to the speed in which the filter processes the measurements rather than the ultimate assignments made. We therefore wish to try the

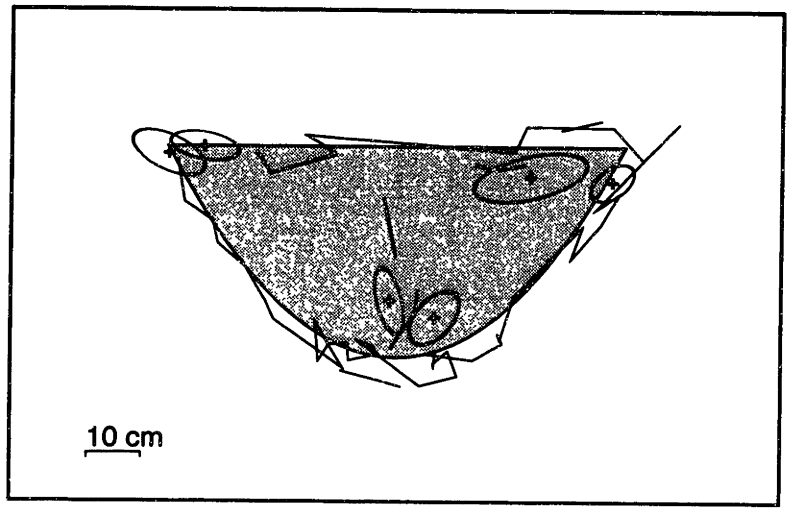


Figure 5.16: Erroneous results due to uncertain motion. The algorithm actually tracks the entire face, but with several omissions. It introduces an extra face of one segment based upon several observations. The poor performance of the algorithm, particularly with respect to the locations of the link points in the face, demonstrates the difficulty of growing position uncertainty $\sigma_0 = 1$ cm and $\sigma_2 = 0.02$.

opposite case, in which we attempt to track the curved face of the parabola by *removing* some of the scans from the sequence used in the parabola experiment of Section 5.2.3.

Figure 5.17 shows the result of two instances of the scene in which we generate scans less frequently. In the first one, we remove approximately every other scan from the original experiment. The algorithm performs robustly under this degree of adversity, finding the best solution for all practical purposes. In the second case, the algorithm fares less well. When we remove two-thirds of the scans, it misses nearly all of the curved face, treating the majority of observations as false alarms. The end result contains a few vertices instead of the face, the most plausible explanation based on the limited data available.

5.5 Summary

We have demonstrated the validity of the model-based multiple hypothesis approach to tracking observations of sharp corners and smooth faces, both flat and curved. The data sets used as input to the algorithm came from experiments performed underwater using a

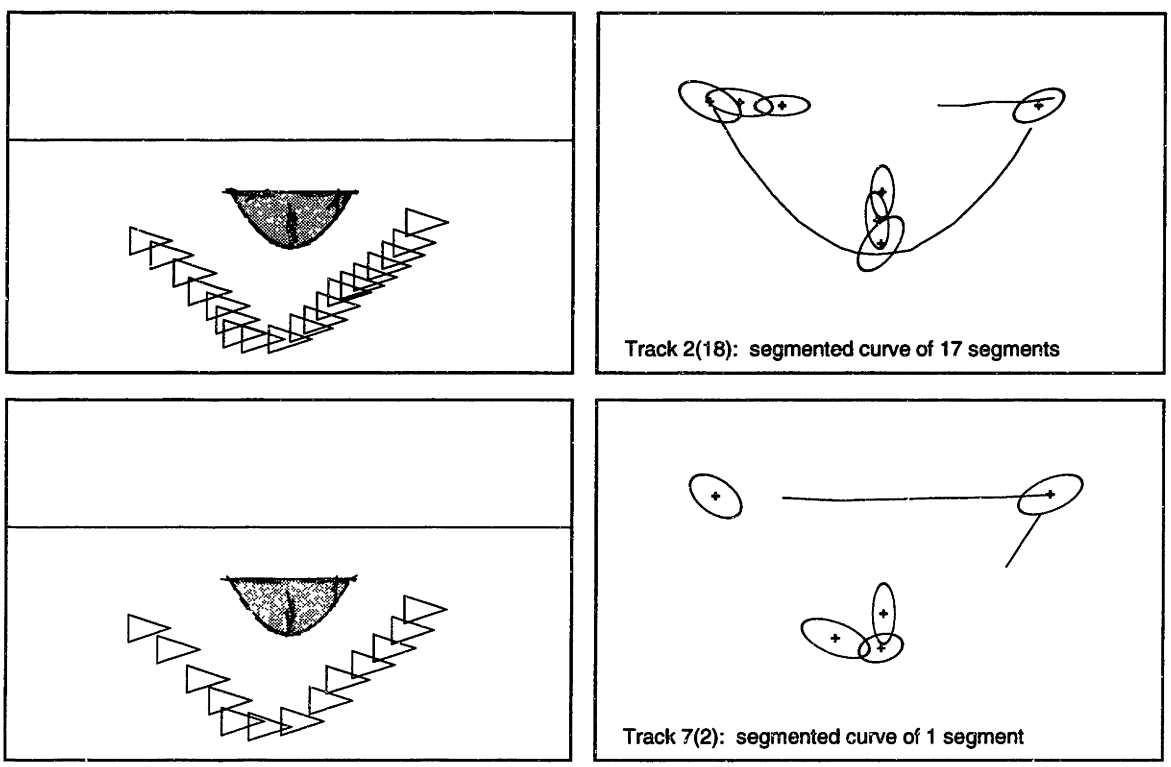


Figure 5.17: Increased distance between scans. *Top row*, when we removed approximately every other scan from the original set, the algorithm still performs reasonably well, assigning 18 RCDs to the curved face. *Bottom row*, removing at least two-thirds of the original scans, the curved face nearly vanishes. The MHT filter explains the observations as being mostly false alarms accompanied by a pair of vertices.

commercially available profiling sonar.

Our failed first attempt at tracking a curved face led us to understand how the distance between the sensor and target, relative to its radius of curvature, influences the performance of the MHT. The parabola experiment described in Section 5.2.3 demonstrates successful tracking of a curved face when the relative distance is reduced.

We have also presented results that reflect conditions contrary to the initial assumptions made in Chapter 2. The issues regarding rough surfaces and uncertain sensor location lead to recommendations for future research as well as the desire to extend the algorithm to three dimensions. A discussion of these points follows in the next chapter.

Chapter 6

Concluding Remarks

To address our original objective of underwater shape reconstruction, we have developed and implemented an algorithm to interpret the observations that result from an acoustic range sensor. In the previous chapter, we demonstrated the validity of the algorithm by applying it to real data produced during a series of underwater experiments. In this chapter, we summarize the specific research contributions made by our implementation of the model-based MHT filter. Finally, in support of our desire to see this work extended such that it finds application in the field of deep ocean object retrieval using an AUV, we offer insights on future research issues.

6.1 Summary of Contributions

We have extended the model-based multiple hypothesis tracking filter described by Cox and Leonard [20], an improved version of the original MHT offered by Reid [63]. Our implementation provides, for the first time, the ability to track measurements that correspond to *curved* faces of objects. The significance of this achievement follows from its explicit focus on explaining the origins of measurements, while simultaneously addressing the inaccuracy in their values. Do they correspond to a sharp corner, a curved face, or random clutter in the environment?

This question is difficult to answer in practice. In regions of high curvature, sequential observations of a face evolve in a manner that creates the illusion that they correspond

to a sharp edge. Based on the elliptical mast experiment described in Chapter 5, we now understand the role of *characteristic range* in determining the curvature of the object relative to the nominal sensor range. The algorithm we developed successfully partitions measurements of this type providing the relative curvature is moderate.

The algorithm owes its success in part to the target models described in Chapter 4. The plane and sculpt target behaviors that track smooth faces account for observations that result from specular reflection of acoustic waves and the vertex target behavior accounts for diffracted waves. Together these models exploit the geometric constraints put forth by the strict alignment conditions of *isolated targets*, as discussed in Section 3.1.2. The representation we use to model the environment in this geometric approach to sonar interpretation explicitly makes available the boundary of objects, allowing efficient interrogation for our long term objective of supporting autonomous object retrieval.

Our approach also provides useful information in real time. The tracking filter produces local geometry estimates of the surrounding scene in the vicinity of the sensor with every iteration. It uses this information to improve the accuracy of bearing angles in existing measurements. We also envision local geometry estimates being applicable to navigation, helping an autonomous vehicle to bound the uncertainty in its location.

The architecture provides a framework which allows complex post-processing, performing cognitive tasks for example. As we have demonstrated in our global geometry synthesis in Section 5.1, the appropriate place to address higher level issues occurs after the tracking filter has suitably partitioned the measurements, thus preventing any bottlenecks in the tracking filter. We advocate this architecture because of its applicability to underwater vehicle operations. Unlike their terrestrial counterparts, free swimming underwater vehicles are always in motion. We seek to develop as robust a scheme as possible to meet the rigorous demands of the underwater environment.

6.2 Future Research Issues

With the philosophy that one should learn to walk before attempting to run, we chose to focus on underwater reconstruction of the shape of two dimensional objects that contain

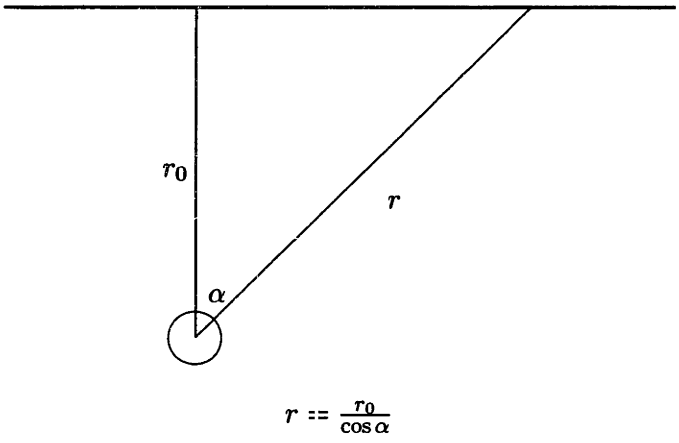


Figure 6.1: Rough wall feature detector. The circle indicates a narrow beam sensor located at a nominal distance r_0 from a rough wall. The bearing angle and range of each return will obey the reciprocal cosine relationship shown.

smooth surfaces. We also made the assumption that precise knowledge of the sensor location was available. As a result, our approach applies more to areas containing manufactured objects than it does to natural settings. Rarely do we find an unstructured ocean environment which meets this description.

Natural settings present significant challenges to our geometric approach to data interpretation. We must therefore offer an opinion on extensions to the approach which offer the potential to model three dimensional scenes which contain rough surfaces as well as smooth. Following these issues, we close with some ideas that address uncertain sensor location.

6.2.1 Rough Surfaces

The structure of our algorithm provides two possible places where we might address the treatment of surface roughness. The first one exists at the feature extraction level. Reviewing figure 3.8 on page 56, we see that the rough wall stands out more prominently than any other feature in the whole scan. So many returns originate from it that we might deduce its presence from a single scan. The appropriate feature detector in this situation applies the reciprocal cosine relationship demonstrated in figure 6.1 to match returns. We then use

the knowledge of its existence to prefilter the raw returns so that the feature extractor only reports a single observation whose bearing angle has normal alignment.

Alternatively, the pattern found in the global geometry shown in figure 5.13 on page 124 suggests a second possibility—a postprocessing approach which combines the face with all of the vertices whose normal distance is sufficiently small. The face then inherits an attribute of roughness enabling the filter to explain subsequent observations. The philosophy behind our overall architecture particularly complements this strategy.

Though perhaps trivial, we must consider a third possibility. We can ignore the shortcoming entirely. Since point scatterers inherently radiate less energy than smooth surfaces, the detection always occurs delayed. As a result, the vertices fall slightly *behind* the face track, which careful inspection of figure 5.13 reveals. In light of these observations, who cares if the filter classifies all of these vertices? Their existence in no way affects the manner in which we use the reconstruction of the scene.

6.2.2 Towards Three Dimensions

One reason for our initial restriction that the environment be two dimensional comes from the choice of sensor. A single-axis profiling sonar only generates 2D data. Recognizing the limitations nevertheless, we need to extend the algorithm to three dimensional objects.

Considerable challenges face us in extending some of the components of our approach. The level of complexity in the local differential properties increases for example, though the overall analogy remains sound—tracking the evolution of an object boundary by processing range measurements into an estimate of its local geometry. Some components, on the other hand, extend quite readily to the 3D case.

The region of constant depth feature, for example, has an obvious 3D counterpart. We now seek to detect plateaus in the resulting 2D scans (also known as depth maps). An RCD_2 contains angles of *azimuth* and *elevation* along with the usual range value, although we maintain our reluctance to trust the angle estimates even when the sensor has a relatively narrow beamwidth.

Several more difficult issues involved in extending the approach to 3D arise when we attempt to solve for the origin of measurements and estimate the underlying geometry

simultaneously. Three dimensional objects contain three different types of object features: faces, vertices and *edges*.[1] Indeed, we must contend with a new class of targets, which is far from intuitive.

The mechanism of diffraction governs the nature of the acoustic energy scattered by an edge, indicating its presence in the scan plane to which it is normal. What we call vertices in 2D actually correspond to the real world edges found in our targets. Because of their vertical alignment and our horizontal sensor, however, the vertex model more appropriately characterizes their behavior.

In 3D, vertices still do exist, largely unchanged from 2D except in that they inherit a third coordinate. They too act as diffractive scatterers, visible in fact from all unobstructed locations.

Faces similarly inherit another degree of freedom; we must now consider their shape in two directions, rather than just the one. One aspect of faces remains constant, we still postulate that the specular reflection model governs their echo formation.

Tracking

Whereas in our earlier investigation we successfully tracked the 2D faces of objects by using the tangent line and circular arc as estimates of the local shape. For a 3D face, we must now provide models to estimate both the *tangent plane* and the two *principal curvatures*, the latter necessary to describe the local shape of sculpted surfaces which curve in both directions.[2]

Track initiation proves rather difficult for faces and edges in the 3D realm. For the vertex and plane target models, we have a straightforward analogy in which initialization directly follows the approach taken in two dimensions. The real difficulty in starting a face or edge track shares the same issue of full state initialization, which we encountered in our earlier discussion of the 2D sculpt behavior model. Finding that the dimensionality of the target state estimate prohibited its initialization from a single measurement, we adopted a

[1] To clear up the terminology, the 3D names have intuitive explanation: faces are object surfaces; edges define the boundaries of faces; and vertices define the endpoints of edges. We exercised artistic license when we borrowed the terms face and vertex to describe 2D features.

[2] We have, perhaps, revealed the origin of the names *plane* and *sculpt* in our tracking primitives.

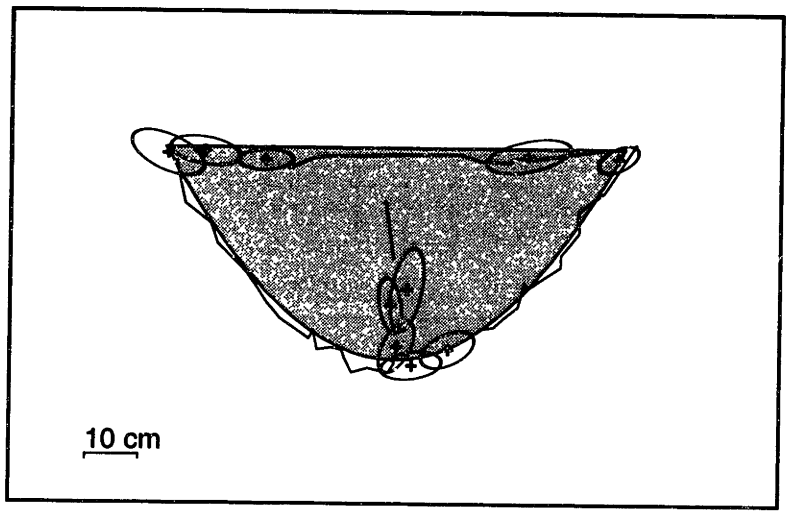


Figure 6.2: Improved results due to better navigation. With bounded uncertainty in the position of the sensor, $\sigma_0 = 1$ cm and $\sigma_2 = 0$, the performance of our algorithm improves significantly (cf. figure 5.16 on page 128).

strategy in which the higher order models evolve out of the existence of lower order ones.

This framework consistently deals with all behavior models similarly to our two dimensional approach, however it does introduce a higher branch factor which in turn slows down the algorithm. To achieve the level of robustness demonstrated in this thesis, we will most likely have to increase the depth of the hypothesis tree to adjust for the delay in initialization of higher order models.

6.2.3 Position Uncertainty

Unbounded growth in the uncertainty of the sensor position, a characteristic of dead reckoning type navigation systems, causes a real problem for autonomous data interpretation, as previously discussed in Section 5.4.2. Externally referenced systems improve upon this difficulty, however, as figure 6.2 demonstrates. Relying on a system with a vanishing drift rate, $\sigma_1 = 0$, but with a small fixed error, $\sigma_0 = 1$ cm, the algorithm fares reasonably well.

A heuristic proposed by Leonard [44] for concurrent map building and navigation offers a possible method to attain bounded uncertainty without using a predeployed positioning system. As the map is built, some environment features emerge with higher accuracy than

others. The potential exists to use these features as navigational beacons in order to reduce the uncertainty in the location. The approach allows refinements to the map only when the position error is sufficiently small, and movement of the vehicle only when the map error is sufficiently small.

To the best of our knowledge autonomous vehicles that have built their own maps and navigated by them have met success only in structured environments. In the final analysis, this issue remains an open challenge to the research community.

Appendix A

Experimental Techniques

In this appendix, we detail the procedure used to calibrate the ST1000 profiling sonar, we describe the mounting apparatus used to locate the sensor precisely, and finally discuss some of the difficulties encountered working underwater and our techniques to overcome them.

A.1 Calibrating the Acoustic Range Sensor

Considering its price, ease of use, and robust design, the ST1000 does a pretty good job of measuring the TOF of an echo pulse. We have already seen that, due to the bandwidth of the demodulation filter, the rise time in the RSS of about 60 μs delays the time of detection. To correct for this, the system engineer [16] subtracts a fixed value, experimentally determined in his laboratory, to shift the TOF back to the desired location at the up crossing of the axis. Repeated observations made during the course of our experiments confirm that the ST1000 measures TOF in a highly repeatable manner for a given target, and with only small variation over the full range of targets used.

Even if we can measure the TOF accurately, we may erroneously calculate the range if we do not accurately know the speed of sound. Here we have a shortcoming of the ST1000, which assumes a fixed value for c of 1475 m/s. We have operated the system in fresh water (salinity $S = 0‰$) at a depth of $z = 0.8$ m and over a temperature range of T =16–24 °C.

Using the empirical model from Clay and Medwin [17],

$$c = 1449.2 + 4.6T - 0.055T^2 + 0.00029T^3 + (1.34 - 0.010T)(S - 35) + 0.016z \qquad \text{(A.1)}$$

the speed of sound for our experiments varies through a range of 1469–1493 m/s.

Our earliest attempt at calibration used a slope-intercept model, in which for a sonar range r_s the corrected range $r_c = ar_s + b$. Using equation (A.1), $a = c/1475$. We determine b by measuring the range to a fixed target at a known distance and comparing it to ar_s. This technique proved inadequate due to the group delay phenomenon (see figure 3.13) which makes b strongly depend on our choice of target. We cannot remedy this because including any *a priori* knowledge of our target severely limits the utility of our approach.

In the end, we correct only for the error in assumed sound speed of the sensor. We can improve on the accuracy of equation (A.1) by inferring the actual speed of sound directly from the scans and the known relative motion of the sensor. We do this by placing a carefully aligned target in the water at some arbitrary location (x_0, y_0). We then echorange from several different locations (x_0, y_i). Let r_i represent the values of these sonar measurements. The least-squares line of best fit through the points (y_i, r_i) has slope equal to $c/1475$. This method gives reasonable accuracy over a wide range of targets, examples of which are shown in figure 3.3, but its limitations are evident.

A.2 Description of the Apparatus

The mounting apparatus designed specifically for these experiments, shown holding the ST1000 in figure A.1, contains a polyethylene sleeve which provides a snug fit to within 5 mil for the sensor housing. The sleeve attaches to one end of an aluminum shaft which has sufficient length to place the sensor head at half the water depth, thus eliminating the effects of multipath from the surface and bottom of the water column.

The top of end of shaft attaches to a two-axis gimbal to ensure that the shaft was accurately aligned in the vertical, which proved of paramount importance in locating the horizontal position of the sensor. By lowering a plumb bob from above, we established the horizontal position not of the sensor itself, but only of the screw holes on the gimbal plate

shown in figure A.5 and figure A.3. Since the sensor was approximately 1 m below the gimbal, a small inaccuracy in shaft alignment became a position error in the range of 1 cm. With the bubble centered in the spirit level, as shown in figure A.5, the shaft was vertical to within 0.1°.

The gimbal attaches to a pair of orthogonal carriages which allow movement in a horizontal plane. The longitudinal rail supporting the overhead carriage defines the x-axis of a right handed coordinate system, positive to the right in figure A.4. Longitudinal position is determined by placing a ruler against the front of the carriage and reading its position on a steel tape which of known reference, as shown in figure A.4. A transverse carriage mechanism defines the y-axis, negative to the right in figure A.3. The transverse position is read directly from the notched indicator in figure A.3. The combination of the system and experimental technique as described allow precise positioning of the vantage point, accurate to the nearest few millimeters.

A.3 Some Minor Difficulties

The first attempt at performing the elliptical mast experiment, made with the mast flooded, resulted in data of poor quality—the RCDs did not line up according to the hand measured location of the object. We realized that the thickness of the aluminum included a region which corresponds to exactly one wavelength (5.3 mm) which happened to line up with the worst part of the data.

A homogeneous layer with thickness equal to an integral number of half-wavelengths, sandwiched between two identical homogeneous half spaces, has a theoretical reflection coefficient of zero under normally incident insonification. An aluminum layer between water and air, on the other hand, is nearly a perfect reflector even when its thickness equals one wavelength (cf. figure 3.12 on page 64). This realization led to closing the bottom end of the mast with a wood, epoxy, and fiber glass cap, which can be seen to the left of the sensor in figure 2.2.

A second attempt, shown in figure A.6, was made with the plugged target. Because the target became quite buoyant, it required being ballasted with approximately 22 kg of

lead. This trial yielded a surprise: data only marginally better, for all practical purposes, than the flooded mast. A more careful analysis of the target, which includes the group delay discussion in Section 3.2, suggests that the water-backed aluminum actually is the better target when considered from a time-of-flight perspective. Figure 3.13 on page 65 corroborates this finding.

The target strength versus group delay debate ultimately became moot when it was discovered that because of its small footprint, the mast wobbled to and fro when placed on the slightly uneven concrete floor of the tank. As little as one degree out of vertical caused a deviation in horizontal position of 1.4 cm between the perceived location, determined by a plumb bob at the top, and the horizontal position at the sensor depth of 0.8 m. This error was indeed the culprit which caused the poor data.

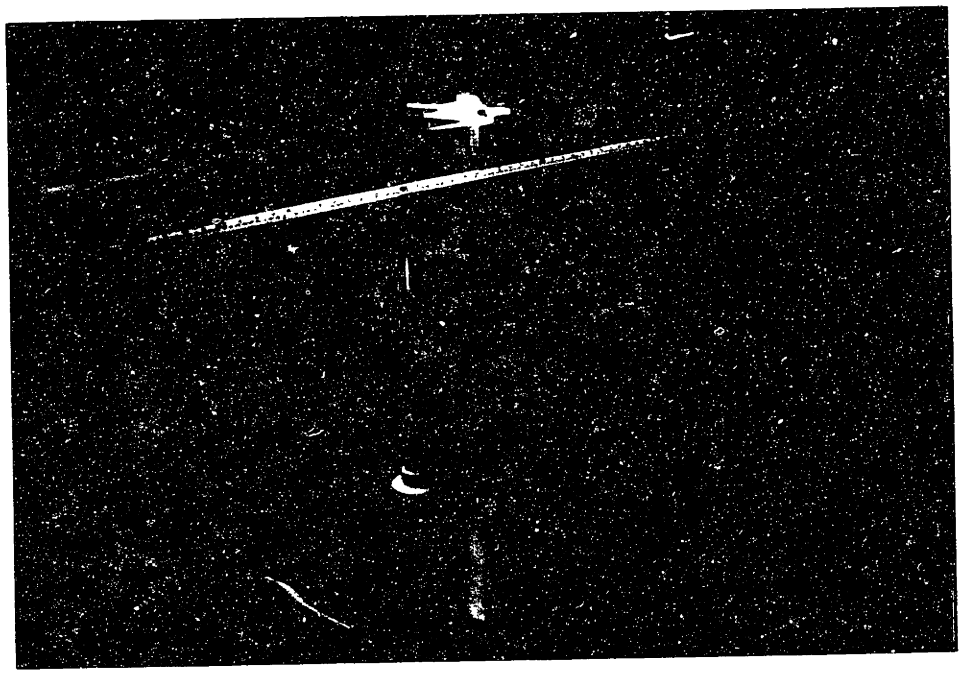


Figure A.1: Sensor mounting system

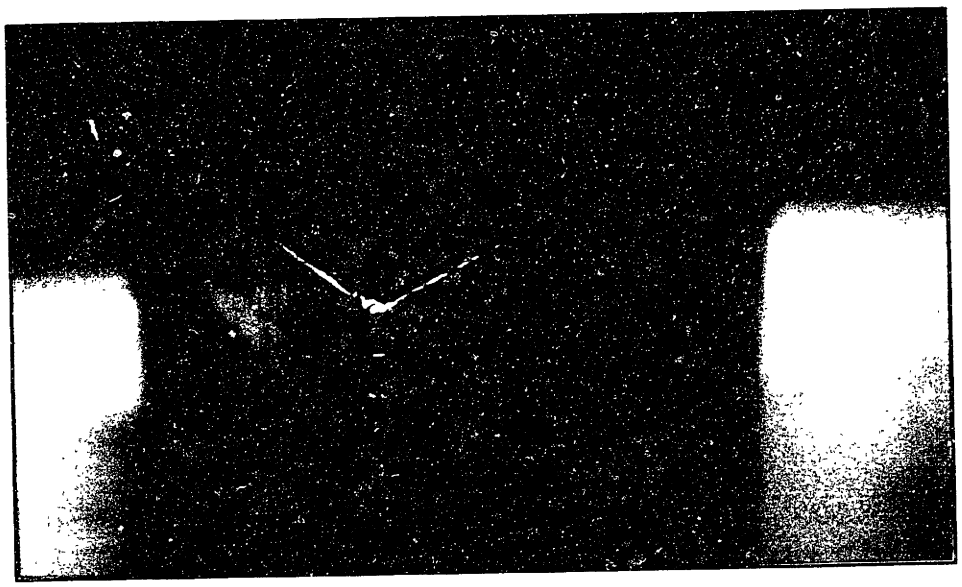


Figure A.2: Triangular prism of first experiment

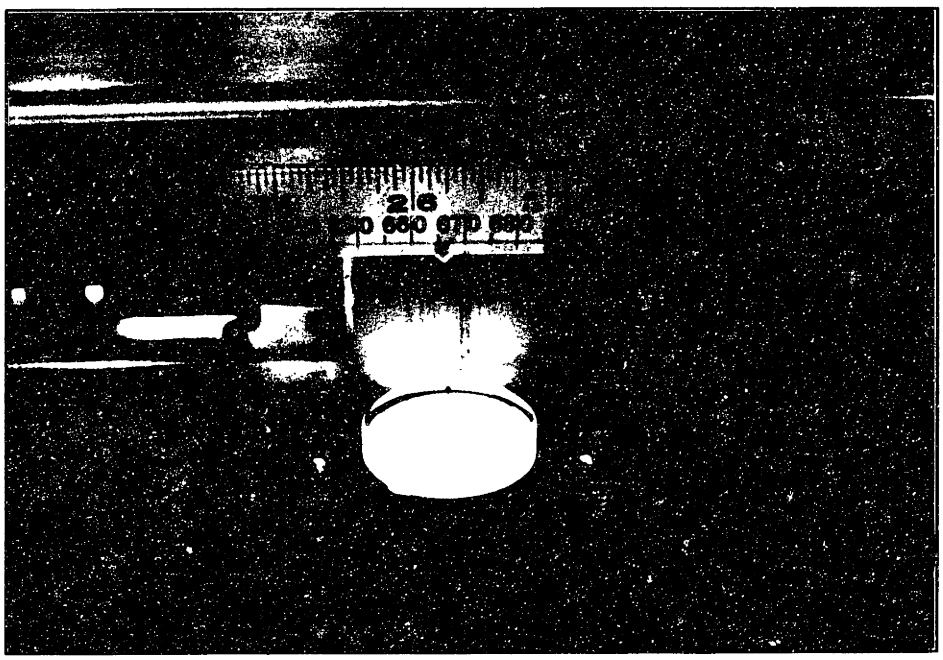


Figure A.3: Transverse positioning, the y-direction

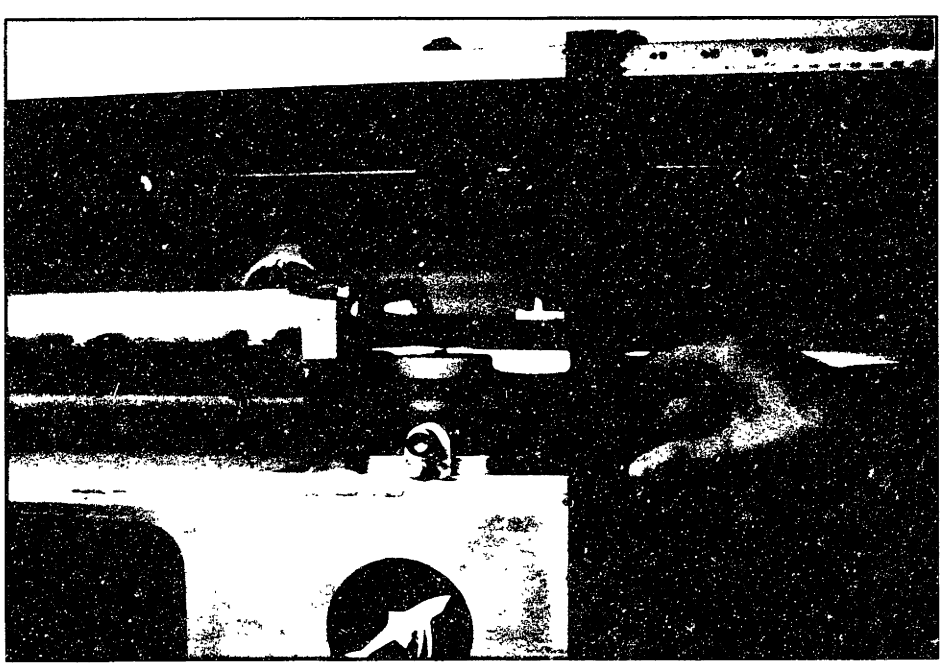


Figure A.4: Longitudinal positioning, the x-direction

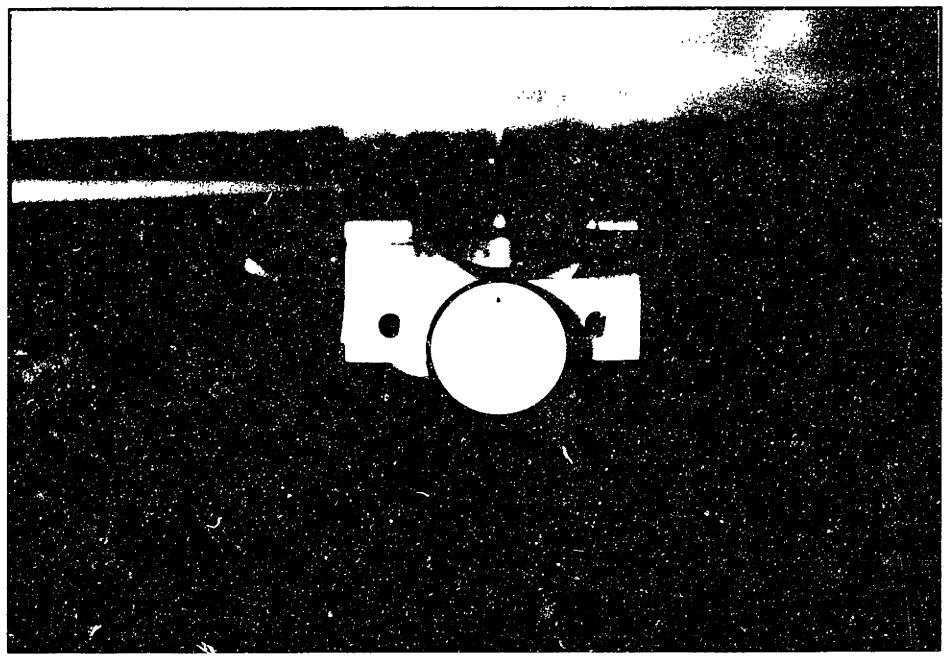


Figure A.5: Vertical alignment of the sensor axis. With bubble centered, this system reduced the vertical alignment error to less than 0.1°.

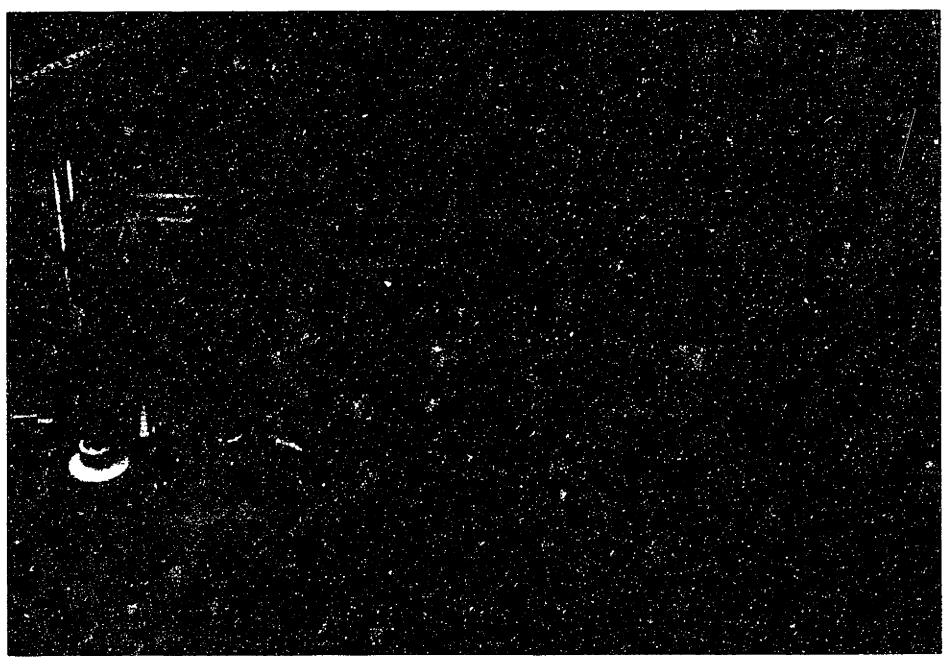


Figure A.6: Elliptical mast in place for second experiment

Bibliography

[1] M. Abramson. A localization algorithm/map matching approach for underwater vehicle navigation. In *Sensor and Navigation Issues for Unmanned Underwater Vehicles*, Cambridge, MA, January 1991. MIT Sea Grant Marine Industry Collegium.

[2] J. Achenbach and A. Gautesen. Geometrical theory of diffraction for three-D elastodynamics. *J. Acoustical Soc. of Am.*, 61(2):413–421, February 1977.

[3] W. Au. *The Sonar of Dolphins.* Springer-Verlag, 1993.

[4] K. Audenaert, H. Peremans, Y. Kawahara, and J. Van Campenhout. Accurate ranging of multiple objects using ultrasonic sensors. In *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pages 1733–1738, May 1992.

[5] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky. An overview of matched field methods in ocean acoustics. *IEEE J. Oceanic Engineering*, 18(4):401–424, October 1993.

[6] A. B. Baggeroer, W. A. Kuperman, and H. Schmidt. Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem. *J. Acoustical Soc. of Am.*, 83(2):571–587, 1988.

[7] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association.* Academic Press, 1988.

[8] J. G. Bellingham, J. W. Bales, D. K. Atwood, M. Perrier, C. A. Goudey, T. R. Consi, and C. Chryssostomidis. Performance characteristics of the Odyssey AUV. In *Proc. of the 8th Intl. Symposiom on Unmanned Untethered Submersible Technology*, pages 37–49. Autonomous Undersea Systems Institute, 1993.

[9] J. G. Bellingham, C. A. Goudey, T. R. Consi, and C. Chryssostomidis. A small long-range autonomous vehicle for deep ocean exploration. In *Proc. of the 2nd Intl. Offshore and Polar Conf.*, pages 461–467, 1992.

[10] L. M. Brekhovskikh and Y. P. Lysanov. *Fundamentals of Ocean Acoustics.* Springer-Verlag, 2d edition, 1991.

[11] W. L. Brogan. Algorithm for ranked assignments with applications to multiobject tracking. *IEEE J. of Guidance*, 12(3):357–364, 1989.

[12] M. K. Brown. Feature extraction techniques for recognizing solid objects with an ultrasonic range sensor. *IEEE J. Robotics and Automation*, RA-1(4):191–205, December 1985.

[13] M. K. Brown. The extraction of curved surface features with generic range sensors. *Intl. J. of Robotics Research*, 5(1):3–18, 1986.

[14] W. S. Burdic. *Underwater Acoustic System Analysis.* Prentice-Hall, 1992.

[15] J. F. Canny. Finding edges and lines in images. Master's thesis, Massachusetts Institute of Technology, June 1983.

[16] M. Cardew. Telephone conversation with author, 1993. System Technologies, Ulverston, UK.

[17] C. S. Clay and H. Medwin. *Acoustical Oceanography: Principles and Applications.* John Wiley & Sons, 1977.

[18] R. F. W. Coates. *Underwater Acoustic Systems.* Macmillan Education, London, 1990.

[19] M. Cooper and R. Whaley. Finding a needle in a haystack. In *Oceans 91 Proc.*, pages 169–174. Oceanic Engineering Society of IEEE, 1991.

[20] I. J. Cox and J. J. Leonard. Modeling a dynamic environment using a Bayesian multiple hypothesis approach. *Artificial Intelligence*, 66:311–344, 1994.

[21] I. J. Cox and M. L. Miller. On finding ranked assignments with application to multitarget tracking and motion correspondence. Technical report, NEC Research Institute, 1994.

[22] T. B. Curtin, J. G. Bellingham, J. Catipovic, and D. Webb. Autonomous oceanographic sampling networks. *Oceanography*, 6(3):86–94, 1993.

[23] G. Dangelmayr. Asymptotic inverse scattering. *Wave Motion*, 6:337–357, 1984.

[24] G. Dangelmayr and F. J. Wright. Geometrical inversion for a scattering curve. *Inverse Problems*, 2:293–305, 1986.

[25] C. Dean and M. Werby. Target shape and material composition from resonance echoes of submerged elongated elastic targets. In *Proc. SPIE Conf. on Automatic Object Recognition*, volume 1471, pages 54–65, Orlando, FL, 1991.

[26] R. V. Denton, A. J. Rockmore, B. Friedlander, and J. S. Malin. An image reconstruction approach to target association. In *Proc. IEEE Asilomar Conf. on Circuits, Systems, and Computers*, pages 304–308, Pacific Grove, CA, November 1978.

[27] M. do Carmo. *Differential Geometry of Curves and Surfaces.* Prentice-Hall, 1976.

[28] L. B. Felson. Target strength: Some recent theoretical developments. *IEEE J. Oceanic Engineering*, 12(2), April 1987.

[29] L. Flax, G. C. Gaunaurd, and H. Überall. Theory of resonance scattering. In W. P. Mason and R. N. Thurston, editors, *Physical Acoustics: Principles and Methods*, volume XV, chapter 3, pages 191–294. Academic Press, 1981.

[30] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE J. Oceanic Engineering*, OE-8:173–184, July 1983.

[31] A. Freedman. The high frequency echo structure of some simple body shapes. *Acustica*, 12(2):259–268, 1962.

[32] A. Freedman. A mechanism of acoustic echo formation. *Acustica*, 12(1):10–21, 1962.

[33] G. C. Gaunaurd. Inverse scattering problems in the acoustic resonance region of an underwater target. In *Proc. SPIE Conf. on Automatic Object Recognition*, volume 1471, pages 30–41, Orlando, FL, 1991.

[34] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization.* Academic Press, 1981.

[35] D. R. Griffin. *Listening in the Dark: The Acoustic Orientation of Bats and Men.* Yale University Press, 1958.

[36] J. C. T. Hallam. *Intelligent Automatic Interpretation of Active Marine Sonar.* PhD thesis, University of Edinburgh, 1984.

[37] B. K. P. Horn. *Robot Vision.* MIT Press, 1986.

[38] M. Kayton, editor. *Navigation: Land, Sea, Air, and Space.* IEEE Press, 1990.

[39] L. Kleeman and R. Kuc. Mobile robot sonar for target localization and classification. Research Report ISL-9301, Intelligent Sensors Laboratory, Department of Electrical Engineering, Yale University, 1993.

[40] J. J. Koenderink. *Solid Shape.* The MIT Press, 1990.

[41] R. Kuc and M. W. Siegel. Physically based simulation model for acoustic sensor robot navigation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-9(6), November 1987.

[42] T. Kurien. Issues in the design of practical multitarget tracking algorithms. In Y. Bar-Shalom, editor, *Multitarget-Multisensor Tracking: Advanced Applications*, pages 43–83. Artech House, 1990.

[43] L. D. Landau and E. M. Lifshitz. *Fluid Mechanics.* Pergamon Press, 2d edition, 1987.

[44] J. J. Leonard and H. F. Durrant-Whyte. *Directed Sonar Sensing for Mobile Robot Navigation.* Kluwer Academic Press, Boston, 1992.

[45] J. Lighthill. *Waves in Fluids.* Cambridge University Press, 1978.

[46] T. Lozano-Pérez. Foreword. In I. J. Cox and G. T. Wilfong, editors, *Autonomous Robot Vehicles*, pages vii–xi. Springer-Verlag, 1990.

[47] G. R. Mackelburg. Acoustic data links for UUVs. In *Oceans 91 Proc.*, pages 1400–1406. Oceanic Engineering Society of IEEE, 1991.

[48] Marquest Telepresence, Bourne, MA. Commercial offer: EXACT acoustic navigation system, 1992.

[49] D. McNamara, C. Pistorius, and J. Malherbe. *Introduction to the Uniform Geometrical Theory of Diffraction.* Artech House, Boston, 1990.

[50] M. L. Miller. *Implementation notes for MHT with multiple target models.* NEC Research Institute, 1993.

[51] D. A. Mindell and D. R. Yoerger. A hybrid parallel control and navigation system for the autonomous benthic explorer (ABE) vehicle. In *Proc. of the 8th Intl. Symposiom on Unmanned Untethered Submersible Technology.* Autonomous Undersea Systems Institute, 1993.

[52] H. P. Moravec and A. Elfes. High resolution maps from wide angle sonar. In *Proc. of the IEEE Intl. Conf. on Robotics and Automation*, pages 116–121, St. Louis, 1985.

[53] C. L. Morefield. Application of 0-1 integer programming to multi-target tracking problems. *IEEE Trans. Automatic Control*, AC-22(6):302–312, June 1977.

[54] S. Mori, C. Chong, E. Tse, and R. Wishner. Tracking and classifying multiple targets without a priori identification. *IEEE Trans. Automatic Control*, AC-31(5):401–409, May 1986.

[55] M. E. Mortenson. *Geometric Modeling.* John Wiley & Sons, 1985.

[56] C. Mullen. The Lucona affair. In *Oceans 91 Proc.*, pages 185–188. Oceanic Engineering Society of IEEE, 1991.

[57] K. G. Murty. An algorithm for ranking all the assignments in order of increasing cost. *Operations Research*, 16:682–687, 1968.

[58] V. Nagarajan, M. R. Chidambara, and R. N. Sharma. Combinatorial problems in multitarget tracking—a comprehensive survey. In *IEE Proc. pt. F, Communications, Radar, and Signal Processing*, volume 134(1), pages 113–118, 1987.

[59] Numerical Algorithms Group, Oxford. *NAG Fortran Library, Mark 14*, 1990.

[60] C. B. Officer. *Introduction to the theory of sound transmission, with application to the ocean.* McGraw-Hill, New York, 1958.

[61] H. Peremans, K. Audenaert, and J. M. Van Campenhout. A high-resolution sensor based on tri-aural perception. *IEEE Trans. Robotics and Automation*, 9(1):36–48, February 1993.

[62] A. N. Popper. Sound emission and detection by delphinids. In L. M. Herman, editor, *Cetacean Behavior: Mechanisms and Functions*, chapter 1, pages 1–52. Robert E. Krieger, Malabar, FL, 1988.

[63] D. B. Reid. An algorithm for tracking multiple targets. *IEEE Transactions on Automatic Control*, AC-24(6):843–854, December 1979.

[64] H. Schmidt. *SAFARI: Seismo-Acoustic Fast field Algorithm for Range Independent environments, User's Guide.* SACLANT ASW Research Centre, La Spezia, Italy, 1987.

[65] H. Schmidt. Numerical modeling in ocean seismo-acoustics. In *Oceans 91 Proc.*, pages 84–92, Honolulu, 1991. Oceanic Engineering Society of IEEE.

[66] H. Schmidt. Numerically stable global matrix approach to radiation and scattering from spherically stratified shells. *J. Acoustical Soc. of Am.*, 94(4):2420–2430, October 1993.

[67] P. Schweizer, W. Petlevich, P. Haley, and J. Oravec. Image processing architectures for autonomous underwater vehicles (AUVs) in mine detection and classification operations. In *Proc. of the 7th Intl. Symposiom on Unmanned Untethered Submersible Technology*, pages 280–293, 1991.

[68] W. E. Shotts and T. M. McNamara, Jr. A blueprint for the use of unmanned undersea vehicles in littoral warfare. *Johns Hopkins APL Technical Digest*, 14(2):148–153, April-June 1993.

[69] W. McC. Siebert. *Circuits, Signals, and Systems.* MIT Press, 1986.

[70] R. A. Singer and A. J. Kanyuck. Computer control of multiple site correlation. *Automatica*, 7:455–463, July 1971.

[71] R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial relationships in robotics. In I. J. Cox and G. T. Wilfong, editors, *Autonomous Robot Vehicles.* Springer-Verlag, 1990.

[72] W. K. Stewart, Jr. *Multisensor Modeling Underwater with Uncertain Information.* PhD thesis, Massachusetts Institute of Technology, 1988.

[73] P. A. Tipler. *Physics.* Worth Publishers, New York, 1976.

[74] G. Trimble, J. Vilaro, D. Okamura, R. Lum, and K. Dutta. Underwater object recognition and automatic positioning to support dynamic positioning. In *Proc. of the 7th Intl. Symposiom on Unmanned Untethered Submersible Technology*, pages 273–279, 1991.

[75] Tritech International, Aberdeen, UK. *ST1000 Technical Notes*, 1992.

[76] S. T. Tuohy. *Geophysical Map Representation, Abstraction and Interrogation for Autonomous Underwater Vehicle Navigation.* PhD thesis, Massachusetts Institute of Technology, 1993.

[77] H. Überall, editor. *Acoustic Resonance Scattering.* Gordon and Breach Science Publishers, 1992. Proc. of a symposium held 22 May 1989 at the Catholic University of America, Washington, D.C. to honor the French pioneers of acoustic resonance experiments.

[78] R. J. Urick. *Principles of Underwater Sound.* McGraw-Hill, 3d edition, 1983.

[79] J. M. Walton. Advanced unmanned search system. In *Oceans 91 Proc.*, pages 1392–1399. Oceanic Engineering Society of IEEE, 1991.

[80] A. Weigel. DSV SEA CLIFF recovery of the UAL 811 cargo door. In *Oceans 91 Proc.*, pages 175–178. Oceanic Engineering Society of IEEE, 1991.