Exploring Scatterer Anisotropy
in Synthetic Aperture Radar
via Sub-aperture Analysis

by

Andrew J. Kim

B.E.E., Georgia Institute of Technology, 1995
S.M.E.E., Massachusetts Institute of Technology, 1997

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

September 2001

Signature of Author: _____

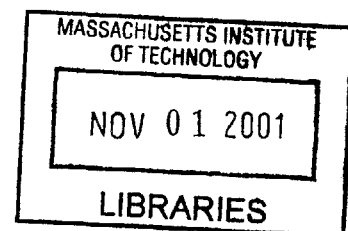Dept. of Electrical Engineering and Computer Science
August 31, 2001

Certified by: _____

Alan S. Willsky
Edwin S. Webster Professor of Electrical Engineering
Thesis Supervisor

Accepted by: _____

Arthur C. Smith
Professor of EECS
Chair, Committee for Graduate Students

BARKER

Exploring Scatterer Anisotropy
in Synthetic Aperture Radar
via Sub-aperture Analysis

by Andrew J. Kim

Submitted to the Department of Electrical Engineering and Computer Science
on August 31, 2001 in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Electrical Engineering and Computer Science

Abstract

Scattering from man-made objects in SAR imagery exhibits aspect and frequency dependencies which are not always well modeled by standard SAR imaging techniques based on the ideal point scattering model. This is particularly the case for high-resolution wide-band and wide-aperture data where model deviations are even more pronounced. If ignored, these deviations will reduce recognition performance due to the model mismatch, but when appropriately accounted for, these deviations from the ideal point scattering model can be exploited as attributes to better distinguish scatterers and their respective targets. With this in mind, this thesis develops an efficient modeling framework based on a sub-aperture pyramid to utilize scatterer anisotropy for the purpose of target classification. Two approaches are presented to exploit scatterer anisotropy using the sub-aperture pyramid. The first is a nonparametric classifier that learns the azimuthal dependencies within an image and makes a classification decision based on the learned dependencies. The second approach is a parametric attribution of the observed anisotropy characterizing the azimuthal location and concentration of the scattering response. Working from the sub-aperture scattering model, we develop a hypothesis test to characterize anisotropy. We start with an isolated scatterer model which produces a test with an intuitive interpretation. We then address the problem of robustness to interfering scatterers by extending the model to account for neighboring scatterers which corrupt the anisotropy attribution. The development of the anisotropy attribution culminates with an iterative attribution approach that identifies and compensates for neighboring scatterers. In the course of the development of the anisotropy attribution, we also study the relationship between scatterer phenomenology and our anisotropy attribution. This analysis reveals the information provided by the anisotropy attribution for two common sources of anisotropy. Furthermore, the analysis explicitly demonstrates the benefit of using wide-aperture data to produce more stable and more descriptive characterizations of scatterer anisotropy.

Thesis Supervisor: Alan S. Willsky
Title: Edwin S. Webster Professor of Electrical Engineering

Acknowledgments

First and foremost, I thank my advisor Alan Willsky. Any words that I use will lack the eloquence to convey how deeply I appreciate Alan's help over the past six years. His guidance and support have significantly helped me mature as a researcher. Not only has he helped my research with his vast technical insight, but he has also helped me to define, analyze, and explain to others the underlying problems that need to be addressed.

I would also like to thank the members of my thesis committee: John Fisher, Jeff Shapiro, and Paul Viola. Their comments and assistance have also significantly contributed to my research. Particular thanks must be made to John Fisher. I have had the great opportunity to interact with John who has served as both a member of my committee and a research scientist in Alan's group. With his ideas and contributions, John has been like a second advisor to me. John has also been a great friend, Orlando tour guide, and football pool participant.

Throughout my years in college, my life has been influenced by many people. I consider myself most fortunate to have made such friends. Walter Sun has a strength of character in him that I have never seen in anyone else. Our friendship started while we were undergraduates and continues to be strong even though separated by thousands of miles these past few years. Every season, I can look forward to disagreeing with Walter on the relative strength of collegiate teams in various sports. Dewey Tucker is another great friend that I have had the pleasure of knowing since our days at Georgia Tech. Though I tease Dewey a lot for "following me to MIT and the SSG lab[1]", I am glad he is around. Dewey is a great friend with whom I have had great times and been able to talk to during difficult times.

Besides Dewey Tucker and John Fisher, there are a number of great people in the SSG lab that I have enjoyed being around these past six years. John Richards has been a great friend, fellow SAR grouplet member, officemate, classmate, SPIE travel companion, and football pool participant. John could always be counted on for his posting of highlights from "The Onion", jokes about Canadians, and Big-10 sports bias. Alex Ihler and I have shared many interests. I most thank Alex for getting me involved in Shotokan Karate. I will miss our discussions and shared efforts in administering the

---

[1]He will probably also "follow" me back to Atlanta.

Contents

List of Figures

List of Tables

Notational Conventions

Mathematical Notation

$\cdot^*$	conjugation, e.g. a^*
$\cdot * \cdot$	convolution, e.g. $x(t) * y(t)$
$\cdot^H$	hermitian transpose
$\xrightarrow[N \to \infty]{\mathcal{P}}$	convergence in probability as N tends to infinity
$\sim$	distributed according to
$\cdot^T$	transpose
$\|\cdot\|_W$	vector norm derived from an inner product with weighting matrix W, i.e. $\|x\|_W \triangleq x'Wx$
$\chi^2(\cdot, \cdot\cdot)$	Chi-square divergence
$D(\cdot \| \cdot\cdot)$	Kullback-Leiber divergence
$\mathbb{E}[\cdot]$	expectation of a random variable/vector
$\mathcal{F}\{\cdot\}$	Fourier transform
$\mathcal{F}^{-1}\{\cdot\}$	inverse Fourier transform
$\mathrm{Im}\{\cdot\}$	imaginary part of a complex number
$\mathcal{O}(\cdot)$	computational order of an algorithm
$\mathrm{Re}\{\cdot\}$	real part of a complex number

Commonly Used Variables

e_i	Euclidean basis vector for component i
$\mathcal{H}$	set of all sub-aperture anisotropy hypotheses associated with a sub-aperture pyramid
I	identity matrix
m	scale with m=0 denoting the finest scale associated with the full-aperture and $m = M$ denoting coarsest scale associated with the smallest sub-aperture
q_M	set of sub-aperture reflectivity measurements at scale M
$S_{m,i}$	sub-aperture at scale m and offset i
S	sub-aperture pyramid
S_m	set of sub-apertures at scale m of a pyramid
$S_{m,i}$	specific sub-aperture at scale m and offset i in a pyramid

Common Acronyms

FFT	fast-Fourier transform
FM	frequency modulated
GLLR	generalized log-likelihood ratio
GLRT	generalized likelihood ratio test
GTD	geometric theory of diffraction
i.i.d.	independent and identically distributed
KL	Kullback-Leibler
k-NN	k nearest neighbor
LLN	law of large numbers
LLR	log-likelihood ratio
LRT	likelihood ratio test
MC	Monte Carlo
ML	maximum likelihood
MSM	multiple scatterer model
pdf	probability density function
PMF	probability mass function
PSNR	peak signal-to-noise ratio
ROC	receiver operating characteristic
RMS	root-mean-square
SAR	synthetic aperture radar
SNR	signal-to-noise ratio

Chapter 1

Introduction

Synthetic aperture radar (SAR)[17, 42, 43, 52] is a microwave radar[58] imaging system frequently used in fields such as remote sensing, surveillance, and automatic target recognition (ATR). Its popularity is due in large part to its ability to attain a high imaging resolution, rapidly cover large areas of terrain, supply its own illumination, and operate in inclement weather. The SAR imaging scenario is illustrated in Figure 1.1 where a radar is mounted on an airborne or spaceborne craft and takes multiple measurements of a scene. In conventional radar, the cross-range resolution is the width of the shaded footprint illustrated. This width is proportional to the distance between the radar and the target area; and, it is inversely proportional to the size of the antenna. SAR achieves a high resolution by effectively manipulating the latter. In particular, it uses the measurements taken from the different positions along the flight path and fuses them together to synthesize an aperture the length of the integrated flight path, thus producing higher fidelity and higher resolution imagery than that of conventional radar imaging.

The application for SAR by which this thesis is motivated is ATR[8, 9, 23]. This involves using attributes of the SAR data to classify the contents of the image scene. Any ATR algorithm can be loosely regarded as a feature extraction and comparison process. The features may be as simple as the reflectivity value at each pixel, e.g. as used in simple image correlation techniques[56]. They may also be as complicated as precise descriptions of the geometrical structure of the target[25, 34, 13] or as abstract transformations of the data which do not have a clear physical interpretation[10]. A motif of this thesis is the development of a feature to be used in ATR which provides both distinguishing information for discrimination and a physically meaningful interpretation. The need for the former is clear in that the feature should help differentiate various targets of interest. The latter is useful because it allows for an intuitive understanding of what information the feature represents thus guiding its use in ATR algorithms. Furthermore, although the proposed features are designed with ATR in mind, the information they convey is useful in many other contexts due to their physical interpretation.

The features with which this thesis is concerned are motivated by a shortcoming of standard SAR imaging. As technology progresses, the desire for higher resolution imagery drives the development of SAR systems with larger synthetic apertures. The syn-
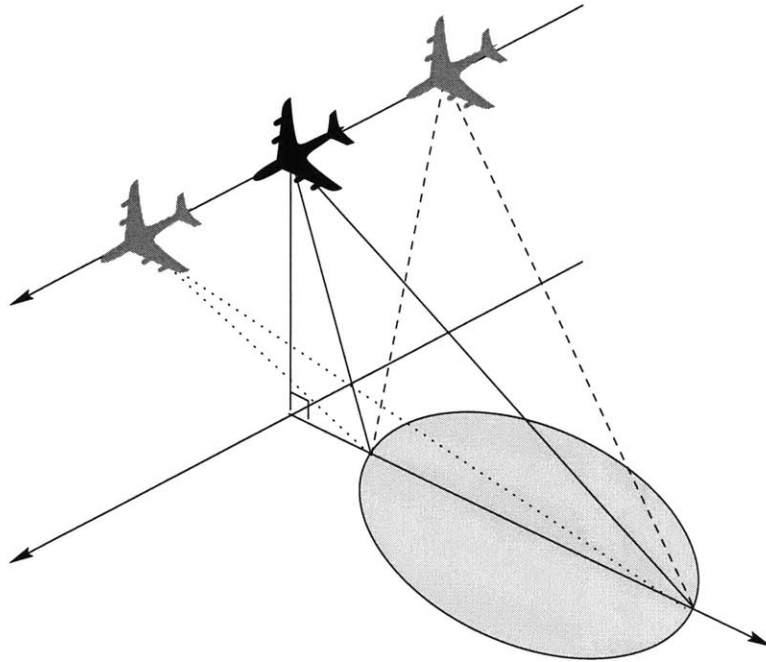
Figure 1.1. SAR Imaging scenario.

thesized aperture is produced with the use of an advanced signal processing technique. As with any signal processing algorithm, its gains are achieved by taking advantage of a model for the data. One such aspect of the model is the assumption of an isotropic point scatterer, i.e. one whose reflectivity is invariant with respect to azimuthal orientation. This assumption may also be equivalently interpreted as saying that all scatterers are smaller than the resolution of the imaging process. Herein lies the problem. Real scatterers have a physical extent to them. Thus, at some point, the scatterers start to span multiple resolution cells as the resolution cell size decreases, or viewed differently, scatterers start to exhibit variable azimuthal responses as the aperture size is increased. In other words, the isotropic assumption begins to break down.

With the increasing availability of wide-aperture data, the issue of how to handle scatterer anisotropy becomes an important issue. A naive (and common) approach is to simply ignore scatterer anisotropy and process the data according to standard the image formation procedure. Although such a method produces better quality images compared to narrow-aperture images, they are under-achieving in the sense that the estimates can be severely biased due to the model mismatch, and more importantly, they do not reveal valuable information that azimuthal variations convey.

■ 1.1 Premise

This thesis is based on the premise that scatterer anisotropy is present and exploitable in SAR. Not only is scatterer anisotropy predicted by scattering physics[48, 61], but it has also been observed in real data[2, 12, 51]. Neglecting anisotropy naturally leads to suboptimal use of the SAR data due to the model mismatch as demonstrated by Yeang *et. al.*[61].

Scatterer anisotropy need not degrade SAR quality. In fact, it can offer significant improvements to SAR imaging and ATR. From an imaging standpoint, a better scattering model allows for more accurate reflectivity estimates since the weighting in the image formation process can be adjusted according to the azimuthal scattering pattern (e.g. with a matched filter). Knowledge of anisotropy may also be used to better suppress interfering scatterers in super-resolution image formation algorithms. In addition to improving reflectivity estimation, knowledge of the anisotropy of a scatterer is useful as a feature in and of itself. For a canonical scatterer, one can infer properties about its geometric shape and size from its anisotropy. Knowledge of the geometry of the individual scatterers on an object may then be considered in tandem to aid the classification of the target under investigation. Anisotropy information may also be useful to characterize the stability of a scatterer. In particular, the specular response from an anisotropic canonical scatterer appears over a smaller range in azimuth leading to an observability that is more sensitive to azimuthal orientation. Knowledge of this anisotropy however allows such sensitivity to be incorporated into algorithms such as peak matching which should improve ATR performance.

■ 1.2 Previous Efforts

Previous efforts to deal with scatterer anisotropy in SAR can loosely be divided into two categories: those which are based on precise canonical scattering models and those based on the more general sub-aperture scattering models. Physical optics (PO)[48] and the geometric theory of diffraction (GTD)[33] give accurate predictions of the scatterer response for what are called *canonical* or *atomic* scatterers. These are scattering primitives which have very simple geometries and include as examples the flat plate, cylinder, sphere, top-hat, dihedral, and trihedral.

Canonical scattering models are a natural tool for exploiting scatterer anisotropy due to their precise modeling. Their application in SAR has generally fallen into one of two classes. The first is their use to estimate the canonical type and the associated parameters of scatterers in SAR data. Moses, Potter, and Chiang take such an approach where they use GTD models. Potter and Moses[45], use a maximum likelihood framework to estimate the scatterer type and associated parameters from image data. This work is extended by Chiang[13] to evaluate the improvement in SAR ATR performance that such estimates could provide. McClure and Carin[40] have also used canonical models to estimate the scatterer type and associated parameters. In their approach, they populate a "dictionary" of basis functions with the GTD responses for

various scatterers of interest in addition to other families of models[1]. They use this dictionary of basis functions in a matching-pursuit framework to decompose SAR data into constituent scatterers. In the mentioned approaches, very high fidelity data is necessary for accurate estimation. This is required because over a single image, it is often difficult to differentiate responses from different geometrical configurations. Performance degrades even more so in the presence of interfering scatterers. Richards[46] addresses this problem in a context in which many images from various orientations are available such as in an automated target model generation framework. He uses the iterative expectation-maximization algorithm to perform scatterer correspondence among the various images and to estimate the parameters of each of the scatterers. The use of a large amount of observation data in this approach successfully allows for accurate estimation of the scatterers.

The success of the approaches mentioned in the previous paragraph depends heavily on the availability of large amounts of data or high fidelity data with little interference among scatterers. However, it is often the task in ATR to classify an object using only one or a few images which contain many scatterers. This motivates the other, less ambitious, use of canonical models which is to use them to improve the images of reflectivity estimates. Benitz[4, 7, 5] takes this approach by employing canonical scatterer models in super-resolution methods. In particular, he uses a canonical scatterer response as the model in a variant on Capon's method and in the MUSIC algorithm. The result is an image of reflectivities specifically tuned for the model scatterer. Because the image is tuned for a particular scatterer, this imaging process is performed for a series of pre-specified scatterer geometries thus giving rise to a vector of images or what Benitz calls high definition vector imaging (HDVI).

Chaney et. al.[12] and Allen et. al.[2] follow approaches similar in spirit to Benitz's methods but with some significant differences. Allen and Chaney are both concerned with low-frequency SAR which has a lower resolution where only large scatterers, measuring on the order of a meter or more, are apparent in the imagery. For a vehicle, these scatterers are typically the broadside response associated with the front, back, or side of the target. However, tree trunks also produce strong responses. Allen uses the differences in azimuthal dependencies between the flat plate and the cylindrical tree trunk to differentiate the two. He does so using a bank of approximate matched filters to generate reflectivity estimates. The matched filters are approximate in the sense that they are based on *sub-aperture scattering models*, i.e. models which assume the azimuthal response is uniform over a section of the aperture and zero elsewhere. In particular, he uses a sequence of overlapping sub-apertures each of which has the same length. Allen uses the resulting sub-aperture reflectivity estimates in two ways. First, to distinguish a vehicle's broadside response from a tree trunk's isotropic response, Allen uses the maximum sub-aperture reflectivity to characterize the scatterer. Although the flat plate and tree trunk produce similar reflectivities when the entire aperture is used,

---

[1]Besides the canonical scatterer models, McClure and Carin use resonance models to account for cavities and chirp models to account for dispersive structures.

the flat plate produces a very directive response resulting in a very strong response concentrated in a few of the sub-apertures, whereas the tree trunk produces an isotropic response resulting in a constant but smaller reflectivity among all the sub-apertures. Thus, the vehicle appears significantly brighter than trees in Allen's image of maximum sub-aperture reflectivities. Allen also uses the sub-aperture reflectivities to estimate the azimuthal location and width of the directive flat plate response. He does so by plotting the sub-aperture reflectivities as the sub-apertures transition from one end of the aperture to the other thus displaying how the response varies in azimuth. The characteristics of the directive response are then measured from this plot.

Chaney independently takes an approach similar to Allen's. In particular, Chaney forms a series of sub-aperture reflectivity estimates and chooses the sub-aperture producing the maximum estimate. Doing this for every pixel, Chaney obtains an image of matched reflectivity estimates. The most significant difference between Allen's and Chaney's approaches is that Chaney varies the size the sub-aperture to adapt to the azimuthal width of the response, whereas Allen uses a fixed pre-specified azimuthal width. Varying the size of the sub-aperture, however, also produces reflectivity estimates with different spatial resolutions. Thus, to limit the artifacts that such changes may produce in the image, Chaney also regularizes the sub-aperture selection to limit the fluctuations as one moves throughout the image.

There are two primary differences between the approach taken by Benitz and those taken by Allen and Chaney. The first is that Allen and Chaney use the simpler sub-aperture scattering model and are thus not overwhelmed by the myriad images they need to generate. They are only concerned with the gross-level scatterer anisotropy which one can observe. The second difference is that Benitz provides a vector of images (one image of estimates under each scatterer model), whereas Allen and Chaney provide a single image (choosing a single estimate from among all the models). The latter is more amenable to direct inclusion into existing image based ATR algorithms, but the fashion in which the estimate is chosen leads to sub-optimal resolution. In particular, Allen uses sub-apertures of a pre-specified length, thus imposing a fixed resolution which is not adapted to individual scatterers. Chaney, in contrast, uses sub-apertures of varying lengths, and, for each pixel, selects the sub-aperture which produces the maximum reflectivity estimate. This approach, however, leads to an excessively coarse resolution because the method neglects the higher noise variance in smaller sub-aperture estimates. In particular because the sub-aperture reflectivity estimate from too small an aperture has an unbiased error, the higher noise variance increases the probability that the maximum reflectivity is produced in one of the smaller sub-apertures.

Smith[51] takes an approach similar to Allen's by dividing the full-aperture into overlapping sub-apertures of constant length. However, instead of looking to detect and measure anisotropy from canonical scatterers, Smith's goal is to characterize the stability of the scatterer. In particular, for each peak location determined through a separate peak extraction step, a number is assigned to the scatterer characterizing the fluctuations of the sub-aperture reflectivities. This peak stability is then used as a

confidence measure for that scatterer in a peak matching classification algorithm. This stability weighting is motivated by the idea that more stable peaks (i.e. those which exhibit less azimuthal variation) should be weighted more heavily since they are more likely to be matched in another image of the target. Although this thesis argues that anisotropy is useful and should not be discarded, Smith's approach is reasonable in its context of application to existing ATR peak matching algorithms which use standard SAR image formation based on the isotropic scattering assumption.

Subotic *et. al.*[53] have also worked with sub-aperture models, but not in the context of anisotropy. In his work, he examines the reflectivities produced by a nested sequence of sub-apertures. Treating the sequence as a multi-scale process, he then examines several statistical tests for the detection of *isotropic* scatterers. Subotic's contribution is significant because of its specific focus on how the varying size of the resolution cell impacts the statistical behavior of the reflectivity estimates. All of the other methods mentioned in this section do not address this issue with the exception of Chaney who addresses it indirectly by regularizing the sub-aperture selection process.

Irving *et. al.*[31] has also proposed modeling SAR data as a multi-scale stochastic process. Instead of varying the size of the sub-aperture to change the resolution, Irving *coherently* averages 2×2 pixel blocks in a recursive fashion to generate a quad-tree representation of the SAR data. He then uses an auto-regressive model in scale to describe the behavior of the *log-magnitude* of the reflectivities on the quad-tree. This is motivated by the idea that clutter is composed of many small independent scatterers and thus should have little scale-to-scale dependence because the random phases on each of the scatterers cause them to interfere with each other in a chaotic fashion as resolution is varied. In contrast, larger scattering objects, such as trees or targets, should demonstrate strong scale-to-scale dependence as they share a common phase over the scatterer and thus constructively interfere to produce a stronger response as scale is coarsened. Multi-scale modeling is thus used by Irving to differentiate anomalous man-made targets from natural clutter according to their scale auto-regressive behavior. This idea is also used for terrain segmentation by Fosgate *et. al.*[24] where grass and forest terrain are differentiated based on their respective multi-scale dependencies. Fosgate also proposes a multi-scale anomaly detector to detect the extremely high scale-to-scale correlation that is characteristic of man-made targets. Kim *et. al.*[36] enhances the segmentation algorithm of Fosgate by incorporating model deviations and subsequently develops a segmentation-based compression algorithm exploiting the revealed dependencies for each terrain class.

■ 1.3 Thesis Objective

The results of both Allen and Chaney demonstrate that anisotropy can be used to improve ATR performance on acquired SAR data. This thesis builds upon their sub-aperture analysis to develop robust methods to exploit scatterer anisotropy in ATR. In particular, we present a more robust representation of the data and a statistical

framework which improves upon that of Allen and Chaney by providing more accurate anisotropy characterizations and higher resolution imagery.

The first main issue to be addressed in this thesis is that of data representation. While it is theoretically true that any invertible transformation of the data is equally suitable for subsequent estimation, the packaging of the data significantly impacts the clarity and practical implementation of the estimation. For the case of SAR, there is a fundamental trade-off in being able to obtain measurements with high azimuthal resolution (i.e. the ability to observe azimuthal variations) and high imaging resolution. The methods mentioned in the previous section do not directly address this issue and thus suffer from this trade-off as they must exchange one form of resolution for the other. However, this thesis proposes a multi-resolution pyramid representation of the SAR data at a variety of resolution trade-offs that permits a decision-directed approach which produces estimates with *both* high imaging *and* azimuthal resolution.

Having addressed the first issue of data representation, this thesis proceeds to its primary goal of exploiting anisotropy for ATR. Even with the sub-aperture representation in place, there are still issues which require consideration. First, there is the issue of how anisotropy is to be exploited to aid ATR. Two approaches are taken towards this end. One is to apply machine learning to solve the classification problem. In this approach, the anisotropy dependencies conditioned on a particular target classes are learned via a nonparametric estimator. The learned densities are then directly used to implement a classifier. The drawback to this approach is that it limits the degree to which physical modeling and intuition can be leveraged to analyze and improve the classifier. In particular, the learned densities are driven by the data and do not convey obvious physical information about the target. Thus, the role which an ATR analyst can play is limited. This motivates the second approach to improving ATR which extends the efforts of Allen and Chaney. In particular, the sub-aperture scattering model is used to describe the concentration of scattering in azimuth. An optimal sub-aperture is chosen and subsequently used to provide an enhanced reflectivity estimate. In addition to the sub-aperture reflectivity estimates, the specifics about the chosen sub-aperture are also provided to convey information about the geometry of the scatterer. The efforts of Allen and Chaney are extended in three significant ways. The first is that their model for the sub-aperture measurements is improved upon thereby producing a more accurate test. The second improvement is that the structure of the sub-aperture pyramid is leveraged to obtain both high imaging and high azimuthal resolution through the use of a decision-directed procedure exploiting the presence of multiple resolutions. The third extension is that the effects of neighboring scatterers are included in our analysis. Addressing the issue of interfering scatterers is essential as they corrupt the observed azimuthal response resulting in anisotropy misclassifications and, in effect, reduced spatial resolution when ignored.

■ 1.4 Thesis Organization

The remainder of this thesis addresses these issues and is organized as follows.

■ 1.4.1 Background

Chapter 2 provides the necessary background in SAR and machine learning needed to address the issues of concern in this thesis. The chapter begins by discussing the fundamentals of SAR. Such details are important in that the processing used to achieve SAR's high resolution are strongly based on the ideal point-scattering model which ignores scatterer anisotropy. SAR image formation is also presented from a beamforming perspective which allows the inclusion of models other than the ideal point scatterer model. The chapter concludes with a description of nonparametric density estimation and how it can be used for classification.

■ 1.4.2 Sub-aperture Analysis

Chapter 3 introduces the sub-aperture pyramid which serves as the fundamental structure for all of the analysis in this thesis. In this structure, the sub-apertures are arranged in a pyramidal fashion according to their size and azimuthal location. The sub-aperture pyramid is used both to estimate reflectivity and define the models for anisotropy. The framework can be interpreted in a multi-scale setting, where each scale provides a different azimuthal versus spatial resolution trade-off. This structure is also viewed in the context of other works. In particular, it can be seen as

- a multiple-basis representation offering different tilings of the azimuthal-image plane[2] à la wavelet packet dictionaries[59].

- an extension of the sub-aperture work by Allen and Chaney,

- an instance of the vector imaging work by Benitz,

■ 1.4.3 Nonparametric Estimation of Anisotropy

Chapter 4 presents an extension of the nonparametric approach of DeBonet *et. al.*[20]. The fundamental change is the use of the sub-aperture pyramid motivated by scattering physics in contrast to steerable wavelets motivated by visual attributes. Because the sub-aperture scattering model approximates the aspect dependent scattering produced by canonical scatterers, the sub-aperture pyramid captures such responses in a clear manner. The advantage of this machine learning approach is that data dependencies are automatically learned with a nonparametric estimator. However, there are several issues concerning the implementation of this nonparametric approach which are addressed in this chapter. Classification performance is measured and analyzed with the MSTAR[1] data set.

---

[2]This is traditionally called the time-frequency plane in signal processing, but is not done so here to prevent confusion with SAR terminology.

■ 1.4.4 Sub-aperture Anisotropy Attribution

Chapter 5 deals with the topic of attributing a given spatial pixel with an anisotropy la-
bel and associated reflectivity estimate. Allen and Chaney choose the sub-aperture with
maximum reflectivity because of its matched-filter interpretation. However, Allen im-
poses a pre-specified sub-aperture size and thus cannot adapt the sub-aperture to match
the width of the azimuthal response. Chaney, in contrast, varies the sub-aperture size,
but he does not account for the resulting change in the sub-aperture noise variance.
Based on the sub-aperture pyramid, a consistent statistical model and test for sub-
aperture anisotropy are presented. The result is an intuitively simple test providing
better imaging resolution and more accurate anisotropy labels than Allen and Chaney.
The anisotropy characterization also reveals why the sub-aperture selection used by
Allen and Chaney leads to an excessively coarse imaging resolution. Several modifi-
cations on the proposed anisotropy test are discussed. The first modification changes
the estimation of an auxiliary parameter to address the lack of an explicit model for
interfering scatterers which distort the azimuthal response and hence the classification.
The second modification also addresses the issue of interfering scatterers but does so
through a telescopic decision-directed testing procedure on the sub-aperture pyramid
which also reduces computational complexity. The third modification addresses devi-
ations from the sub-aperture scattering models through a model perturbation term.
Results demonstrating the utility provided by the anisotropy attribution are provided
on the MSTAR data set.

■ 1.4.5 Anisotropy Phenomenology

The results on MSTAR data at the end of Chapter 5 strongly argue for two different
sources of anisotropy. The study of these two modes is the topic of Chapter 6. These
two classes are called geometric and volumetric scatterers. *Geometric scatterers* are
large structured scatterers such as canonical scatterers and scatterers with slight devia-
tions from canonicity. *Volumetric scatterers* on the other hand are actually a collection
of many unresolvable scatterers which happen to coherently combine to produce an
anisotropic response over the observed azimuth. Not surprisingly, anisotropy from vol-
umetric scatterers is not stable due to the chaotic nature of its azimuthal response, and
thus it provides no utility to ATR. Anisotropy from geometric scatterers is however very
stable and provides significant information about the underlying scatterer, and thus it
provides significant utility for ATR. Unfortunately, it is difficult, if not impossible, to
determine the source of anisotropy based on a single observation from moderate reso-
lution (i.e. moderate aperture) SAR. However, the use of wide-aperture SAR decreases
the incidence of anisotropy from volumetric scatterers by utilizing the increased range
of azimuthal data to assess the stability[3] of the azimuthal response.

---

[3]This notion of azimuthal stability is more general than that of Smith's[51] who considers devia-
tions from the point scattering model as compared to deviations from canonical anisotropic responses
considered in this thesis.

■ 1.4.6 Iterative Anisotropy Attribution

Chapter 7 returns to the topic of anisotropy attribution by improving the modeling and estimation proposed in Chapter 5. In particular, this chapter extends the modeling of neighboring scatterers by allowing the consideration of anisotropic neighbors. Rather than attempting the monumental task of simultaneously estimating the anisotropy of all the pixels, an iterative approach is taken in which the azimuthal contributions from neighboring scatterers is modeled using the anisotropy classification from the previous iteration. Special care must be taken in using the information from the previous iteration so as to appropriately account for the contributions from neighboring scatterers. In particular, a distinction must be made between pixels and scatterers. Consider for example, removing the azimuthal distortion produced by a physically large interfering scatterer that spans several, say N, pixels. If one were to simply subtract off its effect for every pixel, then that would be equivalent to saying that there are N scatterers exhibiting this response (one for each of the pixel locations) instead of the one underlying scatterer. Thus, the modeling of neighboring scatterers must be carefully taken into account in order to prevent over-compensating for their azimuthal distortion. While it is necessary to remove the interference from neighboring scatterers, the effects from on-scatterer pixels should be preserved. *On-scatterer pixels* are pixels corresponding to the same scatterer as the pixel being examined. If their contribution is removed, then another form of distortion is introduced which cancels part of the response from the scatterer in question thus corrupting the anisotropy classification. In order to address these issues, an additional processing step is used which aggregates pixels into their underlying scatterers. The segmentation algorithm used is based on a greedy maximum-likelihood algorithm. The segmentation is then used to distinguish off-scatterer pixels from on-scatterer pixels so that only the contribution from interfering scatterers are removed. The chapter concludes with a presentation of experimental results demonstrating the performance of both the anisotropy attribution and scatterer segmentation.

■ 1.4.7 Conclusions

The thesis concludes with Chapter 8. A summary of the contributions in this thesis is first provided. This is followed by a discussion of opportunities for future research utilizing or extending the anisotropy analysis we have developed.

Chapter 2

Background

The work in this thesis focuses on the development of techniques that exploit non-ideal scattering in synthetic aperture radar (SAR) imagery for the purpose of target recognition. In this background chapter, the fundamentals describing relevant areas of SAR and machine learning are presented. These are broad topics incapable of being summarized in a single chapter. Focus is thus restricted to those aspects which are most pertinent to the work in this thesis. Section 2.1 presents the fundamentals of conventional SAR image formation. Section 2.2 describes SAR from a spectral estimation perspective motivated by array processing which helps provide intuition about SAR. Section 2.3 discusses the issue of nonideal scattering along with some current efforts to address it. Section 2.4 concludes the chapter with a discussion of nonparametric density estimation and how it can be used for the problem of classification. A footnote for each sub-section heading indicates the bibliographical sources from which the material is obtained and where more details may be found.

■ 2.1 Conventional SAR

Active radar imaging systems are frequently used in remote sensing and surveillance applications. SAR is a special case of microwave radar imaging that uses sophisticated signal processing to achieve a high imaging resolution. In the SAR scenario, a radar is mounted on an airborne or spaceborne craft where it illuminates the region of interest and measures the corresponding reflectivity. Using a microwave band radar system has several advantages. Since it is an active imaging system, it provides its own illumination in the form of microwave radiation, thus allowing it to be used for both day and night operation. Furthermore, because microwave radiation is not significantly attenuated by atmospheric components such as common gases, dense fog, or cloud cover, microwave-frequency radar systems produce images that are unimpaired by such natural phenomena.

The radar imaging scenario is illustrated in Figure 2.1 along with some commonly used terms. The *footprint* refers to the ground area illuminated by the radar and represents the region for which the reflectivity is being measured. The *target field* is the cumulative area covered by the footprint as the radar is moved. In the case of *spotlight SAR*, the radar is continuously steered such that the area of illumination is constant,

in which case, the target field and the footprint coincide. In the case of *stripmap SAR*, the position of the radar is held fixed with respect to the platform, and the target field corresponds to the swath of terrain covered. We will not concern ourselves with the differences between these two modes of SAR as the underlying principles discussed are the same for both cases. *Cross-range* or *azimuth* refers to the direction along the ground which is parallel to the radar's trajectory. Although these refer to the same direction, the term azimuth is generally reserved for discussion of sensor measurements, and the term cross-range is generally reserved for discussion of the target field or image. *Slant-range* is the direction from the radar to center of the footprint. Slant-range is also frequently referred to as *down-range* or simply *range*. The *elevation angle* is the angle between the normal of the ground and the slant-range. The projection of the slant-range onto the ground is termed the *ground-range*.
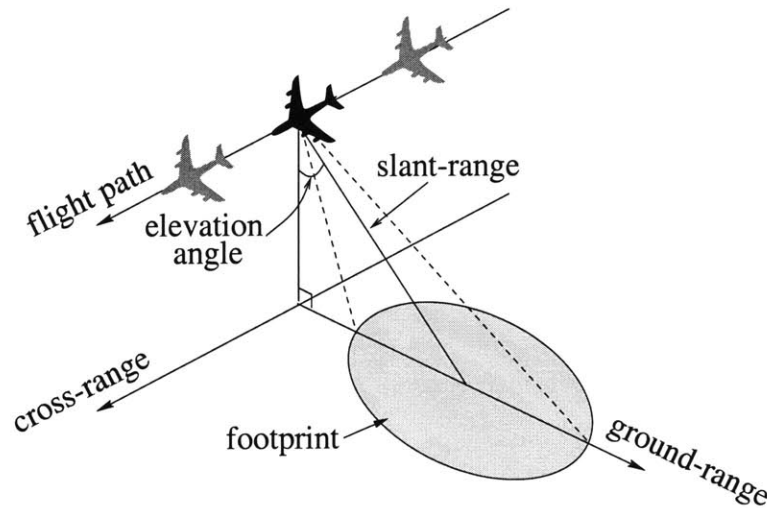
Figure 2.1. Terminology used in SAR imaging.

For each cross-range location, the reflectivity of objects at varying down-range locations is estimated by transmitting a reference signal and measuring the gain and delay of the return signal. The time delay provides the scatterer's range, and the gain provides its reflectivity. Thus, the coordinates in the resulting radar image are slant-range and cross-range as illustrated in Figure 2.2(a). Objects not in this imaging plane are projected onto it with their slant-range distance being reported as the distance from the radar. This imaging plane is "opposite" to that of photographical images where the imaging plane is the plane perpendicular to the wavefront of the transmitted signal, i.e. the coordinates are cross-range and the other orthogonal direction to slant-range as illustrated in Figure 2.2(b).

Without sophisticated signal processing, the range resolution is determined by the pulse width of the transmitted signal. A rectangular pulse of duration T_p, for exam-

(a)   (b)
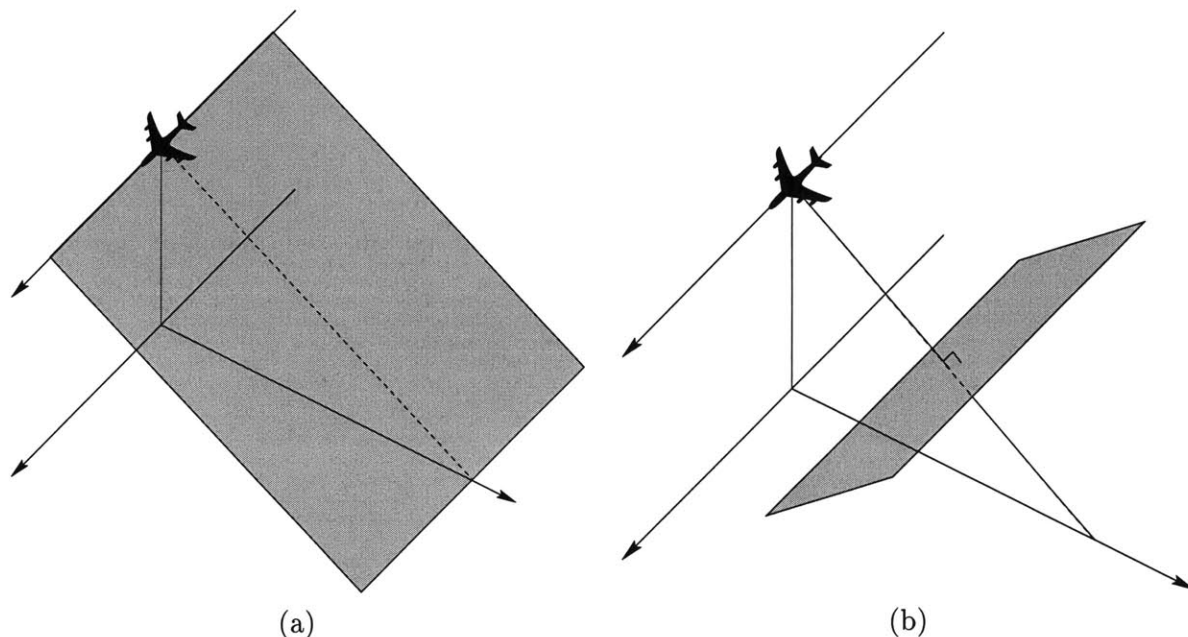
Figure 2.2. Illustration of the imaging planes (shaded region) for (a) SAR and (b) photographical imagery.

ple, would have a range resolution of cT_p where c is the speed of light. Ideally, one would transmit an arbitrarily short pulse signal to achieve a corresponding high range resolution; however, in order to maintain a high SNR on the reflected signal for a sufficient level of detection, the amplitude of the pulse must be increased as its duration is shortened. Practical limitations on equipment restrict the signal that can be produced, and in particular, limit the amplitude that can be generated for such a narrow pulse. However, by using the technique of range compression, which is standard in microwave imaging systems, the reflection from transmitting a high-energy long-duration pulse may be processed so that the resulting resolution is much finer than that corresponding to the duration of the transmitted pulse. This permits both a high SNR and a high range resolution without having to generate a high amplitude short duration pulse.

A drawback of microwave imaging is that with straightforward techniques, the cross-range resolution is poor due to the long wavelength of microwave radiation relative to other sources such as light. The cross-range resolution is dictated by the width of the illuminated radar footprint which is proportional to both the distance from the radar and the wavelength of the signal and is inversely proportional to the size of the radar antenna. For practical scenarios, this resolution can be on the order of several kilometers. SAR overcomes this limit by fusing together multiple measurements taken at different azimuthal locations to synthesize a longer antenna aperture and thus reduce the width of the radar footprint. It is this processing in cross-range that differentiates

SAR from other microwave imaging systems. In the remainder of this section, the techniques used to achieve high range and cross-range resolution are described.

■ 2.1.1 Matched Filter Range Compression [1]

This section explains the use of the matched filter for range compression. As previously mentioned, range compression is a technique that allows the use of a long duration transmission signal which simultaneously provides high range resolution and high SNR. The matched filter is used because it is the optimal filter for detecting a signal in the presence of additive white Gaussian noise. The signal we seek to detect is a time delayed version of the one transmitted where the delay reflects the distance from the scattering object. Consider transmitting the following linear frequency modulated (FM) chirp signal

$$s(t) \triangleq \operatorname{rect}\left(\frac{t}{T_p}\right) \operatorname{Re}\left\{ e^{j2\pi(f_c t + \kappa t^2/2)} \right\} = \operatorname{rect}\left(\frac{t}{T_p}\right) \cos\left(2\pi(f_c t + \kappa t^2/2)\right) \qquad (2.1)$$

where f_c is the *carrier frequency*, κ is the *frequency rate*, T_p is the pulse duration, and

$$\operatorname{rect}(t) = \begin{cases} 1 & |t| \le 1/2 \\ 0 & \text{otherwise.} \end{cases}$$

The phase of the transmitted signal in Eq. (2.1) is

$$\phi(t) = 2\pi(f_c t + \kappa t^2/2)$$

giving an instantaneous frequency of

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}\phi(t) = f_c + \kappa t.$$

Thus, with a duration of T_p, the bandwidth of the signal is $B = |\kappa| T_p$ since the frequency varies from $f_c - \kappa T_p/2$ to $f_c + \kappa T_p/2$ as t varies from $-T_p/2$ to $T_p/2$. The Fourier Transform for $s(t)$ is difficult to compute in closed form, but for large time-bandwidth products BT_p, the following approximation can be used [17]

$$S(f) = \begin{cases} \frac{1}{\sqrt{|\kappa|}} e^{j\,\operatorname{sign}(\kappa)\,\pi/4}\, e^{-j\pi(f-f_c)^2/\kappa} & |f - f_c| < B/2 \\ \frac{1}{\sqrt{|\kappa|}} e^{-j\,\operatorname{sign}(\kappa)\,\pi/4}\, e^{j\pi(f-f_c)^2/\kappa} & |f + f_c| < B/2 \\ 0 & \text{otherwise,} \end{cases}$$

i.e. it is approximately an ideal bandpass signal with quadratic phase.

---

[1] Primary sources: Wehner[58] and Curlander[17].

Consider an ideal point scatterer with reflectivity A at a range R from the radar. For a point scatterer, the reflected signal is a time delay of the transmitted signal where the delay is given by the round-trip propagation time, i.e.

$$v_r = As(t - 2R/c). \tag{2.2}$$

Using the matched filter $h(t) = s^*(-t)$ with frequency response $H(f) = S^*(f)$, the output in the frequency domain is[2]

$$
\begin{aligned}
G(f) &= H(f)V_r(f) \\
&= AS^*(f)S(f)\,e^{-j2\pi 2R/c} \\
&= \begin{cases} A\,e^{-j4\pi R/c} & |f - f_c| < B/2,\ |f + f_c| < B/2 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

which in the time domain corresponds to

$$
\begin{aligned}
g(t) &= A\int_{-f_c-B/2}^{-f_c+B/2} e^{-j4\pi R/c}e^{j2\pi ft}\,df + A\int_{f_c-B/2}^{f_c+B/2} e^{-j4\pi R/c}e^{j2\pi ft}\,df \\
&= 2AB\,\cos(2\pi f_c(t - 2R/c))\,\frac{\sin(\pi B(t - 2R/c))}{\pi B(t - 2R/c)} \\
&= 2AB\,\cos(2\pi f_c(t - 2R/c))\,\operatorname{sinc}(B(t - 2R/c))
\end{aligned}
$$

The $\cos(\cdot)$ term is a high frequency modulation and the $\operatorname{sinc}(\cdot)$ term determines the range focal point and resolution. In particular, the envelope of the matched filter output is a sinc function centered at $t = 2R/c$ which corresponds to a scatterer at a range of R. The first nulls of the sinc envelope occur at $t = 2R/c \pm 1/B$. Thus, if the range resolution is taken to be the distance from the peak to the first null, the resolution is

$$\delta_r = \frac{c}{2B}. \tag{2.3}$$

If the resolution is alternatively defined in terms of the half-power or 3dB width, then taking the distance between the half-power points of the $\operatorname{sinc}(\cdot)$ function, the resolution is

$$\delta_{r\,3\text{dB}} \approx \frac{0.44c}{B}. \tag{2.4}$$

■ 2.1.2 Matched Filter Cross-Range Compression [3]

The use of range compression is standard in microwave radar imaging and allows for a high range resolution. Cross-range resolution however is unaffected as it is governed by the azimuthal width of the radar footprint in cross-range which is

$$\delta_{\text{footprint}} = \frac{\lambda R}{L_p} \tag{2.5}$$

---

[2]The scaling constant $|\kappa|$ has been dropped since it is just a constant gain term.

[3]Primary sources: Munson[43] and Curlander[17].

where λ is the wavelength of the transmitted signal and L_p is the physical length of the aperture. For many scenarios, this corresponds to a resolution on the order of several kilometers. The novelty of SAR is in its fusing multiple azimuthal measurements to synthesize a larger aperture and thus achieve a higher cross-range resolution. The relative displacement between the moving radar platform and the stationary target field gives rise to a chirp-like term in cross-range allowing range compression to be performed in this dimension as well. The situation is illustrated in Figure 2.3 where an airborne SAR makes regularly spaced measurements illuminating all points of interest as it follows a linear path. Because the velocity of the platform is negligible compared to that of the transmitted pulse, the radar can be considered to be stationary for the measurement taken at each $y = n\Delta$ for integers n.
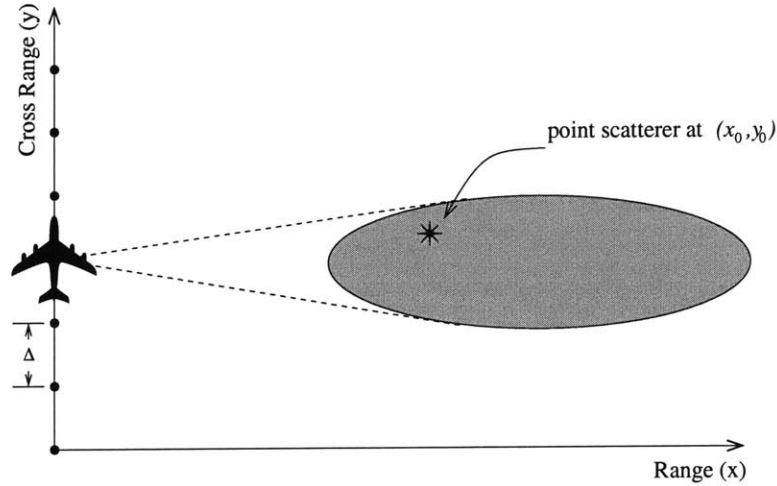
Figure 2.3. Multiple cross-range sensor measurements of the target field.

Again, consider transmitting the chirp signal in Eq. (2.1). Now, consider a point scatterer with reflectivity A at location (x_0, y_0) in down-range and cross-range respectively. The received signal corresponding to the measurement taken at $y = n\Delta$ is

$$v_r(t, n\Delta) = As\left(t - \frac{2R(x_0, y_0 - n\Delta)}{c}\right)$$

where $R(x_0, y_0 - n\Delta)$ is the distance from the nth sensor at $(0, n\Delta)$ to the scatterer at (x_0, y_0). Both n and t convey time but on different scales. In particular, the variable t is associated with the propagation of the transmitted signal which travels at the speed of light, whereas n is associated with the movement of the radar which travels at the speed of the aircraft or spacecraft. Because the duration over which t is measured is much shorter than that for n, t is called *fast time* and n is called *slow time*.

Performing the range compression in fast time with the matched filter $h_r(t) = s^*(-t)$ on the nth sensor measurement gives the output in the frequency domain as

$$G_r(f, n) = H(f)V_r(f, n\Delta)$$

$$= AS^*(f)S(f)\, e^{-j2\pi 2R(x_0, y_0 - n\Delta)/c}$$

$$= \begin{cases} A\, e^{-j4\pi R(x_0, y_0 - n\Delta)/c} & |f - f_c| < B/2,\ |f + f_c| < B/2 \\ 0 & \text{otherwise} \end{cases}$$

which corresponds in the time domain to

$$g_r(t, n) = A \int_{-f_c - B/2}^{-f_c + B/2} e^{-j4\pi R(x_0, y_0 - n\Delta)/c} e^{j2\pi f t}\, df + A \int_{f_c - B/2}^{f_c + B/2} e^{-j4\pi R(x_0, y_0 - n\Delta)/c} e^{j2\pi f t}\, df$$

$$= 2AB \cos\left(2\pi f_c \left[t - \frac{2R(x_0, y_0 - n\Delta)}{c} \right] \right)\, \text{sinc}\left(B \left[t - \frac{2R(x_0, y_0 - n\Delta)}{c} \right] \right)$$

$$= 2AB\, \text{Re}\left\{ e^{j2\pi f_c(t - 2R(x_0, y_0 - n\Delta)/c)} \right\}\, \text{sinc}\left(B \left[t - \frac{2R(x_0, y_0 - n\Delta)}{c} \right] \right).$$

Again, the matched filter output has the form of a modulated sinc pulse with mainlobe width $c/2B$ in down-range. In SAR imaging situations, the azimuthal extent of the footprint is narrow relative to the range x_0. Thus, the range R in the slowly varying sinc expression can be approximated[4] as x_0[42]. Such a coarse approximation is not sufficient for the rapidly varying modulation term where deviations of a fraction of a wavelength can have a profound effect. For this term, a first order approximation can be used[5][42]

$$R(x_0, y_0 - n\Delta) = \sqrt{x_0^2 + (y_0 - n\Delta)^2}$$

$$\approx x_0 + \frac{(y_0 - n\Delta)^2}{2x_0}.$$

Substituting these approximations for R in $g_r(t, n)$ yields

$$g_r(t, n) \approx 2AB\, \text{Re}\left\{ e^{j2\pi f_c\left(t - \frac{2x_0 + (y_0 - n\Delta)^2/x_0}{c} \right)} \right\}\, \text{sinc}(B(t - 2x_0/c))$$

$$= 2AB\, \text{Re}\left\{ e^{j2\pi f_c(t - 2x_0/c)} e^{-j2\pi f_c(y_0 - n\Delta)^2/(x_0 c)} \right\}\, \text{sinc}(B(t - 2x_0/c)).$$

Due to the sinc contribution, the response $g_r(t, n)$ is significant only when t is approximately equal to $2x_0/c$, i.e. when focused on the scatterer in down-range. For such a value of t, the approximation to $g_r(t, n)$ is

$$g_r(2x_0/c, n) \approx 2AB\, \text{Re}\left\{ e^{-j2\pi f_c(n\Delta - y_0)^2/(x_0 c)} \right\}$$

---

[4]When this approximation does not hold, there is said to be *range walk* and special processing needs to be performed to compensate for it.

[5]Again, the assumption of $x_0 \gg |y_0 - n\Delta|$ is necessary for the sufficiency of this approximation.

which is similar in form to a delayed version of the transmitted chirp in Eq. (2.1) with the time variable denoted as $n\Delta$, a carrier frequency of 0, and a frequency rate of $-2f_c/(x_0 c)$. A rect($\cdot$) function is absent from the above equation for $g_r(2x_0/c, n)$, so instead of processing it with a matched filter (spanning all n), it is cross correlated with a windowed version of the chirp signal. This filter is

$$h_a(n; x_0) = \text{Re}\left\{ e^{j2\pi f_c n^2/(x_0 c)} \right\} \text{rect}(n\Delta/L_a)$$

where L_a is the length of the synthetic aperture and implicitly determines the number of cross-range measurements used since they are assumed to be uniformly spaced. Note that the filter is parameterized with the range location, in addition to slow-time, because the frequency rate of the chirp in the filter varies with range to correctly focus in azimuth. Because multiple measurements are used, the effective aperture length L_a is much larger than the physical aperture, and thus the name synthetic aperture radar.

By, approximating the cross correlation of $g_r(2x_0/c, n)$ with $h_a(n; x_0)$ as the autocorrelation of $h_a(n; x_0)$ and subsequently following similar steps as for the range compression, the time domain output becomes

$$g_{r,a}(2x_0/c, n) = \frac{4ABL_a f_c}{x_0 c} \text{sinc}\left(\frac{2L_a f_c}{x_0 c}(n\Delta - y_0) \right).$$

As for the matched filtering in range compression, when the cross-range resolution is taken as the distance between the peak of the sinc and the nearest null, this gives an azimuthal resolution of

$$\delta_a = \frac{cx_0}{2f_c L_a} = \frac{\lambda_c x_0}{2L_a}. \tag{2.6}$$

Alternatively, using the half-power bandwidth as the resolution,

$$\delta_{a\,3\text{dB}} \approx \frac{0.44cx_0}{f_c L_a} = \frac{0.44\lambda_c x_0}{L_a}. \tag{2.7}$$

The expression for resolution in Eq. (2.6) is quite similar to that for conventional radar in Eq. (2.5), but note that Eq. (2.6) uses the synthesized aperture length L_a, which is controlled as a parameter, whereas Eq. (2.5) uses the physical aperture length L_p which is limited by physical constraints.

■ 2.1.3 Deramp Pulse Compression [6]

This section presents a slightly different algorithm called the deramp-FFT method to approximate the matched filter image formation for SAR data. The purpose of its inclusion here is not because of its popularity, but more importantly, because it helps provide a bridge to connect SAR to ideas in spectral estimation for far field array processing. This association is further described in Section 2.2.

---

[6]Primary sources: Curlander[17] and Munson[43].

Again, consider transmitting the chirp $s(t)$ in Eq. (2.1). Performing range compression involves convolving the received signal in Eq. (2.2) with the matched filter $h(t) = s^*(-t)$. However, since the signal $s(t)$ has a frequency which varies with time, efficient linear time-invariant (LTI) processing techniques cannot be used. The idea of deramp-FFT pulse compression is to process the received signal in a computationally efficient fashion while still maintaining the performance of the matched filter. Towards this end, let us start by defining a version of the transmitted signal with an extended pulse duration, i.e.

$$s_e(t) \triangleq \mathrm{Re}\left\{ e^{j2\pi(f_c t + \kappa t^2/2)} \right\} \ \mathrm{rect}(t/T_e) = \cos\left(2\pi(f_c t + \kappa t^2/2)\right) \ \mathrm{rect}(t/T_e)$$

where T_e is the duration which will be specified shortly. This extended chirp is delayed, subsequently multiplied by the received signal $v_r(t)$, and then bandpass filtered to produce $\tilde{v}_r(t)$ as illustrated in Figure 2.4. The delay T_r is an arbitrary, but known, reference and is usually taken to be the round trip delay to the center of target field. Figure 2.5 conveys the intuition behind deramp pulse compression. Both the received and extended signals have a frequency which is linear in time. Because the delay in the extended signal corresponds to a known location, in this case the center of the target field, the position of the scatterer can be inferred from the constant frequency difference between the received and extended signals, i.e. the vertical separation Δf illustrated in Figure 2.5. To detect this difference, the deramp-FFT multiplies the extended reference signal with the received signal. Taking the product of sinusoids results in other sinusoids with frequencies which are the difference and sum of the originals. Thus, keeping only the lower frequency sinusoid reveals the location of the scatterer which is the reason for the mixing and filtering in Figure 2.4. A maximum amount of overlap in time between the received and reference signals would be beneficial because of the increased SNR. The received signal v_r is thus mixed with the extended signal s_e, instead of the transmitted signal s, in order to obtain this maximum overlap in the presence of the unknown delay in the reflected signal.
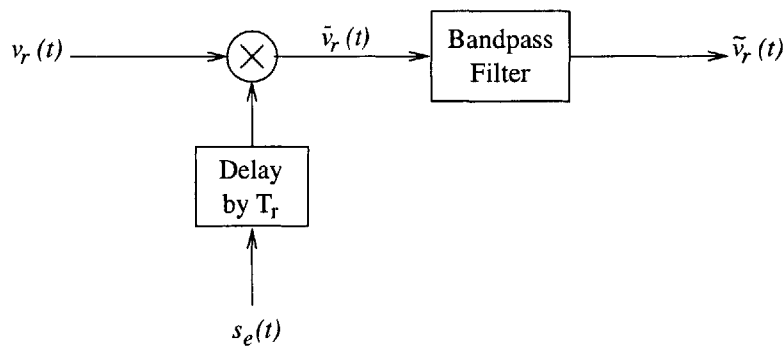
Figure 2.4. Mixing the received signal with a delayed extended chirp.
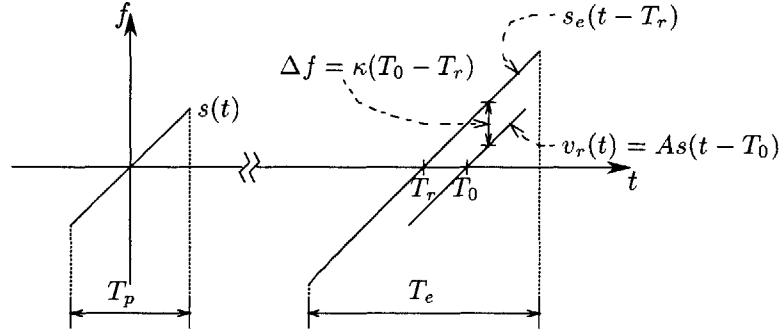
Figure 2.5. Mixing signals to produce a sinusoid whose frequency is proportional to the distance from the center of the footprint.

A more precise description of the deramp-FFT method is now presented. To start, let T_e be sufficiently large such that the support of the delayed and extended chirp $s_e(t - T_r)$ contains the support of the received signal $v_r(t)$ for all scattering locations of interest. Then, the result of the mixing is

$$\bar{v}_r(t) = s_e(t - T_r)s(t - T_0)$$

$$= \begin{cases} A\cos\left(2\pi\left[f_c(t - T_r) + \frac{\kappa(t-T_r)^2}{2}\right]\right)\cos\left(2\pi\left[f_c(t - T_0) + \frac{\kappa(t-T_0)^2}{2}\right]\right) & |t - T_0| < T_p/2 \\ 0 & otherwise \end{cases}$$

$$= \begin{cases} A\cos\left(2\pi\left[f_c(2t - T_r - T_0) + \kappa\frac{2t^2+T_r^2+T_0^2-2t(T_r+T_0)}{2}\right]\right) \\ \quad + \cos\left(2\pi\left[f_c(T_r - T_0) + \kappa\frac{T_r^2-T_0^2+2t(T_0-T_r)}{2}\right]\right) & |t - T_0| < T_p/2 \\ 0 & otherwise. \end{cases}$$

By restricting the range over which a scatterer may contribute to the returned signal, the value of T_0 is bounded. A bandpass filter can thus be specified such that only the lower frequency sinusoid is passed when the scatterer is in the region of interest. Application of such a filter produces the output

$$\tilde{v}_r(t) = \begin{cases} A\cos\left(2\pi\left[f_c(T_r - T_0) + \kappa\frac{T_r^2-T_0^2+2t(T_0-T_r)}{2}\right]\right) & |t - T_0| < T_p/2 \\ 0 & otherwise \end{cases}$$

$$= \begin{cases} A\cos\left(2\pi\left[f_c(T_r - T_0) + \kappa\frac{T_r^2-T_0^2}{2} + \kappa(T_0 - T_r)t\right]\right) & |t - T_0| < T_p/2 \\ 0 & otherwise \end{cases} \tag{2.8}$$

which is called the *deramped signal*. This deramped signal is a constant frequency sinusoid over the reflected pulse width T_p. The first two terms in the argument of the deramped signal's sinusoid are constant phase terms. The third, however, term is a frequency proportional to $T_0 - T_r$, or equivalently, the displacement of the scatterer from the reference location. Thus, knowledge of the frequency is equivalent to knowing

the location of the scatterer. A Fourier transform can thus be applied to $\tilde{v}_r(t)$ to determine the signal strength at each frequency. The precision with which T_0 can be measured using the Fourier transform over the interval T_p is $1/(|\kappa|T_p)$ which translates to a range resolution of $\delta_r = c/(2|\kappa|T_p)$. This is the same as Eq. (2.3) for the matched filter processing. The application of the deramp-FFT processing to cross-range follows identically, and also results in a Fourier transform of a similarly processed signal. Thus, instead of using a time-varying matched filter to obtain reflectivity estimates for each image location, one can take the FFT of a signal which has been mixed and filtered as in Figure 2.4 which significantly reduces the amount of computation required.

■ 2.2 Spectral Estimation and SAR [7]

Recently, there has been much interest in beamforming methods for SAR[37, 3, 4, 7, 5, 11, 16, 21], much of it dealing with super-resolution techniques which try to achieve higher resolutions than those imposed by the Fourier limit in Eqs. (2.3) and (2.6). Beamforming methods for SAR are developed using the deramped-SAR framework where the estimation of the location and reflectivity of a scatterer can be interpreted as a spectral estimation problem.

The discussion in this section concentrates on the spectral estimation approach to SAR image formation using the deramped signal as opposed to the direct matched filtering approach presented in Sections 2.1.1 and 2.1.2. The spectral estimation formulation is equally applicable for both down-range and cross-range processing, and thus no distinction will be made between the two. The spectral estimation framework is based on the far-field array processing scenario illustrated in Figure 2.6(a) for a uniform linear array (ULA) composed of N regularly spaced collinear sensors. Many other array configurations exist[37], but this is one of the simplest and is the one which relates to SAR. In far-field array processing, one considers a radiation source[8] at a distance sufficiently far that the incident wave to the sensor array can be modeled as a plane wave. The problem is then to find the signal strength and angle of arrival of the incident wave which is accomplished by adjusting the weighting coefficients of the sensors in the array to produce an antenna pattern like that illustrated in Figure 2.6(b). In particular, the antenna pattern is produced by coherently combining delayed versions of the received signals where the delay for each sensor increases linearly as the array is traversed to account for the propagation delay. The value of the delay increment between sensors determines the direction of focus for the array as illustrated by ω_0 in Figure 2.6(b). In addition to the sensor delay, which corresponds to linear phase across the sensor array, the magnitude of the gain on each sensor can be adjusted to modify the shape of the antenna pattern produced by the array. Viewed from a signal processing perspective, with a constant magnitude gain on each sensor, the impulse response from a infinitely

---

[7]Primary sources: Krim[37], DeGraaf[21], and Cox[16].

[8]The source itself need not be an active source such as in astronomical settings but may be a passive source whose reflectivity is measured as in radar settings.

long array corresponds to a delta function whose location is determined by the linear phase (or delay increment) of the array. However, since there are only a finite number of sensors, such a filter will exhibit sidelobe behavior due to the "windowing" of the array. To mitigate these sidelobe effects, the magnitudes of the sensors are often adjusted to suppress interference from other directions. The result is a response whose Fourier transform looks like the one illustrated in Figure 2.6(b) which is plotted in polar coordinates instead of rectangular coordinates. The mainlobe is often called the *beam* of the antenna pattern, and this process of designing and directing the beam through adjustment of the sensor coefficients is called *beamforming*.
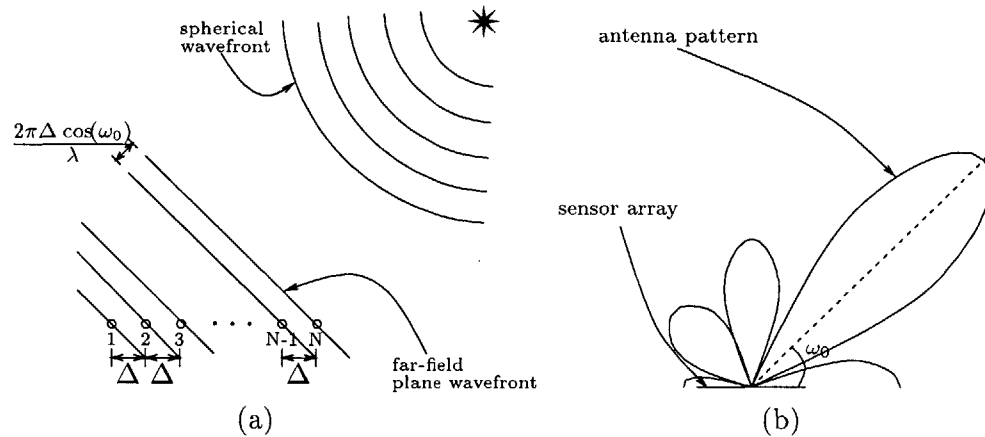
Figure 2.6. Illustration of the array processing scenario for a uniform linear array. (a) Far-field source and incident plane waves. (b) Antenna pattern for a given set of sensor weights.

It should be noted that although the SAR sensor arrangement constitutes a ULA, SAR does not conform to the far-field array processing scenario, and thus beamforming techniques developed for the far-field ULA cannot be directly applied to measured SAR data. The difference between the two situations is that in SAR, the scatterer and radar are in close proximity so that the spherical wavefront cannot be accurately approximated as a plane wave. Thus, instead of detection of a linear phase (or sinusoid component) as in the far-field ULA, one has to detect a quadratic phase in the sensors in SAR. However, techniques developed for far-field array processing can still be applied to *deramped* SAR data for which the quadratic phase has been removed by the mixing operation in Figure 2.4. This mixing operation also has the effect of mapping the entire aperture, which typically has a narrow angular span, to the $180°$ field of view used in the far-field ULA framework.

In presenting beamforming techniques for SAR, the superfluous radar parameters from the deramped signal in Eq. (2.8) are omitted. In particular, the quadrature component of the deramped signal is also assumed to be available, and the phase is incor-

porated into the now complex reflectivity to write the simplified signal as

$$x(t) = A_0 e^{j\omega_0 t} + n(t) \tag{2.9}$$

where ω_0 is the unknown frequency to be estimated (conveying scatterer location), A_0 is the reflective gain from the scatterer, and $n(t)$ is zero-mean circularly complex white Gaussian noise with variance $2\sigma^2$ which represents the response from clutter in the target field. As the measurements are taken at discrete time samples $(t_1, t_2, \ldots, t_N)$, bold face letters are used to denote collections of sampled quantities, e.g. $\boldsymbol{x}$ denotes the vector of all samples of $x(t)$. Since the received data $\boldsymbol{x}$ is a random vector, it has an associated correlation matrix which is denoted as R. The normalized modeled response for a scatterer corresponding to frequency ω_0 is termed the *steering vector* and is denoted by $\boldsymbol{a}(\omega_0)$. In particular, the steering vector for an ideal point scatterer with associated frequency ω_0 is

$$\boldsymbol{a}(\omega_0) = \left[e^{j\omega_0 t_0}, \ e^{j\omega_0 t_1}, \ e^{j\omega_0 t_2}, \ldots \right]. \tag{2.10}$$

This allows the samples of $x(t)$ to be written as

$$\boldsymbol{x} = A_0 \boldsymbol{a}(\omega_0) + \boldsymbol{n}$$

for the case of a single scatterer. For the multiple scatterer case, the received signal can be written as

$$\boldsymbol{x} = \sum_{m=1}^{M} A_m \boldsymbol{a}(\omega_m) + \boldsymbol{n} \tag{2.11}$$

where M is the number of scatterers.

The design objective in beamforming is to generate a highly directive gain peak (or beam) and steer it to the desired location through adjustment of the weighting coefficients. These coefficients are commonly taken as the solution to an optimization problem for a linear filter. Letting $\boldsymbol{h}^*(\omega)$ denote the set of sensor weights applied to $\boldsymbol{x}$ to estimate the energy at frequency w, the output $y(\omega)$ is defined as

$$y(\omega) \triangleq \boldsymbol{h}^H(\omega) \boldsymbol{x}.$$

A frequently used quantity which is useful in defining optimal filters is the power of the output $y(\omega)$

$$\begin{aligned} P(y(\omega)) &\triangleq \mathbb{E}\big[|y(\omega)|^2\big] \\ &= \boldsymbol{h}^H(\omega) \, \mathbb{E}\big[\boldsymbol{x}\boldsymbol{x}^H\big] \, \boldsymbol{h}(\omega) \\ &= \boldsymbol{h}^H(\omega) R \boldsymbol{h}(\omega). \end{aligned}$$

Using these entities, there are many different beamformers that one could design. In the following sub-sections, two of the more popular ones are discussed.

■ 2.2.1 Conventional Beamformer

The first filter described is the *conventional beamformer*. This is motivated by the idea that one wants to maximize the SNR of the filter output. In particular, we seek the filter with unity energy which maximizes the output energy, i.e.

$$
\arg\max_{h(\omega):\|h(\omega)\|_2=1} \{P(y(\omega))\} = \arg\max_{h(\omega):\|h(\omega)\|_2=1} \left\{ \mathbb{E}\big[h^H(\omega)xx^H h(\omega)\big] \right\}
$$

$$
= \arg\max_{h(\omega):\|h(\omega)\|_2=1} \left\{ |A_0|^2 h^H(\omega)aa^H h(\omega) + 2\sigma^2 h(\omega)^H h(\omega) \right\}
$$

$$
= \arg\max_{h(\omega):\|h(\omega)\|_2=1} \left\{ h^H(\omega)aa^H h(\omega) \right\}. \tag{2.12}
$$

The solution for this maximization problem is

$$
h_{conv}(\omega) = \frac{a(\omega)}{\sqrt{a(\omega)^H a(\omega)}}. \tag{2.13}
$$

Because this filter maximizes the SNR of the output, it is an implementation of the matched filter. Furthermore, because $a(\omega)$ is an exponential of frequency w, the output $h^H(\omega)x$ corresponds to a Fourier coefficient of x, and the collection of outputs over all w corresponds to the Fourier transform of x. This formulation supports the reasoning for taking the Fourier transform of $\tilde{v}_r(t)$ at the end of Section 2.1.3 as it implements an approximation of the matched filter[9].

■ 2.2.2 Capon's Beamformer

The ideas for reflectivity estimation presented thus-far only consider the situation in which there is a single point scatterer. Super-resolution techniques, however, take into account the presence of other dominant scatterers to suppress interference and better resolve closely spaced sinusoids. Because the conventional beamformer is only concerned with the estimation of a single scatter, it is not a super-resolution technique. This sub-section describes *Capon's method* which is a popular super-resolution technique that tries to suppress the effects of nearby interfering scatterers to achieve a higher resolution. In particular, Capon's method determines the filter for each frequency ω as the solution to the following optimization problem

$$
\arg\min_{h(\omega)} \{P(y(\omega))\}
$$

$$
\text{s.t. } h^H(\omega)a(\omega) = 1. \tag{2.14}
$$

The idea here is to choose the filter which minimizes the total energy received while having unity gain for the frequency ω. Thus, interference from nearby strong scatterers is suppressed as the resulting filter attempts place nulls at locations producing high

---

[9]From the formulation in Eq. (2.11), this is exactly the matched filter and not just an approximation, but recall that several approximations were made in Section 2.1.3 to formulate SAR into this model.

energy contributions. Using standard optimization techniques, the solution for Capon's beamformer is

$$h_{Capon}(\omega) = \frac{R^{-1}a(\omega)}{a^H(\omega)R^{-1}a(\omega)}. \tag{2.15}$$

Note that this filter makes use of the received data via R to detect and null interfering scatterers. Thus, one cannot simply take $R = 2\sigma^2 I$ as that would ignore the data and simplify to the conventional beamformer. In array processing, R is typically estimated from the received data as the sample correlation matrix of the sensor data. In SAR however, this poses a problem since there is only one sample of the sensor data, i.e. the one image, which results in a correlation matrix with rank one. Typically, this obstacle is bypassed by using the assumption that the sensor domain data are stationary, and thus the sensors can be divided into multiple groups where each group can be treated as a distinct sample of data[3, 37]. For more information on Capon's methods and its variants, including applications to real SAR data, the reader is referred to [16, 4, 7, 5].

■ 2.3 Nonideal Scattering [10]

All the discussion up to this point has been based on the ideal point scattering model, i.e. scattering which is frequency and aspect independent, resulting in the received signal being modeled as a simple delay of the transmitted signal regardless of orientation. However, even the most simple scattering primitives (spheres, plates, dihedrals, trihedrals, cylinders, cones, *etc.*), exhibit either frequency and/or aspect dependences. The degree to which these dependencies manifest themselves depends on the geometry and physical dimensions of the object. Recently, there has been much interest in determining and exploiting the nonideal scattering behavior of simple scatterers especially for the purpose of target classification[5, 40, 45, 13, 46]. This interest is particularly strong in applications using high resolution wide-band and wide-aperture data where these dependencies are more pronounced.

Accurate scattering response models for canonical scatterers can be obtained from physics-based models such as physical optics (PO)[48] or the geometric theory of diffraction (GTD)[33, 48]. For specific scatterer geometries, these models provide a closed form approximation for the response as a function of scatterer size and orientation. The PO model is obtain from the solution of a ray-tracing model including the effects of off-specular responses. The GTD model extends the PO model by also accounting for surface currents and edge effects. The two models are nominally the same except at near tangent incident angles where the GTD model is significantly better. This thesis is primarily concerned with gross level scattering behavior around the specular response where the scatterer is strongly observable. In this regime, the simpler PO model is sufficient and thus the GTD model is not discussed further. To provide an idea of the different responses that scatterers produce, the PO model for some simple canonical

---

[10]Primary sources: Ruck[48], Benitz[5], and Chaney[12].

scatterers are cited from Ruck[48]. The first is the rectangular flat plate illustrated in Figure 2.7(a) where a flat plate of size $a \times b$ is centered at the origin and the radar backscatter is measured from the direction given by ϕ and θ. For this configuration, the radar cross section[11] (RCS) is given by

$$\sigma = \frac{4\pi(ab)^2}{\lambda^2} \left[\sin(\theta) \, \frac{\sin(ka\cos(\theta)\,\cos(\phi))}{ka\cos(\theta)\,\cos(\phi)} \, \frac{\sin(kb\cos(\theta)\,\sin(\phi))}{kb\cos(\theta)\,\sin(\phi)} \right]^2 \qquad (2.16)$$

where $\lambda = c/f$ is the wavelength, $k = 2\pi\lambda$ is the wavenumber, c is the speed of light, and f is the frequency of the signal. This response is plotted for a 1m×1m plate with $f = 9.6\text{GHz}$[12], $\phi = 0°$, and $\theta \in [-30°, 30°]$[13] in Figure 2.8(a). The PO model for a sphere of radius a centered at the origin as illustrated in Figure 2.7(b) is

$$\sigma = \pi a^2 \left[1 - \frac{\sin(2ka)}{ka^2} - \frac{\cos(2ka)}{2k^2 a^2} \right] \qquad (2.17)$$

The azimuthal response for $a = 1$m is shown in Figure 2.8(b). Another canonical scatterer is the right circular cylinder with radius a and height h centered at the origin as illustrated in Figure 2.7(c). The PO response for this scatterer is

$$\sigma = kah^2 \left[\cos(2\theta) \, \frac{\sin(kh\,\sin(2\theta))}{kh\,\sin(2\theta)} \right]^2 \qquad (2.18)$$

which is plotted in Figure 2.8(c) for $a =$0.5m and $h =$1m with $\phi = 0°$. Figure 2.7(d) illustrates the dihedral canonical scatterer composed of two rectangular plates of size $a \times b$ conjoined at the edge of length b such that the two plates form a 90° angle. The PO response for this scatterer scatterer is

$$\sigma = \pi \left[\frac{4ab}{\lambda} \cos(\theta) \, \sin(\phi) \, \frac{\sin(kb\,\sin(\theta))}{kb\,\sin(\theta)} \right]^2 . \qquad (2.19)$$

This response is plotted in Figure 2.8(d) for $a =$0.5m and $b =$1m when $\theta = 0°$.

Given canonical scatterer models, a natural inclination is to use them to improve reflectivity estimates. Extending the matched filtering approach in Sections 2.1.1 and 2.1.2 to handle nonideal scattering would be quite difficult both analytically and computationally. However, nonideal scattering could be incorporated by using the predicted response in place of the point-scatterer steering vector in Eq. (2.10) in the beamformers presented earlier. The general solution for the conventional beamformer in Eq. (2.13) and Capon's beamformer in Eq. (2.15) still hold under these more general steering vectors. Benitz [7, 5] takes such an approach in his high-definition vector imaging

---

[11]RCS is a measure of the reflected power.

[12]This is the same as the MSTAR center frequency (see Appendix B).

[13]Such a large azimuthal range is used for illustrative purposes and is beyond the capabilities of current SAR systems. Typical azimuthal spans in modern systems are as low as 1° and as high as 30°. Furthermore, the PO approximations start to lose their accuracy beyond the displayed range.

(a) rectangular plate   (b) sphere

(c) cylinder   (d) dihedral
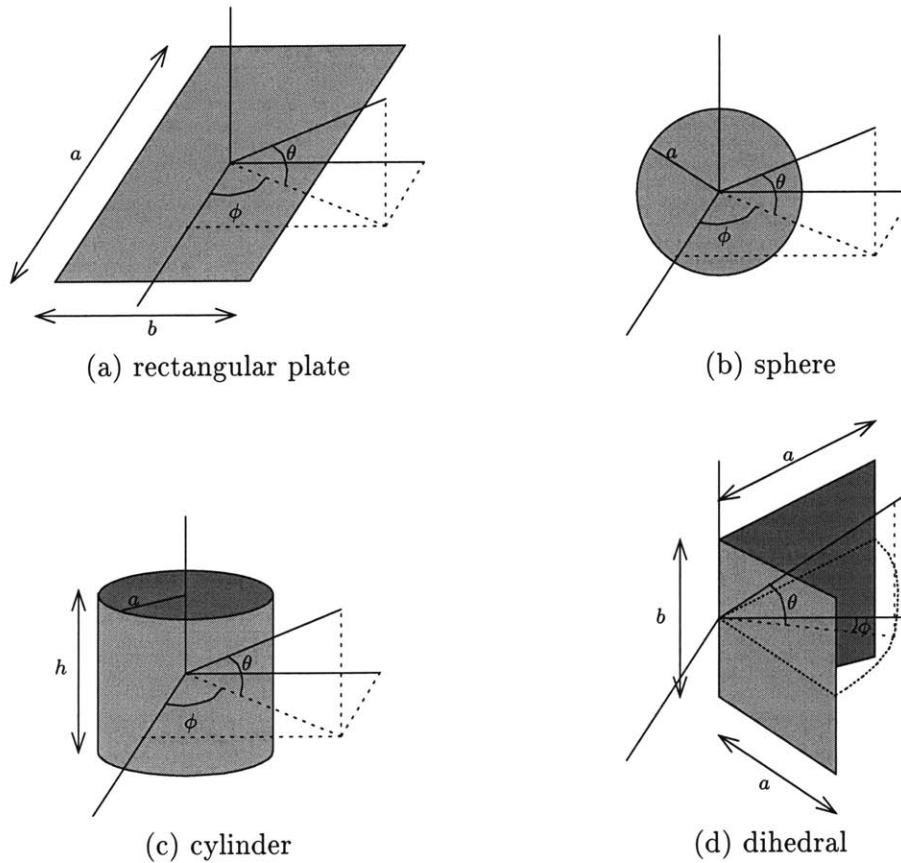
Figure 2.7. Illustrations of canonical scatterers.

(HDVI) technique where he forms a sequence of super-resolution images using a variant on Capon's algorithm. Each image is formed using the steering vector based on a pre-specified canonical scatterer whose size is fixed apriori. Although this can provide information about the specific scatterers in the data, the amount of computation involved can become prohibitive. The responses given by the PO and GTD models are quite precise and relatively highly parameterized. Each steering vector used depends not only on the type of canonical of scatterer sought (e.g. dihedral, flat plate, *etc.*) but also the physical dimensions and orientation of the scatterer. The large number of possible configurations leads an infeasible number of images to generate. Furthermore, many of differences between these scatterer models may be slight and nearly unobservable in the data, particularly if the scene is being imaged from a single viewpoint[14].

There are also complications in using Capon's method and other super-resolution methods which depend on the data correlation matrix R to suppress interfering scat-

---

[14]This is in contrast to having multiple images of the same scene from various points on the viewsphere such as in the context used by Richards[46].

(a) rectangular plate

(b) sphere

(c) cylinder

(d) dihedral
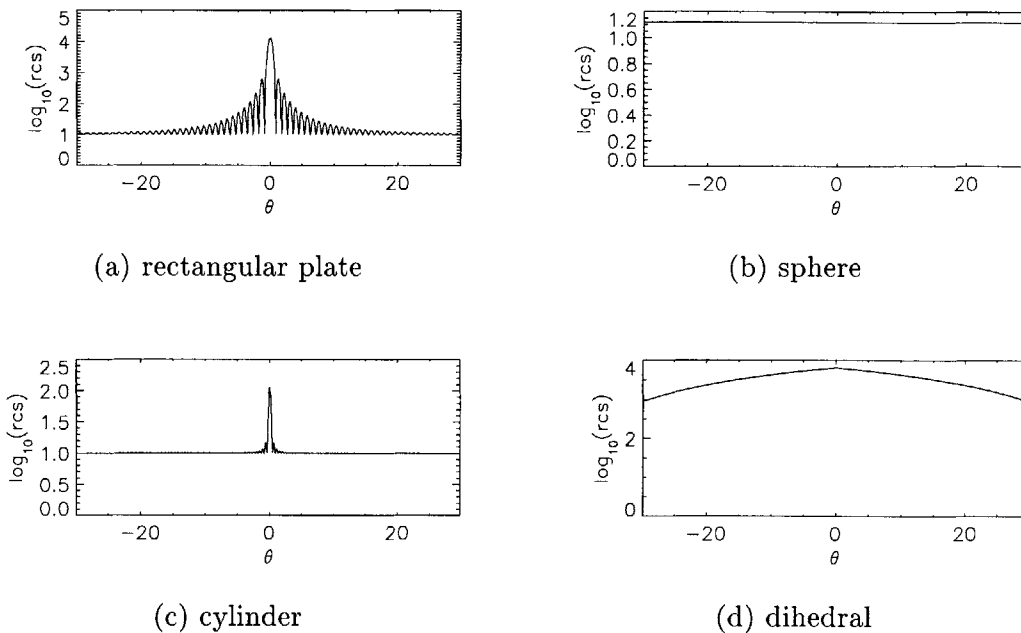
Figure 2.8. Azimuthal responses for canonical scatterers from the PO model.

terers[15]. As mentioned in Section 2.2.2, the data correlation matrix is estimated by dividing the sensors into multiple regions and taking the average of outer-products of the data vector x. However, this averaging is based on the assumption of stationarity among the sub-bands (or sub-apertures) which does not hold if the scatterers exhibit frequency (or azimuthal) dependencies. Thus, these methods inherently assume that the interfering scatterers are ideal point scatterers, i.e. that the response from the scatterer is the same in each of the divided regions. Furthermore, for physically large scatterers that span several resolution cells, these techniques may produce undesirable attributes as they try to resolve a single distributed scatter into multiple sources. In particular, when the scatterer modeled in the steering vector is too small, super-resolution techniques try to suppress the contribution from the other pixels covering the scatterer as they are interpreted as interfering scatterers. The result is a breaking up of the physically distributed scatterer into multiple scatterers. In such a situation however, the resolution should actually be coarsened, not refined, to a degree consistent with the size of the scatterer to take full advantage of the information in the reflected signal.

For wide-aperture and wide-band data, the necessity for super-resolution techniques is arguable since the image already has a high resolution, e.g. on the order of a few inches. In these cases, the overall impact of super-resolution techniques may be to cause more harm than good as the model mismatch for distributed scatterers distorts the estimates

---

[15]Note that the conventional beamformer is *not* in this group.

as larger scatterers are broken into smaller scatterers. Rather than forming myriad estimates based on a large assortment of canonical scatterer types, sizes, and orientations, it is reasonable to take a more general but physically meaningful characterization of the scattering response. This is the approach taken by Chaney et. al.[12] who propose to model scatterer anisotropy in a coarse manner capturing the fact that for many scatterers, anisotropy is related to their physical size. To simplify the amount of computation, they approximate the aspect dependent scattering as zero from all azimuthal directions except over a section of the aperture where the gain is constant, i.e.

$$a(\omega) = \begin{cases} A_0 & \omega_1 \leq \omega \leq \omega_2 \\ 0 & \text{o.w.} \end{cases}$$

for some ω_1 and ω_2. The matched filter in azimuth is then a windowed FFT (in the beamforming framework) and is equivalent to processing with just a subset of the full-aperture. The weighting across the sub-aperture is uniform as in the point scatterer case and is not exactly tuned to any particular scatterer. Although, this implies that optimal processing is not performed, myriad images need not be generated for all the scattering types sought and general differences in scatterer anisotropy can be detected. Chaney et. al. utilize sub-aperture processing by adapting the size and location of the sub-aperture to achieve a better reflectivity estimate for anisotropic scatterers. For example, the broadside flash phenomena associated with a large plate produces a scattering response which is concentrated in a small section of the aperture. Thus, by only using the section of the aperture in which most of the returned energy is concentrated, a higher fidelity estimate of the reflectivity can be achieved since only the portion with a high SNR is used. Viewed in the image domain rather than the sensor domain, this can be interpreted as adjusting the amount of spatial coarsening based on the spatial extent of the scatterer. For each pixel in their algorithm, they compute a set of reflectivity estimates from various sub-apertures and choose the one with the highest reflectivity. The resulting image is then thresholded to detect bright anisotropic scatterers. One drawback to this approach is that their selection criterion for the sub-aperture is based on a model which does not properly take into account the effect of sub-aperture length on the variance of the sub-aperture reflectivity estimate. The resulting images produced by their criterion have an excessively coarse resolution implying that too small an aperture is frequently used. Furthermore, Chaney only uses the reflectivity estimate and discards the properties of the chosen sub-aperture, i.e. its size and location. By ignoring these specifics, possible information about the scatterer's size and orientation is lost. From the duration of the flash (in terms of aperture length), the size of the scatterer can be inferred since scatterer size is inversely proportional to the duration of the flash. From the center of the flash, the orientation of the scatterer can be inferred. For example, the peak response of a flat plate should occur at broadside.

■ 2.4 Machine Learning

The field of machine learning deals with the use of computers to "learn" patterns in the data and make decisions based upon the knowledge gained. This thesis is concerned with the classification subset of machine learning. Obviously, knowledge of the underlying probability distribution of the classes involved is fundamental to any good probabilistic algorithm. However, these distributions are not known and must be estimated. Another issue is how to use these estimated distributions. This section presents the relevant background needed to address these problems. First, the topic of density estimation is discussed. This is followed by a presentation of some fundamental concepts in information theory. These concepts are then used to show the asymptotic equivalence between the likelihood ratio and the information theoretic concept of Kullback-Leibler (KL) divergence. Finally, a classification approach comparing estimated densities via KL divergence is presented.

■ 2.4.1 Kernel Based Density Estimation [16]

Knowledge of probability densities is essential in the field of stochastic signal processing. Thus, the relevant densities must be estimated to some degree. Some tasks require only a few statistics whereas others require the entire pdf, and the former often make strong assumptions about the underlying density. Techniques for density estimation broadly fall into one of two categories. The first is called *parametric* estimation. In this case, a particular parametric form for the density is assumed and its parameters are estimated from data. For example, one may assume the underlying density is a Gaussian and use the data to estimate its mean and covariance. The limitations of such parametric techniques are clear in that the functional form assumed imposes a strong restriction. The class of *nonparametric* density estimators make few assumptions on the density (such as regularity) and lets the number of parameters describing the density grow unbounded with the amount of data. It is often advantageous to use a nonparametric estimator when little is known about the underlying phenomena producing the sample values, and thus, one has no reason to assume a particular family of density functions. For example, visually viewing a nonparametric estimate of the density (e.g. a histogram) is often a useful first step in choosing one or more functionals to describe the data and possibly associate underlying causes with different parts of the distribution. However, the wealth of unlimited choices for the density makes this an ill-posed estimation problem. Even when a particular degree of regularity is assumed, estimation is difficult in regions where only a few local samples are available.

Before describing a common technique for density estimation, some notation must be defined. Lowercase boldface letters are used to denote random vectors, e.g. x, and samples of it are denoted with a corresponding uppercase letter, e.g. X. Calligraphic letters, e.g. $\mathcal{X}$, are used to denote the corresponding sample space. The distribution for

---

[16]Primary sources: Silverman[50] and Fukunaga[26].

a random variable is denoted by $F(\cdot)$ and the corresponding pdf[17] is denoted by $f(\cdot)$. When the random variable involved is not clear from the argument or context, it will be explicitly placed as a subscript under F and f.

A histogram is an example of a nonparametric density estimator. One is free to choose the number of uniformly space bins to get any degree of flexibility in the resulting estimate which is limited only by the number of data samples. The Parzen density estimator is a natural generalization of this idea. Instead of partitioning the domain into uniformly divided regions, a localized function called a *kernel* is centered at each data sample, and all the kernels are summed to produce the density estimate as illustrated in Figure 2.9. The Parzen density estimate for f given N i.i.d. samples $X_1, \ldots, X_N \sim F$ is defined as

$$\hat{f}(X) = \frac{1}{Nh^d} \sum_{i=1}^{N} K\left(\frac{X - X_i}{h}\right) \tag{2.20}$$

where $K(\cdot)$ is the kernel, h is the *kernel width* or *bandwidth*, and d is the dimensionality of x. The kernel used must be nonnegative and integrate to one in order for $\hat{f}$ to be a valid pdf. In particular, kernels that are well localized and symmetric are commonly preferred as one would expect that each data sample tells more about the probability in the neighborhood of that sample than the probability at some distant value. An attribute of this technique is that when a pdf with a finite second moment is used as the kernel, the density estimator is consistent, i.e. $\hat{f}$ converges to f in probability, if the kernel width is chosen as an appropriately decreasing function of N [28]. The estimated density function will also inherit the same differentiability properties as that of the kernel. Thus, if smoothness of the density is important, then an infinitely differentiable kernel such as the Gaussian pdf could be used.
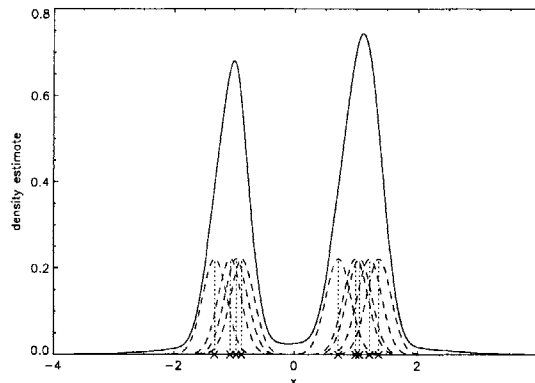
Figure 2.9. Example of Parzen density estimation.

---

[17]It will be assumed that the pdf exists.

■ 2.4.2 Information and Entropy [18]

Although the ideas in information theory originated in the communications field, the application of its results extend far into many other fields. The central ideas in information theory regard how much "information" is present in a sequence of random variables and how efficiently that sequence can be coded, where efficiency is measured in terms of the size of the coded sequence and robustness to noise. Intuitively, this is analogous to the idea of classification by which this thesis is motivated. The object to be identified can be considered the "information" in the signal; the coded sequence can be thought of as the features used to identify the object; the distortions in the image data along with the data lost due to projecting onto the feature set can be thought of as the noise in the communications channel. The goal in ATR is to choose the best feature set (code) that efficiently conveys the object's identity (information), but which is robust to imaging noise and data not captured by the feature set (channel noise). In this section, a basic review of some fundamental information theoretic concepts are presented.

First, the concept of quantifying information is considered. Given an event E that occurs with probability p, we wish to quantify the amount of *surprise* $S(p)$ in learning that E has occurred. It can be shown that if the surprise satisfies the following four axioms, then it has the form

$$S(p) = -c \log(p)$$

where the constant c is an arbitrary positive integer, taken to be 1, and the base of the logarithm is arbitrary, generally taken to be 2. The four axioms are

- *Axiom 1:* $S(1) = 0$. That is, there is no surprise in observing a sure event.

- *Axiom 2:* $S(p)$ is a strictly decreasing function of p. That is, the occurrence of less likely events generate more surprise.

- *Axiom 3:* $S(p)$ is a continuous function of p. That is, small changes in the probability of an event correspond to small changes in the surprise of observing that event.

- *Axiom 4:* $S(pq) = S(p) + S(q)$, where q is the probability of an event F independent of E. That is, the surprise of observing two independent events is the sum of their surprises.

Now, consider a discrete random variable y that can take on values $Y \in \mathcal{Y}$ with probabilities $p(Y)$. The *entropy* of y is defined as the expected value of the surprise of the outcome, i.e.

$$H(y) = \mathbb{E}[-\log(p(y))] = -\sum_{Y \in \mathcal{Y}} p(Y) \log(p(Y)). \tag{2.21}$$

---

[18]Primary sources: Cover[15] and Ross[47].

As its name implies, entropy is a measure of randomness of the outcomes of the random variable. Extending this idea of entropy to continuous random variables, the *differential entropy* of a random variable distributed according to the pdf f is defined as

$$h(x) = \mathbb{E}[-\log(f(x))] = -\int_{\mathcal{X}} \log(f(\boldsymbol{X}))\, dF(\boldsymbol{X}). \tag{2.22}$$

Another useful concept from information theory useful for ATR is the continuous form of the Kullback-Leibler (KL) divergence between two pdf's f and g. The KL divergence is defined as

$$D(f\|g) = \int f(\boldsymbol{X}) \log\left(\frac{f(\boldsymbol{X})}{g(\boldsymbol{X})}\right) d\boldsymbol{X} \tag{2.23}$$

which is the expected value, under f, of the log-likelihood ratio of f and g.[19] It is useful to think of the KL divergence as a distance between two pdf's, even though it does not qualify as a metric since neither the symmetry condition nor the triangle inequality hold. One property of metrics that it does possess is that the KL divergence between two pdf's is always nonnegative and it is zero if and only if the two pdf's are equal.

Another useful property of the KL divergence is its relation to the log-likelihood ratio (LLR). This relationship is attractive because a thresholding of the LLR, i.e. a likelihood ratio test (LRT), is the optimal[20] decision rule for a hypothesis test. In particular, the hypothesis test considered is: given a sequence of samples $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N\}$, we wish to choose one of two hypotheses to describe the nature of the samples

$$H_1 : x_i \text{ i.i.d. } \sim f = f_1 \text{ or}$$
$$H_2 : x_i \text{ i.i.d. } \sim f = f_2.$$

where f is the density producing the samples and f_1 and f_2 are two specified hypothesis densities. It is relatively straightforward to show that

$$\mathbb{E}_f\left[\log\left(\frac{f_1(x)}{f_2(x)}\right)\right] = D(f\|f_2) - D(f\|f_1), \tag{2.24}$$

i.e. the differences between KL divergences in Eq. (2.24) is the expected value of the LLR. This equivalence motivates the classification approach discussed in the next subsection which uses the KL divergence between densities, estimated from samples, as the basis for a decision rule.

■ 2.4.3 KL-Based Classification Technique [21]

To conclude the section on machine learning, several of the machine learning components presented are used to describe a classifier which asymptotically approaches the

---

[19]For continuity, $0 \log\left(\frac{0}{0}\right)$ is defined to be 0.

[20]The LRT is optimal in the sense that for a specified cost structure on the different classification decisions under the true hypothesis, the rule which minimizes the expected cost is an LRT.

[21]This section is a generalization of the approach described by DeBonet[20] whose specifics are described in Appendix A.

likelihood ratio test without assuming parametric forms for any of the underlying data. The setup for the problem is as follows. The learning process is based on L labeled sets $\{S_1, \ldots, S_L\}$ of reference data each containing N samples. For each l, all the samples in S_l are generated as i.i.d. samples from the labeled but unknown density f_l. Then, presented with another set $S_0 = \{X_1, \ldots, X_N\}$ of data consisting of i.i.d. samples drawn from one of the densities f_l, the problem is to choose under which f_l the samples of S_0 were generated.

The first task in this approach is to compute $\hat{f}_l$, the density estimate for f_l, for each $l \in \{0, \ldots, L\}$. Because no parametric forms of the densities are to be assumed, a consistent nonparametric density estimator[22], such as the Parzen density estimator, should be used.

Motivated by the equivalence of the expected value of the LLR and KL divergence in Eq. (2.24), the following derivation shows that the minimization of KL divergences is asymptotically equivalent to maximum likelihood testing for the problem defined at the beginning of the section. Starting from the maximum likelihood (ML) detector for the density generating the data set S_0:

$$\hat{l} = \arg \max_{l \in \{1, \ldots, L\}} \{f_l(X_1, \ldots, X_N)\}$$

$$= \arg \max_l \left\{ \prod_{i=1}^{N} f_l(X_i) \right\}$$

$$\approx \arg \max_l \left\{ \prod_{i=1}^{N} \hat{f}_l(X_i) \right\}$$

$$= \arg \max_l \left\{ \prod_{i=1}^{N} \frac{\hat{f}_l(X_i)}{\hat{f}_0(X_i)} \hat{f}_0(X_i) \right\}$$

$$= \arg \max_l \left\{ \prod_{i=1}^{N} \frac{\hat{f}_l(X_i)}{\hat{f}_0(X_i)} \right\}$$

$$= \arg \max_l \left\{ \sum_{i=1}^{N} \log \left(\frac{\hat{f}_l(X_i)}{\hat{f}_0(X_i)} \right) \right\}$$

$$= \arg \min_l \left\{ \frac{1}{N} \sum_{i=1}^{N} \log \left(\frac{\hat{f}_0(X_i)}{\hat{f}_l(X_i)} \right) \right\}$$

$$\approx \arg \min_l \left\{ \int \hat{f}_0(X) \log \left(\frac{\hat{f}_0(X)}{\hat{f}_l(X)} \right) dX \right\}$$

$$= \arg \min_l \left\{ D(\hat{f}_0 \| \hat{f}_l) \right\}. \tag{2.25}$$

---

[22]For other consistent density estimators, the reader is referred to Silverman[50].

All approximations in Eq. (2.25) are meant in the sense that the argument of the $\arg\max\{\cdot\}$ is approximate. The first approximation, based on the consistency of the density estimator, is that the estimated pdf's are close to the true underlying pdf's. The second approximation is that the relative frequency of each $X \in \mathcal{X}$ converges[23] to the estimated probability $\hat{f}_0(X)$. The result in Eq. (2.25) thus says that the choosing the $\hat{f}_l$ which is "closest" to $\hat{f}_0$ in terms of the KL divergence is asymptotically equivalent to performing ML detection. Furthermore, if a rejection hypothesis is incorporated via

$$\hat{l} = \begin{cases} l^* \triangleq \arg\min_l \left\{ D(\hat{f}_0 \| \hat{f}_l) \right\} & D(\hat{f}_0 \| \hat{f}_{l^*}) \leq T \\ \text{none of the available hypothesis are true} & D(\hat{f}_0 \| \hat{f}_{l^*}) > T \end{cases} \qquad (2.26)$$

for a threshold T, then this decision rule is asymptotically equivalent to selecting the pdf which minimizes the KL divergence only if

$$\mathbb{E}_{\hat{f}_0}\left[\log\left(\hat{f}_l(x) \right) \right] > T$$

i.e. the estimated log-likelihood is greater than T.

---

[23]There are actually two asymptotic results being used here. Both the relative frequency of occurrence in S_0 and the density estimate converge in probability to the true pdf and thus to each other. The former is by the weak law of large numbers, and the latter is by consistency of the density estimator.

Chapter 3

Sub-aperture Representation

This chapter introduces the sub-aperture pyramid which forms the cornerstone of the modeling and analysis of scatterer anisotropy in this thesis. The sub-apertures in the pyramid are used to represent the SAR data with sub-aperture reflectivity measurements and to represent anisotropy scattering hypotheses. The latter topic is left for later chapters while this chapter focuses on the data representation that the sub-aperture pyramid provides.
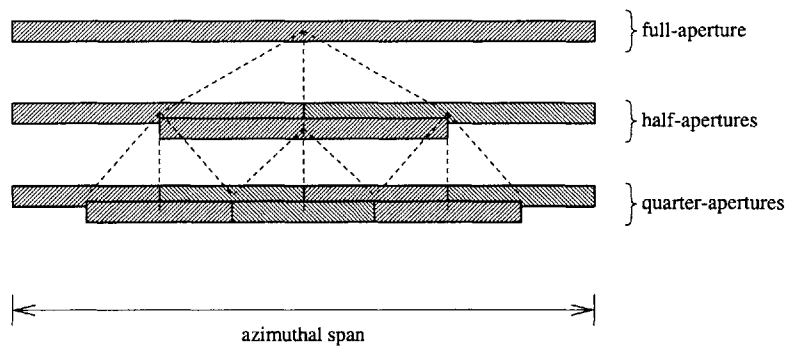
Figure 3.1. Illustration of a sub-aperture pyramid.

The intuitive idea of the sub-aperture pyramid is to use a set of sub-apertures covering the entire aperture to represent the SAR data in a way that reveals scatterer anisotropy. The sub-apertures are arranged in a pyramidal fashion as shown in Figure 3.1 with the full-aperture at the top and successive levels containing increasingly smaller sub-apertures, where at each scale, the entire aperture is covered by sub-apertures. The motivation for using such a structure is best illustrated with the examples shown in Figure 3.2. Here, the sub-aperture representations of three azimuthal responses with different degrees of anisotropy are depicted by using the amount of shading to convey the intensity of the sub-aperture reflectivity estimates. The first example is an isotropic response for which each sub-aperture returns an identical reflectivity estimate. The second response is highly anisotropic which is revealed by the progression of increasing reflectivity estimates forming a "streak" down the middle of the pyramid. The third response is mildly anisotropic which is revealed by the sub-aperture esti-

mates in the middle of the aperture capturing the most energy. In moving from larger to smaller sub-apertures, in this case, once we reach a scale for which sub-aperture size no longer exceeds the width of the azimuthal flash, the reflectivities cease to increase as the pyramid is traversed downwards. Such a leveling off indicates that the examined sub-apertures are too narrow in azimuth. Thus, the level at which this transition occurs indicates the degree of anisotropy of the underlying scatterer. It is through the patterns illustrated by these examples that the sub-aperture pyramid captures scatterer anisotropy and reveals the degree and azimuthal location of the anisotropy.
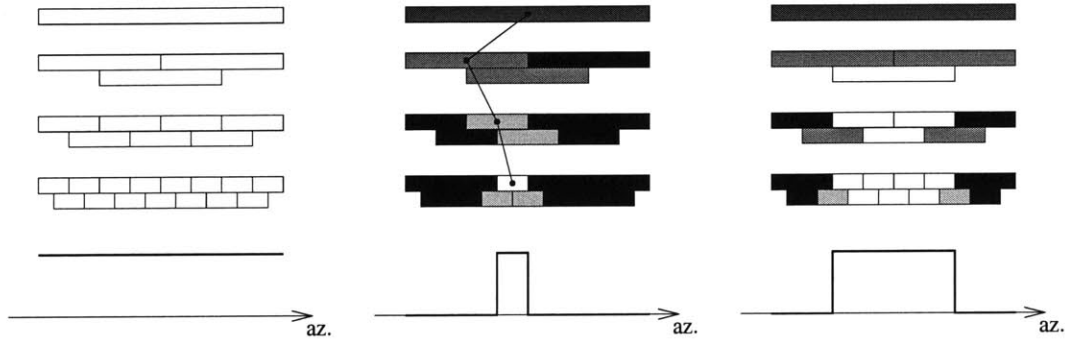
Figure 3.2. Illustration of how different azimuthal responses (bottom) are represented on the sub-aperture pyramid (top).

Based on the illustrations in Figure 3.2, one may think that the smallest sub-apertures provide the best measurements for detecting and characterizing anisotropy. While true for the case of a single scatterer, this is not true when there are multiple scatterers are present. The reason is that smaller sub-apertures have a corresponding coarser imaging resolution, as discussed in Section 2.1, resulting in more interference in the estimated reflectivities. This thesis thus proposes the sub-aperture pyramid as a means to achieve both a high cross-range imaging resolution and high azimuthal resolution. Note that the sub-aperture pyramid only applies to azimuthal data resulting in a multiresolution representation in cross-range[1]. In particular, the pyramid is *not* applied in frequency as the entire bandwidth is always used to generate reflectivity estimates, i.e. range is always imaged at the finest resolution. This variation in resolution in only one dimension is in contrast to the approaches proposed by Irving *et. al.*[31], Fosgate *et. al.*[24], and DeBonet *et. al.*[20] which are multiresolution in both down-range and cross-range.

The remainder of this chapter presents the specifics of the sub-aperture pyramid in more detail. First, the structure of the pyramid is defined and interpreted. Then, sub-aperture reflectivity estimation is discussed. The chapter ends with interpretations of the sub-aperture pyramid in other contexts illustrating the attractive properties it possesses.

---

[1]Frequency dependencies are not explored in this thesis because complications arise in simultaneously performing the analysis in azimuth and frequency as discussed in Section 8.2.

■ 3.1 Sub-aperture Pyramid

Each sub-aperture can be associated with an interval of the real line. In particular, the full-aperture is denoted by the interval $[0,1) = \{t|0 \leq t < 1\}$. Then, the sub-aperture can be defined with simple set theoretic notions. The intuitive idea of the sub-aperture pyramid is to generate an over-complete covering of the full-aperture with sub-apertures that can be arranged in a pyramidal structure. Formally, the sub-aperture pyramid is taken to be a set S of sub-apertures with the following structure. The set S is partitioned into smaller sets S_m for $m \in \{0,\ldots,M\}$ all the elements of which have a particular sub-aperture length with larger values of m conveying smaller sub-aperture lengths as illustrated in Figure 3.3. S_0 refers to the singleton set consisting of the full-aperture at the top of the pyramid, and S_M refers to the set of the smallest sub-apertures at the bottom of the pyramid. A second subscript on S denotes a specific sub-aperture at the given scale. For what follows later in the thesis, the following conditions are imposed on S:

(A) $\forall S_{m,i}$, $S_{m,i} = [a,b) \subset [0,1)$ for some a and b with $0 \leq a < b \leq 1$,

(B) $S_0 = \{[0,1)\}$,

(C) $\forall S_{m,i}, \exists$ a partition $\mathcal{P}(m,i) \subset S_M$ of $S_{m,i}$, and

(D) $\forall S_{m,i}$ with $m \geq 1$, $\exists S_{m-1,j}$ such that $S_{m,i} \subset S_{m-1,j}$.

The first condition simply restricts each sub-aperture to be a connected subset of the available aperture. The use of connected sets is motivated by the predominance of concentrated unimodal scattering in azimuth produced by canonical scatterers. The second condition states that the full-aperture resides at the top of the pyramid which allows for consideration of the minimal degree of anisotropy, and more importantly, produces the finest imaging resolution permissible. The third condition ensures that each sub-aperture can be represented by a partition of smallest-sized sub-apertures at the bottom of the pyramid. This partitioning permits for simplified analysis in the sub-aperture attribution discussed in Chapter 5. The fourth condition asserts that each sub-aperture, except the full-aperture, is contained within a parent sub-aperture. Herein, the term *sub-aperture pyramid* refers to one satisfying conditions (A)-(D).

To illustrate the idea of the sub-aperture pyramid, three examples are presented. The first is the disjoint, half-aperture pyramid shown in Figure 3.3(a). Here, each sub-aperture is divided into two halves for each change in scale thus producing a dyadic partitioning of the aperture. At each scale, this pyramid uses the minimum number of sub-apertures to achieve the covering for a given sub-aperture length. Such a minimal representation, however, is not robust to shifts in the azimuthal response. For example, if the rightmost azimuthal response in Figure 3.2 were applied to the disjoint covering, then the response could not be well captured with the set of half-apertures, and one would have to resort to using smaller sub-apertures which have a coarser imaging resolution. An approach to addressing to such a problem is to increase the amount

of sub-aperture overlap which is done for the half-overlapping, half-aperture pyramid shown in Figure 3.3(b). Here, each sub-aperture is divided into three half-overlapping sub-apertures. Although the half-overlapping pyramid is still susceptible to shifts in azimuth, the impact of such shifts is considerably reduced. One could increase the amount of overlap even more so if desired. For example, one could use the sub-aperture pyramid shown in Figure 3.3(c) where the sub-apertures have a constant offset of $1/8$ for each level.
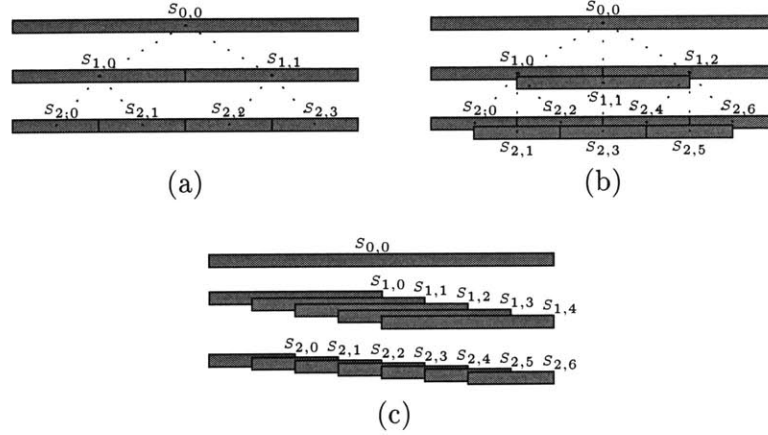
Figure 3.3. Three examples of sub-aperture pyramids. (a) Disjoint, half-aperture pyramid. (b) Half-overlapping, half-aperture pyramid. (c) Eighth-incrementing, half-aperture pyramid.

■ 3.2 Sub-aperture Reflectivity

To quantify azimuthal variations, each sub-aperture in the pyramid is used to measure the localized (in azimuth) reflectivity for a specified location in the image scene. This is done using the azimuthal response which for a given location is formed by first extracting the azimuthal sensor measurements from range compressed data for a given range location. This produces a 1-D signal along the aperture which is not cross-range compressed. The *azimuthal response* is then obtained by demodulating this signal for its specific cross-range location. The purpose in using this partially processed signal is that it focuses the energy in down-range and removes the distracting modulation on the signal caused by the offset in cross-range. The result is an azimuthal response like those presented for the canonical scatterers in Figure 2.8 which only have to be integrated to generate the reflectivity measurement. In particular, each sub-aperture $S_{m,i}$ generates an associated measured reflectivity $q_{m,i}$ according to

$$q_{m,i} = \int_{S_{m,i}} a(\omega) \, d\omega \qquad (3.1)$$

where $a(\omega)$ represents the azimuthal response for the given location.

This sub-aperture measurement is not used as a reflectivity estimate due to lack of normalization. An estimate has an implicit model which in this thesis is taken to be the same as that used by Allen[2] and Chaney[12], i.e. the sub-aperture scattering model whose response is uniform over the aperture and zero elsewhere. The *reflectivity* for this model refers to the amplitude on boxcar model. Thus, assuming white noise across the aperture, the maximum likelihood reflectivity estimate is the average of the azimuthal response over the sub-aperture, i.e.

$$A_{m,i} = \frac{1}{L_{m,i}} \int_{S_{m,i}} a(\omega) \, d\omega = \frac{q_{m,i}}{L_{m,i}} \tag{3.2}$$

where $L_{m,i}$ is the length of the sub-aperture $S_{m,i}$. It is generally more convenient to work with the sub-aperture measurements $q_{m,i}$ because these measurements can be related to each other by simple addition. In particular, if $\mathcal{P}(m,i)$ is a partition of $S_{m,i}$, then $q_{m,i}$ can be written as

$$q_{m,i} = \sum_{S_{M,j} \in \mathcal{P}(m,i)} q_{M,j}$$

without having to worry about scaling factors. However, when displaying images of reflectivity estimates, the normalized sub-aperture reflectivity estimate will be used.

The motivation for using the sub-aperture pyramid is that it should reveal distinguishing aspect dependencies in the scattering. To see that it achieves this goal, consider a metal sphere which has a strong response in all directions and thus produces a strong consistent reflectivity estimate from each of the sub-apertures. However, as depicted in Figure 3.4, a flat plate produces a specular response which is significantly stronger when oriented broadside with respect to the radar. Thus, the reflectivity estimates vary across the sub-apertures with the largest estimates coming from the sub-apertures oriented broadside to the plate. Furthermore, because various sized sub-apertures are used, the duration of the broadside flash is also captured in this representation. To see this, note that for all sub-apertures which reside within the main-lobe of the response, the reflectivity estimates are consistently large. As the sub-aperture is expanded, however, the additional signal energy received is relatively insignificant thereby lowering the reflectivity estimate which is normalized with respect to sub-aperture length in Eq. (3.2). Not only does using an excessively large sub-aperture lower the normalized reflectivity estimate, but doing so also results in a noisier estimate because the additional portion of the aperture used in this estimate is dominated by noise[2]. To illustrate that this anisotropic phenomena is present in real data as well, Figure 3.5 shows images of sub-aperture reflectivity estimates for a three level, disjoint, half-aperture pyramid. The scene contains a BMP-2 tank from the MSTAR public release dataset[1] which uses

---

[2]To be precise, it is actually the SNR of the estimate that decreases when too much of the aperture is used. The estimator variance itself actually decreases because the coherent averaging is done over a larger aperture, but the squared-mean of the estimate decreases even more since no additional signal energy is being incorporated.

a 2.8° aperture. Even in this narrow aperture setting, it is clear that the large scatterer at the front of the vehicle produces an anisotropic response which is captured in the middle section of the sub-aperture pyramid.
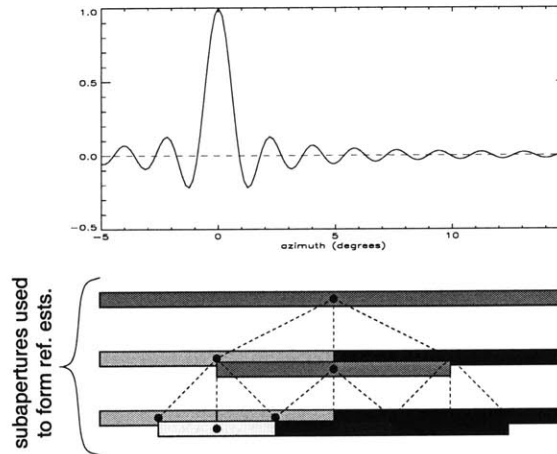
Figure 3.4. The response of a $1m \times 1m$ flat plate and a depiction of the reflectivity estimate for each of the sub-apertures. Lighter shaded sub-apertures convey larger reflectivity estimates.
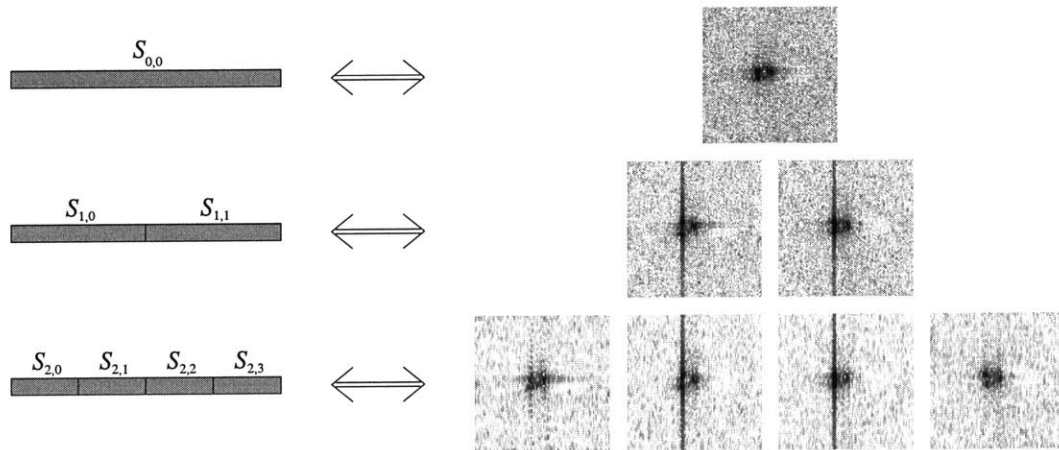
Figure 3.5. Left: Disjoint, half-aperture pyramid. Right: Corresponding sub-aperture images of a BMP-2 at a 17° elevation and 0° azimuth. For each image, the front of the vehicle is the portion nearest the left edge of the image. The azimuthal flash due to the front of the tank can be seen from the quarter-aperture images to be predominantly contained in the middle section of the aperture. However, while providing more azimuthal information about the scattering, these smaller sub-aperture images suffer from severe sidelobe artifacts as can be seen from the front plate response corrupting the reflectivity estimates in neighboring cross-range pixels.

The differences among the three pyramids in Figure 3.3 lie in the degree of sub-aperture overlap. Each level of the disjoint pyramid in Figure 3.3(a) is critically sampled

in the sense that it has the minimal covering of the full-aperture. The other two pyramids are "oversampled" in azimuth as they populate the pyramid with additional overlapping sub-apertures that introduce no new information, although their presence more clearly reveals the properties of "misaligned" azimuthal flashes which do not exactly line up with the divisions of the minimal disjoint pyramid representation. This example illustrates the purpose of the redundant overlapping sub-apertures which is to provide robustness to azimuthal shifts of the data, i.e. slight rotations of the target. The appropriate azimuthal oversampling for a pyramid is an open issue which involves balancing model approximation error and computational burden. The optimal trade-off between these two costs is an issue not addressed in this thesis. In our experiments, the half-overlapping, half-aperture pyramid works reasonably well and will be used for the remainder of the thesis.

■ 3.3 Interpretations

The sub-aperture pyramid is an over-complete representation of the SAR data. The reason for this over-completeness is to aid the robust utilization of anisotropic scattering. To help the understanding of the sub-aperture pyramid, this section comments on the representation in the context of existing works.

■ 3.3.1 Time-Frequency Plane Tilings

Because the integration in Eq. (3.2) is only performed over the sub-aperture $S_{m,i}$, this corresponds to a Fourier coefficient of windowed data. The use of smaller sub-aperture windows thus produces estimates with a coarser imaging resolution. This trade-off is exactly the Fourier time-frequency uncertainty principle applied to SAR data, i.e. with purely linearly processing, one cannot obtain both high azimuthal (time) and high cross-range (frequency) resolution due to the inverse relationship between the two. Thus, each level of the pyramid provides a different azimuthal–cross-range resolution trade-off which is best illustrated with the time-frequency plane tilings used in signal processing and wavelets in particular[59]. Each level of the pyramid produces a tiling of the azimuthal–cross-range plane as depicted in Figure 3.6 where the dimensions of the tiles represent the amount of resolution in each domain, and each tile must span the same amount of area. Note that the tilings in Figure 3.6 are for the disjoint, half-aperture pyramid with critically sampled image domain data, i.e. there is no oversampling in either domain. If one were to oversample, then the tiles would have a corresponding overlap in the respective domain[3]. The tiling for the full-aperture in Figure 3.6(a) is best suited for isotropic scatterers for which one is interested in obtaining high cross-range resolution but not azimuthal resolution. The tiling for the half-apertures in Figure 3.6(b) offers a higher degree of azimuthal resolution but at the cost of imaging

---

[3]Such oversampling is useful in application but is not illustrated here in order to clearly present the resolution trade-off involved.

resolution. The exchange further progresses with the use of the quarter-apertures as illustrated in Figure 3.6(c).

The presence of various azimuthal–cross-range resolution representations in the sub-aperture pyramid is desirable when large distributed scatterers, which produce anisotropic responses, are present in the imagery. In these cases, the benefits of precise estimation of scatterer location can be outweighed by the information conveyed in the azimuthal variations. Thus, it is expected that the best representation varies with the scatterer at hand, and having all the representations in a structured framework is beneficial. This idea is illustrated for a moderately sized scatterer spanning multiple pixels in Figure 3.7 where the degree of shading indicates the intensity of the scattering in (a) and that of the estimates in (b)-(d). The true azimuth–cross-range span of this scatterer is displayed in Figure 3.7(a). The corresponding reflectivity estimates using the different levels of the sub-aperture pyramid are displayed in Figure 3.7(b)-(d). The full-aperture accurately captures the spatial extent of the scatterer but not the intensity. In this case, the reflectivity estimates are biased towards artificially small values because the scattering is modeled as uniform over the full-aperture and thus the reflectivity estimate in Eq. (3.2) is normalized by an excessively large aperture length. The half-aperture representation in Figure 3.7(c) best captures the scattering response because the size of the tiles in this representation is closest to the azimuth–cross-range span of the underlying scatterer, i.e. this representation provides the closest model match. The result is that the scatterer is accurately reported as coming from the first half-aperture, and the scatterer's reflectivity and spatial extent are estimated reasonably well. The quarter-apertures in Figure 3.7(d) give an accurate estimate of the scatterer's azimuthal concentration and reflectivity, but a poor representation of the spatial location and extent due to the coarse imaging resolution.
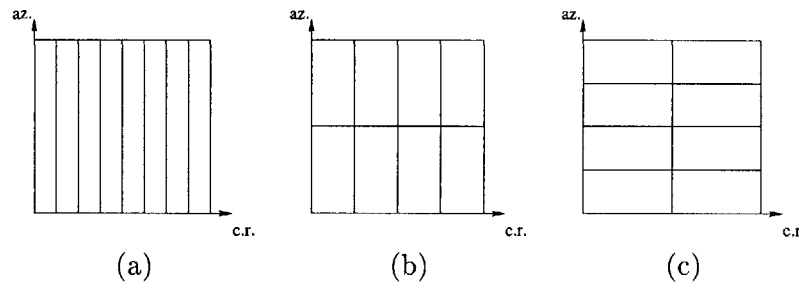
Figure 3.6. Tilings of the azimuth–cross-range plane. (a) Full-aperture tiling. (b) Half-aperture tiling. (c) Quarter-aperture tiling.

Note that the sub-aperture pyramid does not produce tilings like those of the wavelet transform in Figure 3.8(a). A level of a "pyramid" which would produce such a tiling is illustrated in Figure 3.8(b). In particular, this tiling would require different sized sub-apertures within the same level of the pyramid. From a SAR physics point of view, there is no apparent motivation for such a tiling. Because scatterers are imaged from
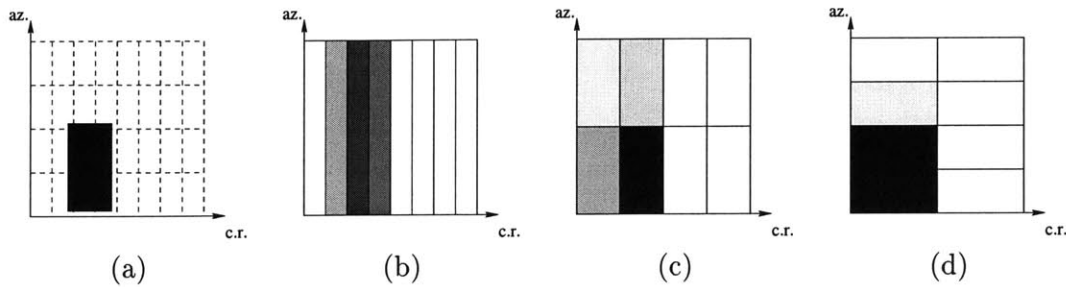
Figure 3.7. Tilings of the azimuth–cross-range plane and how they represent a distributed scatterer. (a) The underlying scatterer's energy coverage in azimuth and cross-range. (b) Full-aperture representation. (c) Half-aperture representation. (d) Quarter-aperture representation.

an arbitrary azimuthal orientation, it is reasonable that at each level of the pyramid, the sub-apertures should be equally-sized and regularly spaced to maximize robustness to the azimuthal uncertainty. Furthermore, the use of varying sized sub-apertures at the same level would complicate the analysis due to the multiple imaging resolutions that would be produced by that level. Hence, such sub-aperture arrangements are not considered in this thesis.
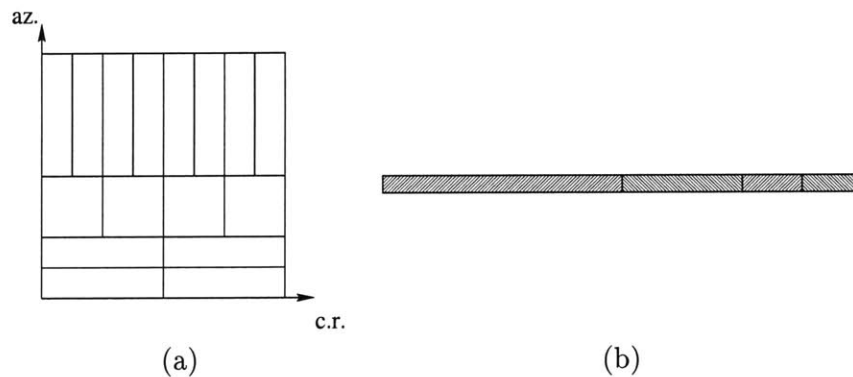
Figure 3.8. (a) Wavelet tiling of the azimuth–cross-range plane. (b) A level of a "pyramid" which would produce such a tiling.

■ 3.3.2 Matched Filter Bank

The reflectivity estimate in Eq. (3.2) is the same as that used by Allen *et. al.*[2] and Chaney *et. al.*[12]. In particular, it is the matched filter estimate for the sub-aperture scattering model. The extension to the representation of Allen and Chaney is in the use of the pyramidal structure to organize the sub-apertures. Like the work of Chaney, this pyramid allows sub-apertures of various sizes and offsets to be used, whereas Allen

uses a fixed sub-aperture size. Thus, Allen's representation corresponds to a particular level of a sub-aperture pyramid.

The sub-aperture pyramid is also similar in spirit to Benitz's high definition vector imaging (HDVI)[5, 6]. For both the sub-aperture pyramid and HDVI, reflectivities are estimated based on various scattering hypotheses and are used to form a vector of reflectivity images. A notable difference between the two approaches is that HDVI uses super-resolution techniques to achieve a resolution that exceeds the Fourier limit in Eqs. (2.3) and (2.6), whereas the sub-aperture pyramid is based on the conventional beamformer and thus does not exceed the Fourier limit. Super-resolution techniques are not used in the pyramid representation because they are not consistent in their handling of anisotropic scatterers. In particular, as discussed in Section 2.3, super-resolution techniques are designed to suppress the interference from neighboring *point scatterers* even when the steering vector is adapted to anisotropic, and hence spatially extended, scatterers. The suppression of interference is specifically tuned to that from point scatterers because of the manner in which the data correlation matrix is estimated. The method assumes stationarity in frequency and azimuth which is equivalent to frequency independent, isotropic scattering. As a result, these methods tend to artificially break up spatially distributed scatterers in the imagery. Because it is the aim of this thesis to develop robust methods to exploit anisotropy, the more consistent conventional beamformer is used to estimate reflectivity.

Chapter 4

Nonparametric Estimation of Anisotropy

This chapter presents a SAR classifier adapted to exploit scatterer anisotropy from the density-based classification idea discussed in Section 2.4.3. The feature vector used in this classifier is obtained from a sub-aperture pyramid of Section 3.1 in order to exploit anisotropic phenomena. In particular, sub-aperture reflectivities are estimated over the entire image to reveal the azimuthal variations produced by the scatterers on a target. The Parzen density estimator is then used to learn the distribution of the sub-aperture reflectivities. The resulting density estimate implicitly describes the anisotropy exhibited by the target in the image. Thus, by comparing the density estimate from an unlabeled target image to density estimates from labeled target images, a classification based upon anisotropy behavior is obtained.

The classification approach presented in this chapter is a modification of the approach proposed by DeBonet et. al.[20] which is summarized in Appendix A. The most significant difference between the two approaches is in the choice of the feature vector. DeBonet uses a multiresolution wavelet-based feature set obtained through the application of oriented edge-filters to the log-magnitude image data at multiple resolutions. In particular, the feature vector for a given location is taken as the set of edge-filter outputs corresponding to the specified location. Use of such a feature set reveals the oriented visual texture of images as discussed by DeBonet et. al. in [19]. Like the wavelet-based feature vector used by DeBonet, the sub-aperture feature vector used in this chapter is also multiresolution in nature due to the use of apertures of varying lengths. However, the use of the sub-aperture feature vector is motivated by radar scattering physics in contrast to image texture.

The remainder of this chapter is organized as follows. Section 4.1 establishes the context for the classification problem considered. Section 4.2 describes the algorithm in detail including construction of the sub-aperture feature vector and implementation issues regarding the nonparametric classifier. Section 4.3 then presents results from the application of the proposed classifier to the MSTAR[1] data set.

■ 4.1 Problem Setting

Prior to presenting the sub-aperture based classifier, this section describes the context in which the classifier will be used. The intended application is that described in the beginning of Section 2.4.3 but specifically adapted to SAR. In particular, the algorithm is designed for classifying a target that is imaged from a single unknown orientation. The feature vectors generated from this image are called *test data*. It is assumed that images of labeled targets for each classification of interest are available for comparison to the test data. The feature vectors generated from images of these labeled targets are called *reference data*. Because anisotropy, by definition, varies in azimuth, it is necessary to have reference images at a variety of azimuths for each vehicle. Each of these images is used to form a model for a vehicle-pose composite hypothesis. The classification algorithm then tests each of these composite hypotheses and returns the vehicle classification as that from the composite hypothesis selected.

■ 4.2 Nonparametric Sub-aperture Based Classifier

This section describes the algorithm used to classify the contents of an image. The process can loosely be broken down into three steps. The first step is to represent the data in a fashion that reveals the azimuthal dependencies we wish to exploit. Section 4.2.1 presents the sub-aperture feature vector used to accomplish this task. The second step is to learn the anisotropy behavior exhibited by a target. Section 4.2.2 describes the application of the Parzen density estimator to the sub-aperture feature vector to learn this behavior. With the anisotropy characterization in hand, a classification decision is then made by comparing the density estimated from the unknown image to densities learned from reference images. Section 4.2.3 describes the use of the KL divergence and Chi-square divergence to compare these densities.

■ 4.2.1 Sub-aperture Feature Vector

Each sub-aperture in the pyramid is used to form an image of log-magnitude reflectivity estimates[1] for the target field. Because the sub-apertures in the pyramid have various lengths, the resolutions of the corresponding sub-aperture images also vary. In particular, the resolution of a sub-aperture image is inversely proportional to the sub-aperture length, and thus, coarser resolution images can be down-sampled at a rate inversely proportional to the sub-aperture length. Only pyramids in which the lengths of larger sub-apertures to smaller sub-apertures form integer ratios are considered in this chapter. The use of integer ratios produces resolution cells whose cross-range extent are also integer multiples between scales. The resolution cells can thus be arranged in a tree-structure based upon the terrain spanned as depicted in Figure 4.1. The *sub-aperture feature vector* for a specified location is generated by taking the collection of all sub-

---

[1]For the remainder of this chapter, when the term *reflectivity* is used, it is taken to refer to the log-magnitude reflectivity.

aperture reflectivities corresponding to the given location. As an example, a pyramid using a half-aperture dividing scheme produces a set of images where pixels can be presented in a dyadic fashion as the resolution is varied as illustrated in Figure 4.1. The feature vector for a given location can then be seen to be a "branch" traversing this tree terminating on the specified location. The collection of feature vectors for the entire image is taken to be the data set $\{X_i\}_{i=1}^N$ where N is the number of pixels in the image. Note that it is immaterial whether a disjoint or overlapping sub-aperture pyramid is used in this example since resolution is determined by the length of the sub-aperture and is independent of the amount of overlap between sub-apertures.
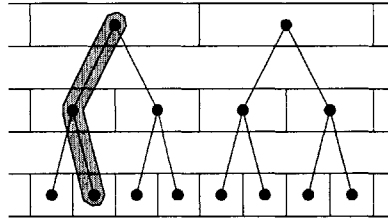
Figure 4.1. The dyadic tree structure for pixels that results from using a half-aperture splitting scheme in the sub-aperture pyramid. The feature vector for the shaded pixel is indicated by the shaded branch.

It is important to understand the difference between the feature vector tree and sub-aperture pyramid, both of which have a pyramidal structure. The sub-aperture pyramid spans the azimuthal domain and is composed of the sub-apertures used to produce reflectivity estimates. Each node on this pyramid corresponds to a sub-aperture. This pyramid has nothing to do with image location. The feature vector tree, in contrast, spans cross-range in the image domain and is used to represent the reflectivity estimates over the entire target field using various sub-apertures. This tree is composed of vector-valued reflectivity estimates (one component for each sub-aperture) for each pixel location. The collection of estimates in a path down this tree constitutes the feature vector for the location corresponding to the node at the bottom of the branch.

Although the feature vector tree and sub-aperture pyramid are defined in different domains, there are relationships between the two. In fact, the topology of the sub-aperture pyramid determines the topology of the feature vector tree and vice-versa. Figure 4.2 illustrates these relations for the half-overlapping, half-aperture pyramid. First, the number of levels in the feature vector tree and sub-aperture pyramid are the same because their levels are defined by resolution or equivalently sub-aperture length. However, the root node of the sub-aperture pyramid corresponds to the leaves of the feature vector tree and the root nodes of the feature vector tree correspond to the leaves of the sub-aperture pyramid. To understand this reversal, recall that the root node of the sub-aperture tree is the full-aperture and use of the full-aperture generates the finest cross-range resolution. Thus, the corresponding entries in the feature vector tree are at the leaves of the tree. Similarly, the shortest sub-apertures are at the bottom of the sub-aperture pyramid, but their corresponding nodes are at the root of the feature

vector tree since these sub-apertures give the coarsest cross-range resolution. Due to this bijective relation for resolution, the same label m is used to denote scale in both the feature vector tree and sub-aperture pyramid. However, to account for the difference in the progression of resolution on the two structures, the ordering of the values of m on the feature vector tree are reversed from that on the sub-aperture pyramid, i.e. $m = M$ for the root node and $m = 0$ for the leaves on the feature vector tree. To prevent any confusion related to scale when discussing the sub-aperture pyramid and feature vector tree, the terms *finer* and *coarser* are exclusively used to refer to resolution in the image domain while the terms *larger* and *smaller* are used to refer to sub-aperture lengths in the pyramid.
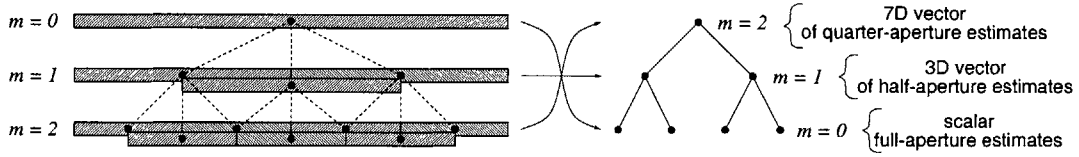
Figure 4.2. The relationship between (left) the half-overlapping, half-aperture sub-aperture pyramid and (right) the corresponding feature vector tree.

The number of offspring at each scale in one structure also effects the other structure. In particular, the number of children that each node on feature vector tree spawns is equal to the ratio of sub-aperture lengths between adjacent levels in the sub-aperture tree. For example, any covering with half-aperture splitting produces a binary tree of feature vectors. Furthermore, the number of sub-apertures for a given level in the sub-aperture pyramid is equal to the dimensionality of the nodes at the associated scale in the feature vector tree since each sub-aperture produces a reflectivity estimate.

■ 4.2.2 Density Estimation

From a collection of sub-aperture feature vectors, the generating density is estimated using the Parzen density estimator described in Section 2.4.1. In particular, the density estimate at a point X is given by,

$$\hat{f}(X) = \frac{1}{NV} \sum_{i=1}^{N} K(X - X_i; h) \qquad (4.1)$$

where V is a normalizing volume factor, $h = [h^{(0)}, \dots, h^{(M)}]$ is the kernel width parameter, and K is the hyper-rectangle kernel given by

$$K(y; h) = \begin{cases} 1 & : \text{if } \forall m, \|y^{(m)}\|_\infty < h^{(m)}, \\ 0 & : \text{otherwise} \end{cases} \qquad (4.2)$$

where $y^{(m)}$ represents the sub-vector of y whose components correspond to scale m, and $h^{(m)}$ represents the kernel width used for all feature vector components at scale m.

Thus, the kernel corresponds to an indicator function conveying whether or not the data point y is in the hyper-rectangle whose side-length is $2h^{(m)}$ for components at scale m. In order for the density estimator in Eq. (4.1) to integrate to one, the normalization factor V must correspond to the volume of the hyper-rectangle. Thus, V is simply

$$V = \prod_{S_{m,i} \in S} 2h^{(m)}.$$

In the approach presented in this chapter, h is a constant parameter. The procedure for determining h is discussed later in Section 4.2.5

The combination of the use of the hyper-rectangle kernel and the tree structure on the feature vectors affords a computationally efficient method for estimating the density. In particular, computing the distances used in the Parzen kernel in a top-down manner on the feature vector tree allows for the density estimation for each X to be done in $\mathcal{O}(N_1 \log(N_2))$ instead of $\mathcal{O}(N_1 N_2)$ where N_1 and N_2 are the number of pixels in down-range and cross-range respectively.

■ 4.2.3 Density Comparison

Two divergence measures are used in this chapter to compare estimated test and reference densities to make a classification decision. A natural method is to use the KL divergence discussed in Section 2.4 via the decision rule in Eq. (2.26), i.e.

$$\hat{l} = \begin{cases} l^* \triangleq \arg\min_l \left\{ D(\hat{f}_0 \| \hat{f}_l) \right\} & : \text{if } D(\hat{f}_0 \| \hat{f}_{l^*}) \leq T \\ \text{none of the available hypotheses are true} & : \text{otherwise} \end{cases} \tag{4.3}$$

which we will call the *KLD rule*. This rule approximates the ML hypothesis test with a "none of the above" decision if no hypothesis is likely enough. Although well motivated, recall that approximating the ML decision rule as selecting the hypothesis with minimum KL divergence in Eq. (4.3) requires the assumption of independent feature vectors. This assumption, however, does not hold due to the multiresolution nature of the feature vector. The violation of this assumption is especially clear in the case of tree-structured features where neighboring feature vectors share coarser scale components as illustrated in Figure 4.1. Thus, motivated by the work of DeBonet et. al.[20], this chapter also investigates the use of the Chi-square divergence as a replacement for the KL divergence. In particular, the KLD rule in Eq. (4.3) is replaced by

$$\hat{l} = \begin{cases} l^* \triangleq \arg\min_l \left\{ \chi^2(\hat{f}_l, \hat{f}_0) \right\} & : \text{if } \chi^2(\hat{f}_{l^*}, \hat{f}_0) \leq T \\ \text{none of the available hypotheses are true} & : \text{otherwise,} \end{cases} \tag{4.4}$$

where χ^2 is the Chi-square divergence given by

$$\chi^2(\hat{f}_l, \hat{f}_0) = \int \frac{\left(\hat{f}_l(X) - \hat{f}_0(X) \right)^2}{\hat{f}_l(X)} dX \tag{4.5}$$

which is a second-order approximation of two times the KL divergence[15] $2D(\hat{f}_0\|\hat{f}_{l*})$. The rule in Eq. (4.4) will be called the χ^2 *rule*. The performance of both the KLD and χ^2 rules are examined in Section 4.3.

■ 4.2.4 Feature Space Discretization

Calculation of both the KL and χ^2 divergences involve an integration over the entire sample space which is computationally intensive due to the nonparametric nature of the density estimates. Thus, an approximation to the integral is taken. As described in Appendix A, DeBonet approximates the χ^2 divergence[2] by replacing the integral with a weighted summation evaluated at the reference data samples. By the law of large numbers (LLN), this estimate of χ^2 divergence converges to the true value. Although this approximation is well motivated, there is a subtle, but significant, drawback to this approach. The problem lies in the fact that the observed data samples arise from a mixture density. In particular, only a fraction of the pixels in the image cover the vehicle to be classified. The rest of the pixels cover some form of clutter such as the terrain, trees, or other nuisance scatterers. Thus, the density generating the observed data samples can be written as the mixture

$$f_l = ag_c + (1-a)g_l$$

where g_l represents the density solely for the target l, g_c represents the density for clutter, and a represents the amount of clutter contamination in the observed data. In many instances, clutter constitutes a significant fraction of the mixture density. For example, the MSTAR data used at the end of this chapter has a contamination factor on the order of 90%. Thus, most of the sample points that DeBonet uses to estimate the χ^2 divergence correspond to clutter.

This misallocation of resources in the approximation motivates this thesis to take an alternative approach to estimating the KL and χ^2 divergences. The approach used here is a Riemannian approximation to the integral instead of a LLN-motivated approximation. In particular, instead of evaluating the summation at reference data sample points, a pre-determined set of points called *evaluation points* are used. The evaluation points are generated using a modification of the k-means algorithm[3][27] to produce a collection of points that are approximately uniformly spaced. Letting k denote the number of evaluation points to choose, the steps in the procedure are illustrated in the flow diagram in Figure 4.3. First, k evaluation points are initialized in Step 1 by setting them to randomly chosen training vectors. Regions are then defined in Step 2 by assigning each training vector to the closest evaluation point where "closeness" is measured in terms of the supremum norm since a hyper-rectangle is used for the density estimation. Each of the evaluation points is then redefined in Step 3 to be the center of its corresponding region, i.e. the center of the smallest bounding box containing all

[2]DeBonet uses the χ^2 divergence instead of the KL divergence as his score function, but his approximation method is easily adapted to approximating KL divergence.

[3]The k-means algorithm is also known as the Generalized Lloyd Algorithm.

the training vectors in that region. A distortion measure is then computed in Step 4 as the root-mean-square (RMS) of the sizes of the regions. A stopping criterion is then checked in Step 5. If the relative change in distortion from the previous iteration has decreased by less than some amount δ, then the algorithm stops, otherwise it returns to Step 2 and continues.

Step 1:
Initialize k region centers

Step 2:
Define new regions by assigning each training feature vector to the nearest center.

Step 3:
Compute the center of each region as the center of the smallest bounding box containing all points in that region

Step 4:
Compute the distortion as the RMS of the region sizes

Step 5:
Is the relative change in distortion $< \delta$?

no

yes

Stop
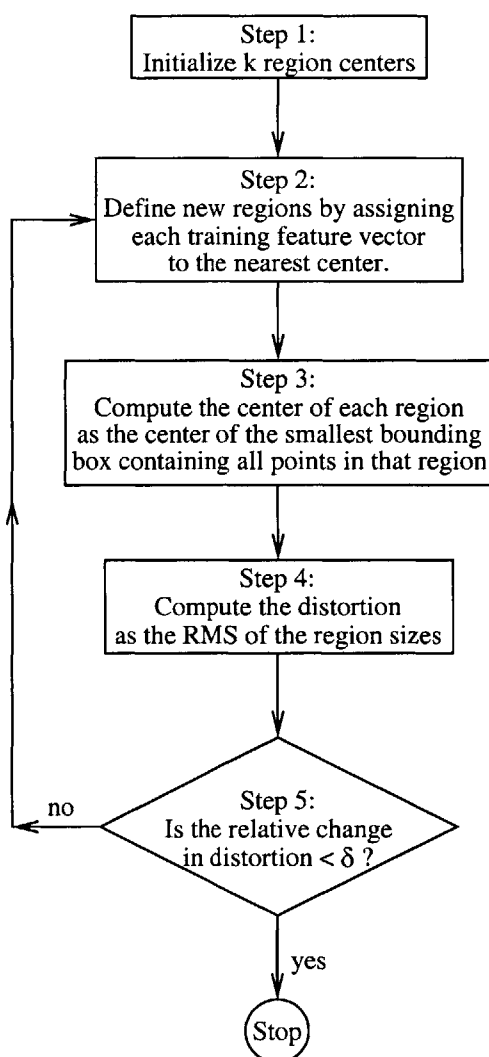
Figure 4.3. Flow diagram illustrating the clustering procedure used to generate a uniform sampling over the union of the supports of the reference densities.

The intuition underlying the algorithm is to adjust the sizes of the box regions until the regions are all the same size, i.e. the evaluation points are uniformly spaced. If there is a region larger than the surrounding ones, then the algorithm will reassign the training vectors that are most distant from the center of the large region to neighboring regions

where the vectors would be closer to a region center. The effect of this reallocation is to decrease the size of the larger region at the cost of increasing the size of the smaller region.

This procedure differs from the standard k-means algorithm in Step 3 and in the distortion measure used. The k-means algorithm computes the *centroid* of each region instead of the *center*. The centroid of a region is the average of the vectors in the region and can be thought of as the center of mass in contrast to the geometric center. The distortion measure use by the k-means algorithm is the RMS of the distances from each training vector to the nearest centroid. The k-means algorithm produces a sampling that is proportional to the underlying density of the training data which results in many closely spaced samples where the density is large and few sparse samples where the density is small. As with DeBonet's χ^2 approximation, such a proportional allocation of evaluation points is undesirable in light of the heavy clutter contamination in the mixture distribution producing the observed data samples.

An advantage of the proposed KL and χ^2 approximation procedure over DeBonet's technique, is that DeBonet's approach may tend to use an excessive number of evaluation points. DeBonet approximates the integral with a summation evaluating the summand at each reference data point. However, for large images, this produces an excessively fine grid. In particular, oversampling smooth densities does not lead to significantly better approximations of the integral but does increase computational cost. However, because the approach proposed in this chapter uses a fixed number k of evolution points, the proposed technique does not oversample in large data sets. Further computational savings over DeBonet's approximation are achieved due to the invariance of the set of evaluation points in the proposed approach. As described in Appendix A, DeBonet uses a set of evaluation points which varies with the hypothesis to which the test data set is compared. Thus, the density for the test data set needs to be estimated several times (once for each hypothesis) in DeBonet's proposed classifier. To illustrate the degree of the computational savings afforded by the evaluation points, the computational orders of the sub-aperture classifier and DeBonet's classifier are compared. For both approaches, the computation is dominated by the density estimation of the test data. Letting N_{test} denote the number of different test data sets, i.e. the number of times the classifier will be run, the computational orders of the two approaches can be broken down as follows. For the approach proposed in this chapter, density estimation for a single point is an $\mathcal{O}(N_1 \log(N_2))$ operation due to the tree structure in cross-range. Evaluating the density at k points for each test data set leads to an

$$\mathcal{O}(N_{test} k N_1 \log(N_2))$$

algorithm to classify all the data. For DeBonet's approach, density estimation for a single point is in an $\mathcal{O}(\log(N_1 N_2))$ operation due to the tree structure in both down-range and cross-range. The density is then estimated for the test data at $N_1 N_2$ points. Furthermore, this set of $N_1 N_2$ points varies with the reference data set to which the

test data set is compared thus leading to an

$$\mathcal{O}(N_{ref}N_{test}N_1N_2 \log(N_1N_2)) \tag{4.6}$$

algorithm to classify all the data. Noting both the omission of the factor N_{ref} and the replacement of N_1N_2 by k (which can be much less than $N_1N_2)^4$, significant computational savings are achieved by using the approximation technique proposed in this section.

■ 4.2.5 Kernel Width Selection

Selection of the kernel width parameter in density estimation is a classic problem. There have been many approaches to address this issue which are based on cross-validation or leave-one-out techniques[5] that minimize a cost criterion such as mean-integrated squared error or entropy [50, 26, 22, 32]. The goal of such methods is to characterize accurately the density as a whole. There is, however, a subtle difference between density estimation for the purpose of characterizing a source distribution and density estimation for the purpose of discrimination. To illustrate this point, consider the two densities f and g depicted in Figure 4.4. If one is only concerned with generally characterizing each source distribution, then for both densities, accurately estimating the region around the mode of the densities is clearly important. However, if the task is to differentiate the behavior of these two densities, then this high probability region is not of much importance as the densities are quite similar in this region. The tails of the densities, in contrast, are very significant since this is where the distinguishing information lies. The importance of such low probability regions is explicit in the KL and χ^2 divergences where the largest contributions come when one of the densities takes on small values. This example is not academic, as constant kernel width density estimators, i.e. those which do not vary the kernel width with the local sparsity of the data, must trade-off estimator variance in the tails for estimator bias in the mode as the kernel width is increased[6]. Such trade-offs will effect the optimal kernel width differently depending on how the resulting density estimates are to be used.

This thesis proposes the following discrimination-based algorithm to determine kernel width[7]. The objective of the algorithm is to find the kernel width parameter h which

---

[4]There is also a computational increase replacing $\log(N_1N_2)$ in DeBonet's approach with $N_1\log(N_2)$ for the algorithm proposed in this chapter, but this change is due to the feature set used and not the integral approximation method.

[5]Cross-validation and leave-one-out[22, 26] are methods used to evaluate the efficacy of a learning rule. In cross-validation, one partitions the data into two sets and uses one set for the "learning" process and the other to estimate the cost of using the learned variables. In leave-one-out, one uses all but one sample for the learning process and calculates the resulting cost for the omitted sample. This partitioning scheme is repeatedly performed omitting a different sample each time. The overall cost is then estimated by averaging the costs from each iteration.

[6]The reader is referred to Silverman[50] for more details on the bias-variance trade-off in density estimation.

[7]At the time this algorithm was developed, the author was not aware of other discrimination-motivated kernel width selection procedures for density estimation. However, since that time, it has
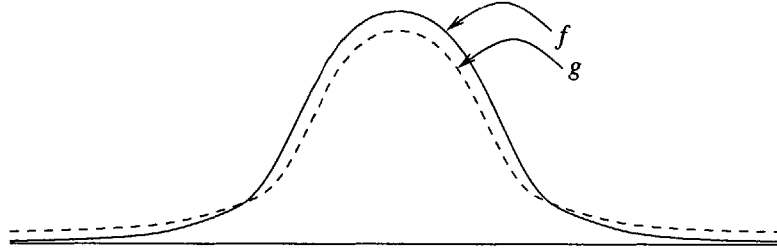
Figure 4.4. Two densities used to illustrate the difference in emphasis between traditional density estimation and density estimation for the purpose of discrimination.

minimizes the probability of false alarm for the decision rule, i.e. Eq. (4.3) or (4.4), under the constraint that a minimum probability of detection is achieved. Towards this end, the coordinate descent algorithm outlined in the flow diagram of Figure 4.5 is employed. The algorithm starts by initializing the set of kernel widths $h = [h^{(0)}, \ldots, h^{(M)}]$ for each scale and setting the scale over which to optimize to $m = M$. Step 2 then adjusts $h^{(m)}$ to minimize the number of false alarms resulting from the decision rule in Eq. (4.3) or (4.4) while maintaining a pre-specified probability of detection. Steps 3 and 4 then decrement m and test if the previous iteration optimized over the finest scale kernel width $h^{(0)}$. If not, then the algorithm proceeds to Step 2 with the new scale. If the previous iteration optimized over the finest scale, then Step 5 checks if the cycle over all the scales resulted in a significant change in the value of h. If there was a significant change, then the optimization cycle is repeated by proceeding back to Step 1. If there was not a significant change in h, then the algorithm assumes a local minimum has been reached and terminates.

■ 4.3 Experimental Results

This section presents empirical results of the algorithms described in this chapter applied to the public release MSTAR data set[1]. This data set contains SAR images generated with an azimuthal span of 2.8°, radar center frequency of 9.6GHz, and a bandwidth of 0.59GHz producing a resolution of one foot in both range and cross-range. A more detailed description of the MSTAR data set can be found in Appendix B.

■ 4.3.1 Kernel Width

Results on the kernel width selection technique described in Section 4.2.5 are presented in this section. The technique is a cross-validation approach where the training data set consists of[8] BMP-2's (c21) and T-72's (132) at a 17° depression, and the testing data set consists of the same vehicles but imaged at a 15° depression. The training

---

come to the author's attention that Vapnik[54, 55] has developed such a criterion based on the principle of Structural Risk Minimization.

[8]Serial numbers for vehicles are placed in parentheses following the vehicle type.
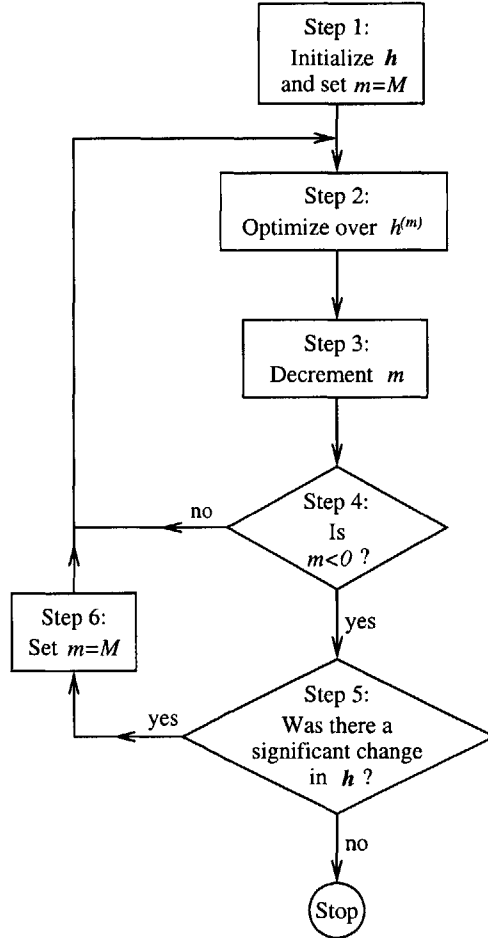
Figure 4.5. Flow diagram illustrating the kernel width selection procedure.

data is also used to determine $k = 512$ evaluation points using the method described in Section 4.2.4. Figure 4.6 displays the number of false alarms using the KLD rule in Eq. (4.3). Each plot in this figure optimizes the kernel width $h^{(m)}$ for a particular scale m while holding the kernel widths for the other scales fixed. The plots displayed are for the final iteration of the algorithm which has settled to a local minimum. The resulting value for h in this case is

$$h_{KL} = [5.3, 2.6, 17.4].$$ (4.7)

Figure 4.7 displays the number of false alarms using the χ^2 rule in Eq. (4.4) as a function of kernel width. Again, the displayed plots are for the final iteration of the algorithm which has settled to the local minimum

$$h_{\chi^2} = [15.0, 14.6, 17.2].$$ (4.8)
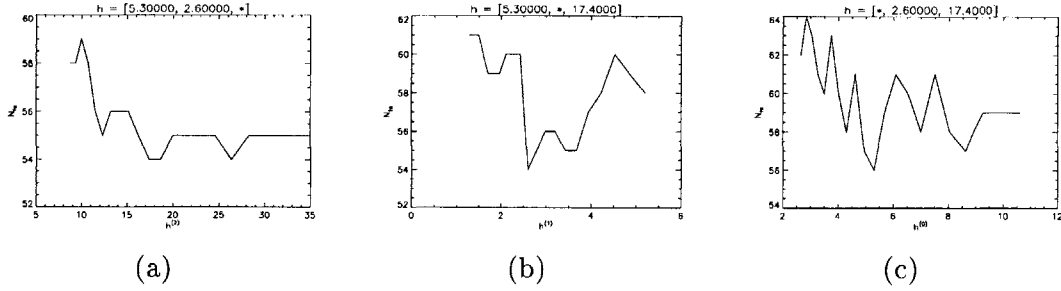
Figure 4.6. Coordinate descent optimization of the kernel width using the KLD rule in Eq. (4.3). (a) Optimizing over the coarsest scale $h^{(2)}$ with $h^{(0)} = 5.3$ and $h^{(1)} = 2.6$. (b) Optimizing over $h^{(1)}$ with $h^{(0)} = 5.3$ and $h^{(2)} = 17.4$. (c) Optimizing over the finest scale $h^{(0)}$ with $h^{(1)} = 2.6$ and $h^{(2)} = 17.4$.

Reasons for the differences between the resulting values of h using the KLD and χ^2 rules are discussed after the presentation of the ROC curves in Section 4.3.2.
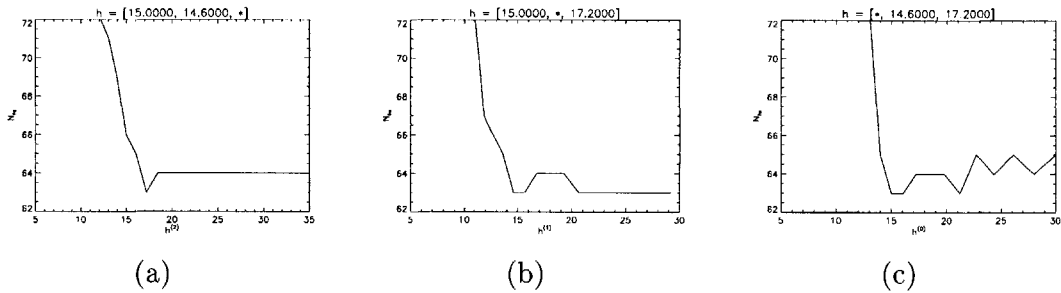
Figure 4.7. Coordinate descent optimization of the kernel width using the χ^2 rule in Eq. (4.4). (a) Optimizing over the coarsest scale $h^{(2)}$ with $h^{(0)} = 15.0$ and $h^{(1)} = 14.6$. (b) Optimizing over $h^{(1)}$ with $h^{(0)} = 15.0$ and $h^{(2)} = 17.2$. (c) Optimizing over the finest scale $h^{(0)}$ with $h^{(1)} = 14.6$ and $h^{(2)} = 17.2$.

■ 4.3.2 ROC curves

Receiver operating characteristic (ROC) curves are used in this section to evaluate the ability of the algorithm proposed in this chapter to discriminate different targets. The reference data for the ROC detection experiments consist of the following vehicles at a 17° elevation: BMP-2 (c21), BTR-70 (c71), and T-72 (132). The testing data consist of the following vehicles at a 15° elevation angle: BMP-2 (9563 and 9566), BTR-70 (c71), T-72 (812 and s7), 2S1 (b01), DRBM-2 (E-71), D-7 (9v13015), T-62 (A51), ZIL-131 (E-12), and ZSU-23/4 (d08). The serial numbers for the testing vehicles are different than those for the training vehicles with the exception of the BTR-70 (c71) in which case there is only one serial number available in the MSTAR data set. It should be noted that although the cross-validation selection of the kernel width parameter h

also involves 15° elevation data, these data are not part of the test data for the ROC detection experiments because different serial numbers are used.

The hypothesis l in which we are interested is the type of the vehicle in the SAR image, e.g. BMP-2, BTR-70, or T-72. The characteristics of the image data, however, vary significantly with respect to the azimuthal pose of the vehicle. This variability may be generated by individual scatterers (i.e. anisotropic scatterers) or by the interaction among several scatterers (e.g. interference, resonance, or occlusion). Regardless of the source, the resulting variability leads to the same vehicle appearing significantly different under various orientations which needs to be accounted for in the model[9]. For example, a BMP-2 tank imaged at a cardinal angle, i.e. at $0°, 90°, 180°$, or $270°$ azimuth, exhibits a highly specular response associated with a front, side, or rear panel that is considerably weaker at other orientations. If one were to model this vehicle with a single pdf (i.e. one not using orientation as a parameter), then one would lose the ability to identify mutually exclusive phenomena in azimuth. To illustrate the drawback of an orientation independent model, note that according to such a model, it would possible to observe the broadside flash from the front, sides, and back of the BMP-2 all in the same image. Such an observation, however, is not physically possible due to the aspect dependence which the model has ignored. To address this issue of capturing dependencies across all azimuths, reference images at particular orientations are used to form individual hypotheses. Each such image is associated with a composite hypothesis composed of vehicle type and azimuthal pose. Performing the hypothesis test in this fashion has the spirit of a generalized likelihood ratio test (GLRT). In fact, if the decision rule utilized likelihood, instead of KL or χ^2 divergence, then this composite hypothesis test would be equivalent to a GLRT. For the experiments presented in this section, the azimuthal coordinate is sampled every 36° for each type of vehicle in the reference data, thus producing a hypothesis space with a total of $N_{ref} = 3 \cdot 10 = 30$ composite hypotheses. The hypothesized azimuthal orientation, however, is discarded and only the vehicle classification is preserved in the final results.

ROC detection curves plotting probability of detection versus probability of false alarm are presented in Figure 4.8(a) for the BMP-2, (b) for the BTR-70, and (c) for the T-72. The results of the KLD rule using the kernel width parameter h_{KL} in Eq. (4.7) are reported with the solid curve in each plot. The results of the χ^2 rule using the kernel width parameter h_{χ^2} in Eq. (4.8) are reported with the dashed curve in each plot. The χ^2 rule clearly outperforms the KLD rule in these ROC curves. To demonstrate that the difference in performance is due to the divergence measure used and not the choice of the kernel width parameter, the KLD rule is reapplied with the kernel width parameter in Eq. (4.8) used with the χ^2 rule. The ROC curves for this modified rule are plotted with the dash-dotted line in each plot. As should be expected, the use of this mismatched kernel width results in poorer performance.

---

[9]While the sub-aperture feature set models azimuthal dependencies, it does so only over the observed azimuth in the training data, i.e. each SAR image. The problem here relates to capturing these dependencies over the entire 360° of azimuth.
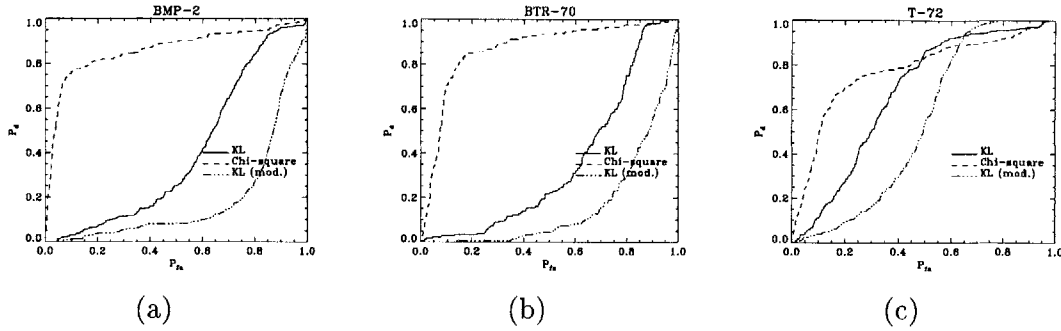
Figure 4.8. ROC detection curves for the (a) BMP-2, (b) BTR-70, and (c) T-72 using the KLD rule with h_{KL} (solid curve), the χ^2 rule with h_{χ^2} (dashed curve), and the KLD rule with with h_{χ^2} (dashed-dotted curve).

Given the kernel width optimization plots in Figures 4.6 and 4.7, it is surprising that the use of the χ^2 rule outperforms the use of the KLD rule in the ROC curves of Figure 4.8. In particular, the use of the KLD rule produces fewer false alarms than the χ^2 rule in the cross-validation algorithm used to determine h. The opposite behavior, however, is observed in the ROC plots in Figure 4.8 where the χ^2 rule produces fewer false alarms. We propose the following explanation for this seemingly contradictory behavior. Recall that the cross-validation learning of h uses the same serial numbers for the BMP-2 and T-72[10]. Because the exact same vehicle[11] is used in the training and testing cross-validation data sets, the similarity in the observed responses is likely to be artificially high. The ROC experiments, however, use vehicles with different serial numbers. Thus, although the observed responses should still be similar due the use of a common vehicle class, there is more intra-class variability due to the use of a different serial number. The higher image similarity of vehicles with the same serial number, as compared to different serial numbers, in the MSTAR data set has been documented by Owirka et. al.[44]. We thus believe on the basis of Figures 4.6, 4.7, and 4.8 that the χ^2 rule is more robust than KLD rule to intra-class deviations. The lower sensitivity of the χ^2 rule to these deviations is consistent with the larger kernel values produced for h_{χ^2} in Eq. (4.8) as compared to h_{KL} in Eq. (4.7). In particular, the KL divergence exploits the higher similarity of the training and testing data to drive h towards smaller values to better distinguish the BMP-2 *serial number c21* and T-72 *serial number 132* which is not equivalent to distinguishing BMP-2's and T-72's in general.

We also believe the over-sensitivity of the KLD rule to be the cause for the non-convex ROC curves for the BMP-2 and BTR-70. In fact, for some values of P_d, the KLD rule performs worse than random guessing for these vehicles. Because the KLD rule

---

[10]Note that because different imaging elevation angles are used for the training and testing data sets, there is no overlap in the two sets in terms of the samples used.

[11]The vehicle used in the training and testing data sets is the same physical vehicle and not just the same type of vehicle.

is not robust to intra-class variations, the rule is trying to detect the specific vehicle (i.e. serial number) that is used in the learning process. The rule, thus, incorrectly declares other instances of the target vehicle as a "none of the above" classification because of the intra-class deviations.

To benchmark the performance of the algorithm proposed in this chapter, ROC detection curves are plotted for the χ^2 rule using the sub-aperture feature set, DeBonet's wavelet-based classifier[20], and the Wright Patterson Air Force Base (WPAFB) standard template matching algorithm[56]. DeBonet's wavelet-based classifier and the sub-aperture-based classifier are comparable in performance, and both techniques outperform the baseline template matching algorithm. Although the sub-aperture-based classifier does not clearly outperform DeBonet's wavelet-based classifier, the performance of the former clearly demonstrates that anisotropy information is useful for SAR ATR. The result is even more impressive given the fact that the MSTAR data set examined has a relatively narrow[12] 2.8° aperture over which to observe anisotropic phenomena.
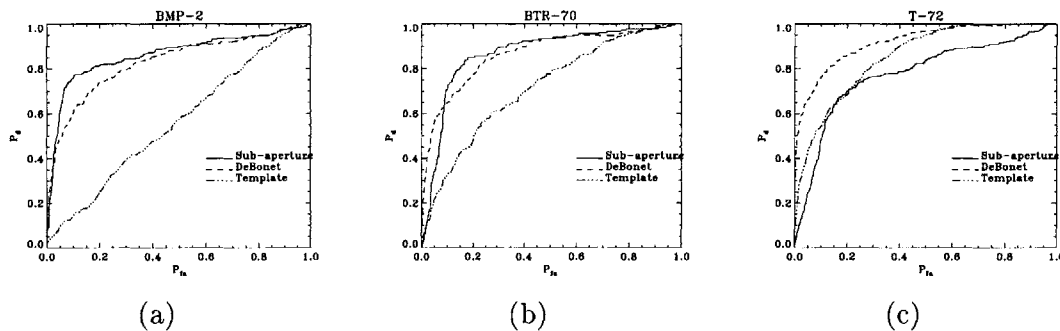
Figure 4.9. ROC detection curves for the (a) BMP-2, (b) BTR-70, and (c) T-72 using the χ^2 rule with h_{χ^2} (solid curve), DeBonet's algorithm (dashed curve), and the WPAFB template matching algorithm (dashed-dotted curve).

---

[12]Figure 3.4 in Chapter 3 illustrates just how narrow this aperture is. The azimuthal response of the 1m×1m flat-plate has a null-to-null width of about 2°. Thus, only large scatterers will produce strongly observable anisotropy in the narrow=aperture MSTAR data. In wide-aperture data, however, the anisotropy from smaller scatterers are more observable and should enhance the utility of the classifier.

Chapter 5

Sub-aperture Anisotropy Attribution

The previous chapter explored the use of the sub-aperture pyramid in a nonparametric learning algorithm to recognize targets based upon their anisotropy behavior. However, besides the use of the sub-aperture feature set, little guidance based on intuition and understanding of radar scattering physics could be provided to the algorithm. This chapter takes a different approach to exploit scatterer anisotropy for SAR ATR. In particular, a feature that characterizes the azimuthal location and concentration of unimodal scattering is extracted from the data. Scattering models such as the PO model and the GTD predict that for many scatterers, there is an inverse relationship between the physical size and the azimuthal energy concentration of the scatterer[33, 48]. The responses for several common canonical scatterers are given in Section 2.3. Not only is there an inverse relation between scaling in the azimuthal and image domains, but many canonical scatterers also produce a unimodal response along the aperture for which azimuthal energy concentration is a natural first-order characterization. The proposed anisotropy feature may thus be used to provide insight about the physical structure of the object under investigation thereby aiding target classification.

The high level problem setting considered here is the same as that in the previous chapter. In particular, we are presented with a single image of SAR data and wish to exploit the anisotropy exhibited by the target to aid the classification process. Anisotropy is exploited in this chapter by attributing the degree of anisotropy, i.e. azimuthal concentration, which conveys information about scatterer size. The basis for this anisotropy attribution is the sub-aperture scattering model used by Chaney et. al.[12] and Allen et. al.[2]. This model is preferred over more detailed models, such as those based on PO and the GTD, because it provides a simple, yet intuitive, characterization of azimuthal scattering. Furthermore, robust estimation of the parameters in the more complicated PO and GTD models is especially difficult in the presence of interfering scatterers given a limited amount of data, i.e. one image per target. Although based on the sub-aperture model proposed by Chaney and Allen, the work presented in this chapter significantly extends their efforts. These extensions include: improving the model for the sub-aperture measurements, accounting for the varying resolution cell sizes among the sub-apertures, modeling the interfering effects of neighboring scatterers, and developing a nonlinear decision-directed sub-aperture selection procedure that preserves high imaging resolution.

The remainder of this chapter is organized as follows. Section 5.1 presents the measurement model and resulting hypothesis test for an isolated scatterer. Section 5.2 extends the work to account for multiple scatterers to remove the interference from neighboring scatterers. Section 5.3 then presents a telescopic, decision-directed modification to the testing procedures presented in Sections 5.1 and 5.2 which improves robustness to interfering scatterers and provides simultaneous high azimuthal resolution and high imaging resolution. Section 5.4 concludes the chapter by applying the new anisotropy attribution to MSTAR data.

■ 5.1 Anisotropy Attribution – Isolated Scatterer Model

The sub-aperture pyramid is introduced in Chapter 3 which emphasizes the structure of the pyramid and how anisotropy is conveyed through the sub-aperture reflectivities. This chapter extends the use of the pyramid to also form a set of anisotropy models. In particular, the sub-aperture scattering models proposed by Chaney and Allen are used to formulate a hypothesis test for anisotropy where the hypotheses are drawn directly from the sub-aperture pyramid.

■ 5.1.1 Observation Model

The anisotropy hypothesis tests presented in this chapter are for a pre-specified scattering location. The location could come from a peak extraction process (resulting in peak scatterer anisotropy attribution) or a pre-specified grid of points (producing an image of the exhibited anisotropy). As is the case for calculating sub-aperture reflectivity in Section 3.2, the dependence of the location parameter is removed through the use of the azimuthal response $a(\omega)$ which corresponds to a range-compressed, but not cross-range compressed, 1-D signal demodulated for the specified cross-range location.

As discussed in Section 2.3, the sub-aperture scattering model used by Chaney and Allen models the azimuthal response as a constant over a sub-aperture and zero elsewhere. The set of candidate sub-apertures in this chapter is restricted to those composing the sub-aperture pyramid. In particular, each sub-aperture $S_{m,i}$ in the pyramid is used to define an associated sub-aperture anisotropy hypothesis $H_{m,i}$ whose modeled response may be written as

$$H_{m,i}: \quad a(\omega) = A1_{S_{m,i}}(\omega). \tag{5.1}$$

where A is the complex scatterer reflectivity and $1_{S_{m,i}}(\omega)$ is the indicator function for the sub-aperture $S_{m,i}$. This model is clearly an idealization of anisotropic scattering, but because we are only seeking a general characterization of anisotropy, this model serves our purpose. Although these hypotheses are used to measure anisotropy, the ideal isotropic scattering hypothesis is included in the hypothesis set since the full-aperture is included in the sub-aperture pyramid. The set of all possible hypotheses associated with the sub-aperture pyramid will be denoted as $\mathcal{H}$.

A reasonable choice of observations to test these hypotheses would be the set of all the measured sub-aperture reflectivities $\{q_{m,i}\}$ where $q_{m,i}$ is given in Eq. (3.1) as

$$q_{m,i} = \int_{S_{m,i}} a(\omega) \; d\omega \,. \tag{5.2}$$

From this definition of the $q_{m,i}$ and the partition property on the pyramid in Section 3.1, it is sufficient to consider the vector $q_M \triangleq \{q_{m,i}\}_{m=M}$ of measurements from the smallest sub-apertures. This vector is sufficient because any sub-aperture reflectivity $q_{m,i}$ can be computed from q_M by summing all the $q_{M,j}$ for which the associated $S_{M,j}$ form a partition of $S_{m,i}$, i.e.

$$q_{m,i} = \sum_{j \left| S_{M,j} \in \mathcal{P}(m,i) \right.} q_{M,j} \,. \tag{5.3}$$

For example, in the three-level, half-overlapping, half-aperture pyramid in Figure 3.3(b), $q_{1,1} = q_{2,2} + q_{2,4}$ since $S_{2,2}$ and $S_{2,4}$ form a disjoint union of $S_{1,1}$. Because q_M is a sufficient statistic, this vector is taken as the observation on which the anisotropy hypothesis test is based.

The value of the observation vector q_M under hypothesis $H_{m,i}$ with a normalized reflectivity of $A = 1$ is denoted as $b(m,i)$, and the jth element of $b(m,i)$ is given by

$$
\begin{aligned}
b(m,i)_j &= \int_{S_{M,j}} 1_{S_{m,i}}(\omega) \; d\omega \\
&= \lambda(S_{M,j} \cap S_{m,i}),
\end{aligned}
\tag{5.4}
$$

where λ denotes Lebesgue measure. Thus, the observed response is proportional to the amount of overlap between the sub-aperture $S_{M,j}$ used to measure the reflectivity and the sub-aperture $S_{m,i}$ used to model the azimuthal response.

Up to this point, noise in the measurements has been omitted. The noise is now modeled as an additive white process along the aperture. In particular, the scattering model conditioned on anisotropy hypothesis $H_{m,i}$ is

$$q_{M,j} = \int_{S_{M,j}} \left[A 1_{S_{m,i}}(\omega) + \eta(\omega) \right] d\omega \,, \tag{5.5}$$

where $\eta(\omega)$ is circularly complex white Gaussian noise with spectral density $2\sigma^2$. Eq. (5.5) leads to the model

$$q_M = Ab(m,i) + \varepsilon, \quad \text{with } \varepsilon \sim \mathcal{N}(0, 2\sigma^2 \Lambda), \tag{5.6}$$

where ε is the measurement noise in the smallest sub-apertures and Λ is the normalized noise covariance matrix inherited from the sub-aperture pyramid. From Eq. (5.5), it follows that the noise ε in Eq. (5.6) is characterized as a zero-mean circularly complex

Gaussian. The covariance between elements of ε are dictated by the amount of sub-aperture overlap. In particular, the elements of the normalized matrix Λ are given by the amount of overlap between the sub-apertures used to generate the measurements, i.e.

$$[\Lambda]_{i,j} = \lambda(S_{M,i} \cap S_{M,j}). \tag{5.7}$$

For the case of the disjoint pyramid, Λ is the a scalar multiple of the identity matrix, i.e. $\Lambda = 2^{-M}I$. For the case of the half-overlapping half-aperture pyramid, Λ is the Toeplitz matrix

$$\Lambda = \frac{1}{2^M} \begin{bmatrix} 1 & .5 & & \\ .5 & \ddots & \ddots & \\ & \ddots & \ddots & .5 \\ & & .5 & 1 \end{bmatrix}. \tag{5.8}$$

■ 5.1.2 Hypothesis Test Example

Prior to developing the anisotropy hypothesis test in its general form, the generalized log-likelihood ratio (GLLR) is determined for the special case of a three-level, disjoint, half-aperture pyramid as an example. For simplicity, only three of the sub-aperture hypotheses are examined. These hypotheses are the full-aperture hypothesis $H_{0,0}$, the left-most half-aperture hypothesis $H_{1,0}$, and the left-most quarter-aperture hypothesis $H_{2,0}$. The model for the observation vector q_2 for each of these hypotheses is the Gaussian random vector described in Eq. (5.6). Note that there is the unknown reflectivity parameter A in the observation model. To address this unknown parameter, a GLLR is used which replaces A with its maximum likelihood (ML) estimate under each hypothesis, i.e. for hypothesis $H_{m,i}$, A is taken as the matched-filter estimate $A = q_{m,i}/L_{m,i}$. The generalized likelihood $\ell_{m,i}$ for each of the three hypothesis can then be specified as follows

$$\ell_{0,0} = p\left(q_2 | H_{0,0}, A = q_{0,0}\right)$$

$$= \frac{1}{|2\pi\sigma^2|^{D/2}} \exp\left[\frac{-1}{4\sigma^2}\left(q_2 - \frac{q_{0,0}}{4}[1,1,1,1]^T\right)^H \Lambda^{-1} \left(q_2 - \frac{q_{0,0}}{4}[1,1,1,1]^T\right)\right]$$

$$= \frac{1}{|2\pi\sigma^2|^{D/2}} \exp\left[\frac{-1}{16\sigma^2}\left(\left|q_{2,0} - \frac{q_{0,0}}{4}\right|^2 + \left|q_{2,1} - \frac{q_{0,0}}{4}\right|^2 + \left|q_{2,2} - \frac{q_{0,0}}{4}\right|^2 \right.\right.$$

$$\left.\left. + \left|q_{2,3} - \frac{q_{0,0}}{4}\right|^2\right)\right]$$

$$\ell_{1,0} = p\left(\boldsymbol{q}_2 \middle| H_{1,0}, A = 2q_{1,0}\right)$$

$$= \frac{1}{|2\pi\sigma^2|^{D/2}} \exp\left[\frac{-1}{4\sigma^2}\left(\boldsymbol{q}_2 - \frac{q_{1,0}}{2}[1,1,0,0]^T\right)^H \Lambda^{-1}\left(\boldsymbol{q}_2 - \frac{q_{1,0}}{2}[1,1,0,0]^T\right)\right]$$

$$= \frac{1}{|2\pi\sigma^2|^{D/2}} \exp\left[\frac{-1}{16\sigma^2}\left(\left|q_{2,0} - \frac{q_{1,0}}{2}\right|^2 + \left|q_{2,1} - \frac{q_{1,0}}{2}\right|^2 + |q_{2,2}|^2 + |q_{2,3}|^2\right)\right]$$

$$\ell_{2,0} = p\left(\boldsymbol{q}_2 \middle| H_{2,0}, A = 4q_{2,0}\right)$$

$$= \frac{1}{|2\pi\sigma^2|^{D/2}} \exp\left[\frac{-1}{4\sigma^2}\left(\boldsymbol{q}_2 - q_{2,0}[1,0,0,0]^T\right)^H \Lambda^{-1}\left(\boldsymbol{q}_2 - q_{2,0}[1,0,0,0]^T\right)\right]$$

$$= \frac{1}{|2\pi\sigma^2|^{D/2}} \exp\left[\frac{-1}{16\sigma^2}\left(|q_{2,1}|^2 + |q_{2,2}|^2 + |q_{2,3}|^2\right)\right]$$

where D is the dimensionality of $\boldsymbol{q}_2$, $(\cdot)^T$ denotes transposition, $(\cdot)^H$ denotes Hermitian transposition, and $p(\cdot)$ is the Gaussian density for the observation vector conditioned on the underlying reflectivity A and anisotropy hypothesis. In each case, the likelihood simply measures the norm of the deviation in the smallest sub-apertures from the expected value under the given hypothesis. For example, under the full-aperture hypothesis, the expected response is to observe one-quarter of the reflectivity in each of the quarter-apertures. Consequently, the likelihood sums the squares of the deviations from this expected value since each sub-aperture measurement $q_{2,i}$ is formed from a disjoint sub-aperture and the measurements are thus modeled as independent.

The generalized likelihoods may be used in a GLLR test as a decision rule. Taking the ratio with respect to the full-aperture generalized likelihood, the GLLR's $\gamma_{m,i}$ can be written as[1]

$$\gamma_{0,0} = 0$$

$$\gamma_{1,0} = \log\left(\frac{p\left(\boldsymbol{q}_2 \middle| H_{1,0}, A = 2q_{1,0}\right)}{p\left(\boldsymbol{q}_2 \middle| H_{0,0}, A = q_{0,0}\right)}\right)$$

$$= \frac{1}{4\sigma^2}\left(\frac{|q_{1,0}|^2}{0.5} - |q_{0,0}|^2\right)$$

---

[1]The calculation involves the use of the partition property, i.e. condition (C) on the pyramid in Section 3.1.

$$\gamma_{2,0} = \log\left(\frac{p\left(q_2\Big|H_{2,0}, A = 4q_{2,0}\right)}{p\left(q_2\Big|H_{0,0}, A = q_{0,0}\right)}\right)$$

$$= \frac{1}{4\sigma^2}\left(\frac{|q_{2,0}|^2}{0.25} - |q_{0,0}|^2\right).$$

There is an interesting transition that occurs in moving from generalized likelihoods to the GLLR's. In particular, it should be noted that the likelihoods $\ell_{m,i}$ involve each of the measurements from the smallest sub-apertures. These measurements manifest themselves in the likelihood via the *incoherent* sum of squared-error magnitudes, i.e. the squared-norm of the deviation from the expected sub-aperture measurements. However, for each GLLR $\gamma_{m,i}$, all the sub-aperture measurements, except $q_{m,i}$ and $q_{0,0}$, cancel out of the expression. In other words, the only sub-aperture measurements that remain in the GLLR are those associated with the sub-aperture hypotheses the GLLR is comparing. In particular, the expression becomes the difference between the energy in each sub-aperture normalized by the sub-aperture length. Thus, if the decision rule is to select the sub-aperture with the highest GLLR, then this rule is equivalent to choosing the sub-aperture with the largest average energy. This interpretation is quite satisfying as it is consistent with the intuition of seeking concentrated unimodal scattering in azimuth.

■ 5.1.3 GLLR

The simple form of the GLLR for the sub-aperture hypotheses in the previous section generalize in a quite straightforward manner. In particular, the GLLR is the difference between the average energy in sub-apertures. To show this fact, however, is quite involved and is the topic of this section. In particular, computation of the generalized likelihoods involve all of the sub-aperture measurements (not just the one sub-aperture measurement as the simplified GLLR may lead one to mistakenly believe). The expression becomes even more complicated when an overlapping sub-aperture pyramid is used because the inverse noise covariance is no longer a scalar multiple of the identity. In fact, the inverse covariance is not even sparse in this case, and thus, *each* of the summands in $\ell_{m,i}$ involves *all* of the smallest sub-aperture measurements.

Although the end expressions for the GLLR's in the previous example are intuitively reasonable, it is somewhat surprising that so many of the sub-aperture terms cancel, even in the more complicated case of an overlapping pyramid. As will be seen in this section, the key to these cancelations is the use of the sub-aperture scattering model which permits the normalized modeled response $b(m, i)$ to be written as a linear function of the noise covariance which ultimately cancels the inverse noise covariance weighting matrix in the calculations.

In order to derive the GLLR for a general sub-aperture pyramid, some notation and simple relations must first be established. Eq. (5.3) provides a representation of larger sub-aperture reflectivity measurements from the set of smallest sub-aperture

measurements due to the pyramid's partition property. It will be more convenient to rewrite Eq. (5.3) as a vector multiplication. To do this, the binary vector $c(\mathcal{P}(m,i))$ is defined. In particular, the components of $c(\mathcal{P}(m,i))$ are defined as

$$c(\mathcal{P}(m,i))_j = \begin{cases} 1 & : \text{ if } S_{M,j} \in \mathcal{P}(m,i), \\ 0 & : \text{ otherwise}, \end{cases} \tag{5.9}$$

i.e. $c(\mathcal{P}(m,i))$ indicates which components of S_M constitute a partition of $S_{m,i}$. Obtaining the sub-aperture reflectivity estimate $q_{m,i}$ from q_M can now be done via

$$q_{m,i} = \sum_j c(\mathcal{P}(m,i))_j q_{M,j}$$

$$= c(\mathcal{P}(m,i))^H q_M. \tag{5.10}$$

Recall from Section 5.1.1 that because of the manner in which the measured reflectivities $q_{m,i}$ are calculated, i.e. integrating over a sub-aperture, the covariance between these two measurements is proportional to the amount of sub-aperture overlap. Similarly, because the boxcar-shaped sub-aperture scattering model is used, the components of the normalized model vector $b(m,i)_j$ are also given by the amount of sub-aperture overlap between $S_{m,i}$ and $S_{M,j}$. Because of the consistency between the reflectivity measurement calculation and the sub-aperture scattering model, the noise covariance and model basis vectors can be written in terms of each other. In particular, it follows from Eq. (5.4) and Eq. (5.7) that

$$\Lambda = \begin{bmatrix} b(M,0) & | & b(M,1) & | & \dots & | & b(M, J_M - 1) \end{bmatrix} \tag{5.11}$$

where J_M is the number of sub-apertures at scale M. From this equation, it follows that

$$b(M,i) = \Lambda e_i \tag{5.12}$$

where e_i is the Euclidean basis vector of appropriate dimension. This relation can be generalized to write $b(m,i)$ for larger sub-apertures in terms of Λ using the vector $c(\mathcal{P}(m,i))$. First, note that for a particular element j of $b(m,i)$,

$$b(m,i)_j = \int_{S_{M,j}} 1_{S_{m,i}}(\omega)\, d\omega \qquad\qquad \text{by definition of } b(m,i)_j$$

$$= \int_{S_{M,j}} \sum_{k|S_{M,k}\in\mathcal{P}(m,i)} 1_{S_{M,k}}(\omega)\, d\omega \qquad\qquad \text{using Eq. (5.3)}$$

$$= \sum_{k|S_{M,k}\in\mathcal{P}(m,i)} \int_{S_{M,j}} 1_{S_{M,k}}(\omega)\, d\omega$$

$$= \sum_{k|S_{M,k}\in\mathcal{P}(m,i)} b(M,k)_j \qquad\qquad \text{by definition of } b(m,i)_j$$

i.e. the normalized model for hypothesis $H_{m,i}$ can be written as a sum of models from partitioned sub-apertures $H_{M,k}$. Thus, for the entire vector

$$
\begin{aligned}
\boldsymbol{b}(m,i) &= \sum_{k \mid S_{M,k} \in \mathcal{P}(m,i)} \boldsymbol{b}(M,k) \\
&= \sum_{k \mid S_{M,k} \in \mathcal{P}(m,i)} \Lambda \boldsymbol{e}_k \qquad \text{using Eq. (5.12)} \\
&= \Lambda \sum_{k \mid S_{M,k} \in \mathcal{P}(m,i)} \boldsymbol{e}_k \\
&= \Lambda \boldsymbol{c}(\mathcal{P}(m,i))
\end{aligned}
\tag{5.13}
$$

thereby obtaining the desired generalization of Eq. (5.12) to larger sub-apertures.

There is one more relation that is useful in the derivation of the GLLR. In particular,

$$
\begin{aligned}
\boldsymbol{b}(m,i)^H \boldsymbol{c}(\mathcal{P}(m,i)) &= \sum_{j \mid S_{M,j} \in \mathcal{P}(m,i)} b(m,i)_j \\
&= \sum_{j \mid S_{M,j} \in \mathcal{P}(m,i)} \int_{S_{M,j}} 1_{m,i}(\omega)\, d\omega \qquad \text{by definition of } b(m,i)_j \\
&= \int_{S_{m,i}} 1_{S_{m,i}}(\omega)\, d\omega \\
&= L_{m,i}
\end{aligned}
\tag{5.14}
$$

which says that the sum of the terms in the modeled response $\boldsymbol{b}(m,i)$ is the integral of the normalized sub-aperture response or equivalently the sub-aperture length.

With the necessary relations in hand, the GLLR for a sub-aperture scattering hypothesis is now derived. Instead of explicitly stating the assumed value of A under each of the hypotheses, for reasons which will become apparent later, it is assumed that the estimated value used for A can be written as $A_{0,0} = \boldsymbol{a}_{0,0}^H \boldsymbol{q}_M$ under $H_{0,0}$ and as $A_{m,i} = \boldsymbol{a}_{m,i}^H \boldsymbol{q}_M$ under $H_{m,i}$. Included in this case are the special instances

$$
\begin{aligned}
A_{0,0} &= q_{0,0} = \boldsymbol{c}(\mathcal{P}(0,0))^H \boldsymbol{q}_M \\
&= \text{the full-aperture reflectivity estimate and} \\
A_{m,i} &= q_{m,i}/L_{m,i} = \frac{1}{L_{m,i}} \boldsymbol{c}(\mathcal{P}(m,i))^H \boldsymbol{q}_M \\
&= \text{the hypothesized sub-aperture reflectivity estimate}
\end{aligned}
$$

that are used in the example of Section 5.1.2. The GLLR can now be obtained as

$$
\gamma_{m,i} = \log\left(\frac{p\left(\boldsymbol{q}_M \middle| H_{m,i}, A = A_{m,i}\right)}{p\left(\boldsymbol{q}_M \middle| H_{0,0}, A = A_{0,0}\right)}\right)
$$

$$
= \frac{1}{2}\left[(\boldsymbol{q}_M - \boldsymbol{b}(0,0)A_{0,0})^H \frac{1}{2\sigma^2}\Lambda^{-1} (\boldsymbol{q}_M - \boldsymbol{b}(0,0)A_{0,0}) \right.
$$

$$
\left. - (\boldsymbol{q}_M - \boldsymbol{b}(m,i)A_{m,i})^H \frac{1}{2\sigma^2}\Lambda^{-1} (\boldsymbol{q}_M - \boldsymbol{b}(m,i)A_{m,i}) \right]
$$

$$
= \frac{1}{4\sigma^2}\boldsymbol{q}_M^H\left[\left(I - \boldsymbol{b}(0,0)\boldsymbol{a}_{0,0}^H\right)^H \Lambda^{-1} \left(I - \boldsymbol{b}(0,0)\boldsymbol{a}_{0,0}^H\right) \right.
$$

$$
\left. - \left(I - \boldsymbol{b}(m,i)\boldsymbol{a}_{m,i}^H\right)^H \Lambda^{-1} \left(I - \boldsymbol{b}(m,i)\boldsymbol{a}_{m,i}^H\right) \right]\boldsymbol{q}_M
$$

$$
= \frac{1}{4\sigma^2}\boldsymbol{q}_M^H\left[\Lambda^{-1} - \Lambda^{-1}\boldsymbol{b}(0,0)\boldsymbol{a}_{0,0}^H - \boldsymbol{a}_{0,0}\boldsymbol{b}(0,0)^H\Lambda^{-1} \right.
$$

$$
+ \boldsymbol{a}_{0,0}\boldsymbol{b}(0,0)^H\Lambda^{-1}\boldsymbol{b}(0,0)\boldsymbol{a}_{0,0}^H
$$

$$
- \Lambda^{-1} + \Lambda^{-1}\boldsymbol{b}(m,i)\boldsymbol{a}_{m,i}^H + \boldsymbol{a}_{m,i}\boldsymbol{b}(m,i)^H\Lambda^{-1}
$$

$$
\left. - \boldsymbol{a}_{m,i}\boldsymbol{b}(m,i)^H\Lambda^{-1}\boldsymbol{b}(m,i)\boldsymbol{a}_{m,i}^H \right]\boldsymbol{q}_M
$$

$$
= \frac{1}{4\sigma^2}\boldsymbol{q}_M^H\left[- \boldsymbol{c}(0,0)\boldsymbol{a}_{0,0}^H - \boldsymbol{a}_{0,0}\boldsymbol{c}(0,0)^H + \boldsymbol{a}_{0,0}\boldsymbol{b}(0,0)^H\boldsymbol{c}(0,0)\boldsymbol{a}_{0,0}^H \right.
$$

$$
\left. + \boldsymbol{c}(m,i)\boldsymbol{a}_{m,i}^H + \boldsymbol{a}_{m,i}\boldsymbol{c}(m,i)^H - \boldsymbol{a}_{m,i}\boldsymbol{b}(m,i)^H\boldsymbol{c}(m,i)\boldsymbol{a}_{m,i}^H \right]\boldsymbol{q}_M \qquad (5.15a)
$$

$$
= \frac{1}{4\sigma^2}\boldsymbol{q}_M^H\left[- \boldsymbol{c}(0,0)\boldsymbol{a}_{0,0}^H - \boldsymbol{a}_{0,0}\boldsymbol{c}(0,0)^H + \boldsymbol{a}_{0,0}\boldsymbol{a}_{0,0}^H \right.
$$

$$
\left. + \boldsymbol{c}(m,i)\boldsymbol{a}_{m,i}^H + \boldsymbol{a}_{m,i}\boldsymbol{c}(m,i)^H - L_{m,i}\boldsymbol{a}_{m,i}\boldsymbol{a}_{m,i}^H \right]\boldsymbol{q}_M \qquad (5.15b)
$$

$$
= \frac{1}{4\sigma^2}\boldsymbol{q}_M^H\left[[\boldsymbol{a}_{0,0} - \boldsymbol{c}(\mathcal{P}(0,0))][\boldsymbol{a}_{0,0} - \boldsymbol{c}(\mathcal{P}(m0,0))]^H - \boldsymbol{c}(\mathcal{P}(0,0))\boldsymbol{c}(\mathcal{P}(0,0))^H \right.
$$

$$
- \left[\sqrt{L_{m,i}}\boldsymbol{a}_{m,i} - \frac{1}{\sqrt{L_{m,i}}}\boldsymbol{c}(\mathcal{P}(m,i)) \right]\left[\sqrt{L_{m,i}}\boldsymbol{a}_{m,i} - \frac{1}{\sqrt{L_{m,i}}}\boldsymbol{c}(\mathcal{P}(m,i)) \right]^H
$$

$$
\left. + \frac{1}{L_{m,i}}\boldsymbol{c}(\mathcal{P}(m,i))\boldsymbol{c}(\mathcal{P}(m,i))^H \right]\boldsymbol{q}_M. \qquad (5.15c)
$$

where Eq. (5.15a) follows from Eq. (5.13) and $L_{0,0} = 1$, and where Eq. (5.15b) follows from the relation in Eq. (5.14).

As in the example in Section 5.1.2, a natural choice for A is the ML estimate under the corresponding hypothesis, i.e. $A_{m,i} = q_{m,i}/L_{m,i} = \boldsymbol{c}(\mathcal{P}(m,i))^H\boldsymbol{q}_M/L_{m,i}$. Using this

value in Eq. (5.15) leads to

$$
\begin{aligned}
\gamma_{m,i} &= \frac{1}{4\sigma^2} q_M^H \left[\frac{1}{L_{m,i}} c(\mathcal{P}(m,i)) c(\mathcal{P}(m,i))^H \right. \\
&\quad - [\sqrt{L_{m,i}} a_{m,i} - \frac{1}{\sqrt{L_{m,i}}} c(\mathcal{P}(m,i))][\sqrt{L_{m,i}} a_{m,i} - \frac{1}{\sqrt{L_{m,i}}} c(\mathcal{P}(m,i))]^H \\
&\quad \left. - c(\mathcal{P}(0,0)) c(\mathcal{P}(0,0))^H \right] q_M \\
&= \frac{1}{4\sigma^2} q_M^H \left[\frac{1}{L_{m,i}} c(\mathcal{P}(m,i)) c(\mathcal{P}(m,i))^H - c(\mathcal{P}(0,0)) c(\mathcal{P}(0,0))^H \right] q_M \\
&= \frac{1}{4\sigma^2} \left[\frac{1}{L_{m,i}} |q_{m,i}|^2 - |q_{0,0}|^2 \right]
\end{aligned}
\tag{5.16}
$$

thus giving the previously stated result. In particular, the GLLR is proportional to the difference between the average energy in the hypothesized sub-aperture and the full-aperture. Furthermore, the expression for the GLLR does not make explicit use of the smallest sub-aperture measurements q_M except to generate the reflectivity of the involved sub-apertures, i.e. $q_{m,i}$ and $q_{0,0}$. Thus, increasing the number of scales or the amount of sub-aperture overlap in the pyramid does not change the statistic.

■ 5.1.4 Comparison to Chaney and Allen

Using a maximum likelihood approach based on the GLLR in Eq. (5.16), the sub-aperture decision rule can now be taken as selecting the sub-aperture with the maximum average *energy*. This test is very similar to the maximum average *reflectivity* selection criterion used by Allen and Chaney but is not the same in general. In particular, Allen and Chaney maximize the normalized reflectivity, or equivalently its square

$$
\begin{aligned}
\gamma_{m,i}^{'2} = |A_{m,i}|^2 &= \left| \frac{q_{m,i}}{L_{m,i}} \right|^2 \\
&= \frac{1}{L_{m,i}} \frac{|q_{m,i}|^2}{L_{m,i}}.
\end{aligned}
\tag{5.17}
$$

Chaney and Allen use this criterion due to its interpretation as a matched-filter and thus its maximum likelihood interpretation. However, such an interpretation requires the assumption that each sub-aperture estimate have the same variance. This assumption does not hold for the model proposed in Eq. (5.5) for which the variance on a sub-aperture reflectivity estimate $A_{m,i}$ is inversely proportional to the sub-aperture length. Furthermore, the assumption of a variance that is constant with respect to sub-aperture length necessarily implies an awkward correlation structure between sub-apertures (even when they are disjoint). Consider, for instance, the case of the full-aperture reflectivity estimate $A_{0,0} = q_{0,0}$ and the reflectivity estimates, $A_{1,0} = 2q_{1,0}$ and $A_{1,1} = 2q_{1,1}$, from

two disjoint half-apertures. Because the two half-apertures form a disjoint union of the full-aperture, the measured reflectivity can be written as $q_{0,0} = q_{1,0} + q_{1,1}$ or equivalently

$$A_{0,0} = \frac{A_{1,0} + A_{1,1}}{2}.$$

However, if $A_{0,0}$, $A_{1,0}$, and $A_{1,1}$ are to all have the same variance, then it must be the case that $A_{0,0} = A_{1,0} = A_{1,1}$, i.e. that the observed *noisy* response in the left and right half-apertures must be exactly the same. This argument can be repeated for smaller sub-apertures producing the result that each sub-aperture reflectivity must be the same! This clearly imposes an unreasonable noise model on the sub-aperture measurements.

To further demonstrate the inappropriateness of the maximum reflectivity statistic in Eq. (5.17), again consider the case of the full-aperture and the two disjoint half-apertures. In this case, the full-aperture hypothesis can never be chosen[2]! To see this fact, we apply the sub-additivity of norms to the full-aperture reflectivity estimate, i.e.

$$|A_{0,0}| = |q_{0,0}| = |q_{1,0} + q_{1,1}|$$

$$\leq |q_{1,0}| + |q_{1,1}|$$

$$\leq 2\max\{|q_{1,0}|, |q_{1,1}|\}$$

$$= \max\{|2q_{1,0}|, |2q_{1,1}|\}$$

$$= \max\{|A_{1,0}|, |A_{1,1}|\} \tag{5.18}$$

i.e. the full-aperture reflectivity estimate can never exceed the maximum half-aperture estimate. Furthermore, the two may be equal only if the first inequality in Eq. (5.18) holds with equality, i.e. $q_{1,0}$ and $q_{1,1}$ have the same phase or one of them is zero. The same argument can be applied recursively on the sub-aperture pyramid to prove that the maximum sub-aperture reflectivity always comes from one of the smallest sub-apertures for the disjoint pyramid[3]. Thus, without the use of a nonzero threshold in a GLRT, the rule used by Allen and Chaney will always select the maximum degree of anisotropy.

In contrast to the maximum reflectivity criterion, using a ML hypothesis test with the GLLR in Eq. (5.16) results in the decision rule of taking the maximum average energy

$$\gamma_{m,i} = \frac{|q_{m,i}|^2}{L_{m,i}}. \tag{5.19}$$

This rule differs from that in Eq. (5.17) in that the latter has an extra scaling factor of $1/L_{m,i}$ which produces an excessively large statistic for smaller sub-apertures,

---

[2]An exception is the case where the two half-aperture reflectivities have the same phase and the decision rule chooses the full-aperture hypothesis in the case of two equal statistics. However, if phase is modeled with a continuous probability distribution, then such an event occurs with zero probability.

[3]The authors believe the same result holds for overlapping pyramids as well, but do not attempt to prove this point as the unsuitability of the criterion in Eq. (5.17) is already clear.

thus leading to too small an aperture being selected. This excessively small aperture then produces a corresponding excessively coarse imaging resolution due to the inverse relation between aperture length and imaging resolution.

■ 5.1.5 Modified GLLR

Though simple and intuitive, the GLLR in Eq. (5.16) is susceptible to the effects of close proximity neighboring scatterers for which the isolated scatterer model in Eq. (5.5) does not account. To see how neighboring scatterers affect the statistic, recall that reflectivity measurements generated from smaller sub-apertures have a coarser imaging resolution. The sub-aperture anisotropy hypotheses are thus not consistent in terms of the resolution cell contents. As an example, consider an isotropic scatterer with reflectivity A located one resolution cell from the scattering center as shown in Figure 5.1. Because the scatterer is located outside the finest resolution cell, the reflectivity $q_{0,0}$ is largely uninfluenced by the scatterer and thus simply measures noise. The coarser scale half-aperture measurements $q_{m,i}$, however, have a large magnitude because the larger resolution cell includes the scatterer. In particular, if no noise were present, then the observed azimuthal response is the complex exponential illustrated. Integrating over the full-aperture gives a reflectivity measurement of zero, as expected, but integrating over a half-aperture produces a measured reflectivity with magnitude $A/\sqrt{2}$. Use of the GLLR in Eq. (5.16) would thus produce an anisotropic classification, even though no scatterer is present in the finest resolution cell.
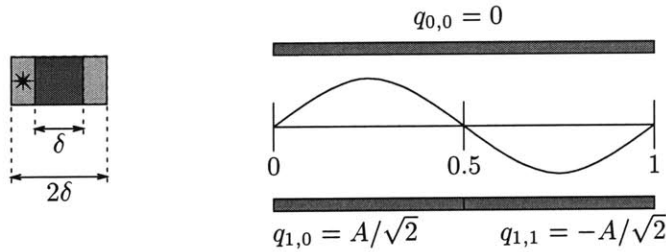
Figure 5.1. Illustration of how an interfering scatterer can cause a false anisotropy classification with the GLLR in Eq. (5.16).

The problem just discussed is a consequence of not modeling the influence of neighboring scatterers. One of the ways in which a neighboring scatterer manifests itself is through the corruption of the estimated reflectivity as the size of the resolution cell varies. To address this problem, instead of choosing the ML estimate $\hat{A} = q_{m,i}/L_{m,i}$ as the unknown reflectivity in the GLLR, we can choose the best reflectivity estimate constrained to lie in the finest resolution cell. In particular, we choose $\hat{A} = q_{0,0}/L_{m,i}$, i.e. the reflectivity estimate using the full-aperture but normalized to account for energy only coming from sub-aperture $S_{m,i}$. Choosing $\hat{A} = q_{0,0}/L_{m,i}$ for *all* hypotheses is

equivalent to using $A_{m,i} = c(\mathcal{P}(0,0))^H q_M / L_{m,i}$ in Eq. (5.15) which produces

$$
\gamma_{m,i} = \frac{1}{4\sigma^2} q_M^H \left\{ \frac{1}{L_{m,i}} c(\mathcal{P}(m,i)) c(\mathcal{P}(m,i))^H \right.
$$

$$
- \frac{1}{L_{m,i}} \left[c(\mathcal{P}(0,0)) - c(\mathcal{P}(m,i)) \right] \left[c(\mathcal{P}(0,0)) - c(\mathcal{P}(m,i)) \right]^H
$$

$$
\left. - c(\mathcal{P}(0,0)) c(\mathcal{P}(0,0))^H \right\} q_M
$$

$$
= \frac{1}{4\sigma^2} \left[\frac{1}{L_{m,i}} |q_{m,i}|^2 - \frac{1}{L_{m,i}} |q_{0,0} - q_{m,i}|^2 - |q_{0,0}|^2 \right]. \tag{5.20}
$$

This statistic is identical to that in Eq. (5.16) except for the extra term comparing the reflectivities $q_{0,0}$ and $q_{m,i}$. Recall that Eq. (5.16) compared the average energy in a sub-aperture to the full-aperture. This modified GLLR accounts for the average energy outside the hypothesized sub-aperture as well. Viewed differently, under hypothesis $H_{m,i}$ and the scattering model in Eqs. (5.5) and (5.6), the values of $q_{0,0}$ and $q_{m,i}$ should simply be noisy perturbations of each other. The new term in Eq. (5.20) penalizes when this is not the case thereby enforcing a consistency of the resolution cell contents. For instance, in the example in Figure 5.1, since the the contribution of each of the half-apertures would be the same, the modified GLLR is equal to zero for all hypotheses, which is reasonable since there is no underlying scattering at the focused location.

■ 5.1.6 Sub-aperture Scattering Model Deviations

The sub-aperture scattering model defined in Eq. (5.1) is a simple model to which real scatterers do not exactly fit. One may question how deviations from this model affect the resulting hypothesis tests. For example, if the scattering has a sinc($\cdot$) dependence in azimuth, then the sidelobes[4] will have a large response for a strong scatterer and may not be well modeled as the additive measurement noise ε in Eq. (5.6). To address this issue, deviations from the sub-aperture scattering model are incorporated into the measurement model. In particular, the underlying *normalized* scattering response is perturbed by white Gaussian noise. This perturbation changes the modeled response in Eq. (5.4) to

$$
\bar{b}(m,i)_j = \int_{S_{M,j}} \left[1_{S_{m,i}}(\omega) + \nu(\omega) \right] d\omega \tag{5.21}
$$

where $\nu(\omega)$ is a white Gaussian process with spectral density $2\rho^2$ and is independent of the measurement noise $\eta(\omega)$. Thus, the modeled measurement vector is now a random vector characterized as

$$
\bar{b}(m,i) \sim \mathcal{N}(b(m,i), 2\rho^2 \Lambda).
$$

---

[4]The term "sidelobes" here refers to those on a sinc-like function in the azimuthal domain and *not* those associated with a neighboring scatterer in the image domain.

This modification results in the new measurement model as

$$H_{m,i} : q_M = A\bar{b}(m,i) + \boldsymbol{\varepsilon}$$
$$= Ab(m,i) + w, \text{ with } w \sim \mathcal{N}\big(0, 2(|A|^2\rho^2 + \sigma^2)\Lambda\big). \tag{5.22}$$

This is essentially the same model as in Eq. (5.6) except that the variance of the noise now depends affinely on the square-magnitude of the underlying scatterer. Furthermore, the variance under each hypothesis scales by the same factor. Thus, the value of the noise variance is relevant only when the GLLR is compared to a threshold other than zero and one wishes to set the threshold based on cost (or prior probability) ratios. This modified measurement model is used throughout the remainder of the thesis.

■ 5.1.7 Experimental Results

To demonstrate the behavior of the anisotropy attribution methods described in this section, the three-level, half-overlapping, half-aperture pyramid is used to attribute synthesized SAR data. A value of $\rho = 0.1$ is used to account for deviations from the sub-aperture scattering model in Eq. (5.21). Anisotropy tests are applied to the synthesized data whose log-magnitude image (formed using the standard image formation process) is displayed in Figure 5.2(a). The position of the radar is to left of the image, and thus cross-range is the vertical dimension. The parameters for the synthesized radar data are set to nominal MSTAR parameters, i.e. an azimuthal span of $2.8°$, center frequency of 9.6GHz, and a bandwidth of 590MHz. The image is composed of 11 scatterers, highlighted by circles in Figure 5.2(b), in circularly complex white Gaussian noise. Each of the scatterers is generated according to the PO model for the flat plate in Eq. (2.16). In particular, the following four scatterer responses are used:

- A flat plate measuring half a resolution cell in cross-range, approximately giving an isotropic response, with a signal-to-noise ratio (SNR)[5] of 30dB. These scatterers are denoted by 'A' in Figure 5.2(b).

- A flat plate measuring two resolution cells in cross-range, approximately giving a half-aperture anisotropic response, with an SNR of 30dB. These scatterers are denoted by 'B' in Figure 5.2(b).

- A flat plate measuring four resolution cells in cross-range, approximately giving a quarter-aperture anisotropic response, with an SNR of 30dB. These scatterers are denoted by 'C' in Figure 5.2(b).

- A flat plate measuring two resolution cells in cross-range, approximately giving a half-aperture anisotropic response, with an SNR of 20dB. These scatterers are denoted by 'D' in Figure 5.2(b).

---

[5]Signal energy is taken as the square-magnitude of the full-aperture reflectivity estimate.

(a)   (b)
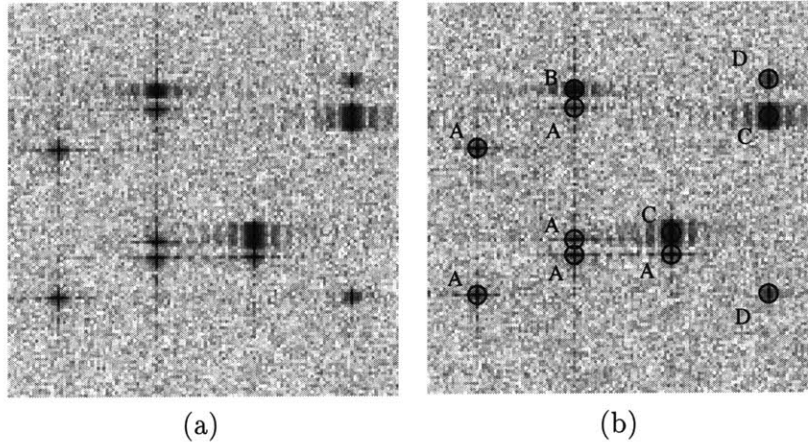
Figure 5.2. Log-magnitude image of the SAR data used to examine the anisotropy attribution methods. (a) Image without scatterer labels. (b) Image with true scatterer labels.

In order to simplify wording, the statistic in Eq. (5.17) used by Chaney and Allen will be called the *reflectivity statistic*, the GLLR statistic in Eq. (5.16) based on the ML estimate $A_{m,i} = q_{m,i}/L_{m,i}$ will be called the *basic GLLR*, and the GLLR statistic in Eq. (5.20) based on $A_{m,i} = q_{0,0}/L_{m,i}$ will be called the *modified GLLR*. These three statistics are applied to the data in Figure 5.2 using the cost matrix in Table 5.1 which specifies a threshold for the GLRT hypothesis test. The particular threshold used corresponds to a situation in which the cost of an error is twice as high when the true underlying hypothesis is full-aperture scattering than for other types of errors. The decision rule is biased in this manner in order to guide attributions for clutter pixels towards the full-aperture decision to produce higher resolution images. Such a biasing is reasonable because clutter is composed of many irresolvable scatterers producing a noise-like response in azimuth and does not conform to the sub-aperture scattering model developed for a single scatterer.

Table 5.1. Cost matrix used to establish thresholds in the GLLR-based decision rules.

true anisotropy \ estimated hypothesis	full-ap.	half-ap.	quarter-ap.
full-ap.	0	2	2
half-ap.	1	0	1
quarter-ap.	1	1	0

The results of the three decision rules are displayed in Figure 5.3. The first row of results, i.e. figures (a) and (b), are for the test based on the reflectivity statistic. The second row of results, i.e. figures (c) and (d), are for the test based on the basic GLLR. The third row of results, i.e. figures (e) and (f), are for the test based on the modified GLLR. For each of these rows, the left-hand column of images (i.e. Figures 5.3(a), (c),

and (e)) displays the image of anisotropy attributions for each of the statistics. There are three levels of shading in this image. White denotes a full-aperture classification, light-gray denotes a half-aperture classification, and dark-gray denotes a quarter-aperture classification. As in Figure 5.2(b), the scattering center for each scatterer is marked with a geometric symbol, but now, the symbol also denotes the degree of anisotropy for the location. In particular, a circle is used to denote a full-aperture classification, a square is used to denote a half-aperture classification, and a triangle is used to denote a quarter-aperture classification. The right-hand column of images (i.e. Figures 5.3(b), (d), and (f)) displays images of the sub-aperture reflectivities from the selected sub-apertures.

As expected, use of the reflectivity statistic in Eq. (5.17) produces an excessive number of quarter-aperture attributions in Figure 5.3(a). With few exceptions, the entire image is classified as quarter-aperture anisotropy, including all the prominent scatterers, even though a cost structure favoring the full-aperture classification is used. The use of the basic GLLR performs significantly better than the reflectivity statistic. The basic GLLR correctly attributes 10 of the 11 scatterers, incorrectly attributing the weak scatterer labeled D in the upper-right portion of the image. The reflectivity statistic, in contrast, correctly attributes only the two quarter-aperture scatterers labeled C. Furthermore, the basic GLLR attributes considerably fewer pixels in clutter as anisotropic thereby improving the image resolution and decreasing the reflectivity variance for these pixels. Like the basic GLLR, the modified GLLR correctly attributes most of the scatterers in Figures 5.3(e). In particular, the only scatterer incorrectly attributed is again the weak scatterer labeled D in the upper-right portion of the image. This scatterer is misclassified by both the basic and modified GLLR's because of the interference from the strong quarter-aperture neighboring scatterer labeled C. Note that the other scatterer labeled D in the lower-right portion of the image is correctly attributed as half-aperture since this scatterer is isolated and thus unaffected by distant scatterers. The modified GLLR also further reduces the number of anisotropic attributions in clutter.

Looking at the images in the right-hand column of Figure 5.3, it is clear that imaging resolution has been lost by all the methods involved. Even the proposed GLLR statistics, which correctly attribute most of the scatterers, result in an excessively coarse resolution around *all* of the scatterers. The loss of resolution is due not to the attribution on the dominant scatterers themselves but the incorrect attribution of pixels in close proximity to the scatterers. In particular, close examination reveals that clutter pixels adjacent to a scatterer are always classified as anisotropic. The resolution cells for these anisotropic attributions are coarser due to the smaller sub-apertures used. In fact, the resolution is often coarse enough that the resolution cells expand to cover the interfering scatterer thus producing a strong reflectivity estimate and loss of imaging resolution.

The misclassification of these clutter pixels is due to interference from the neighboring scatterer which must be addressed if the anisotropy attribution is to be used in an
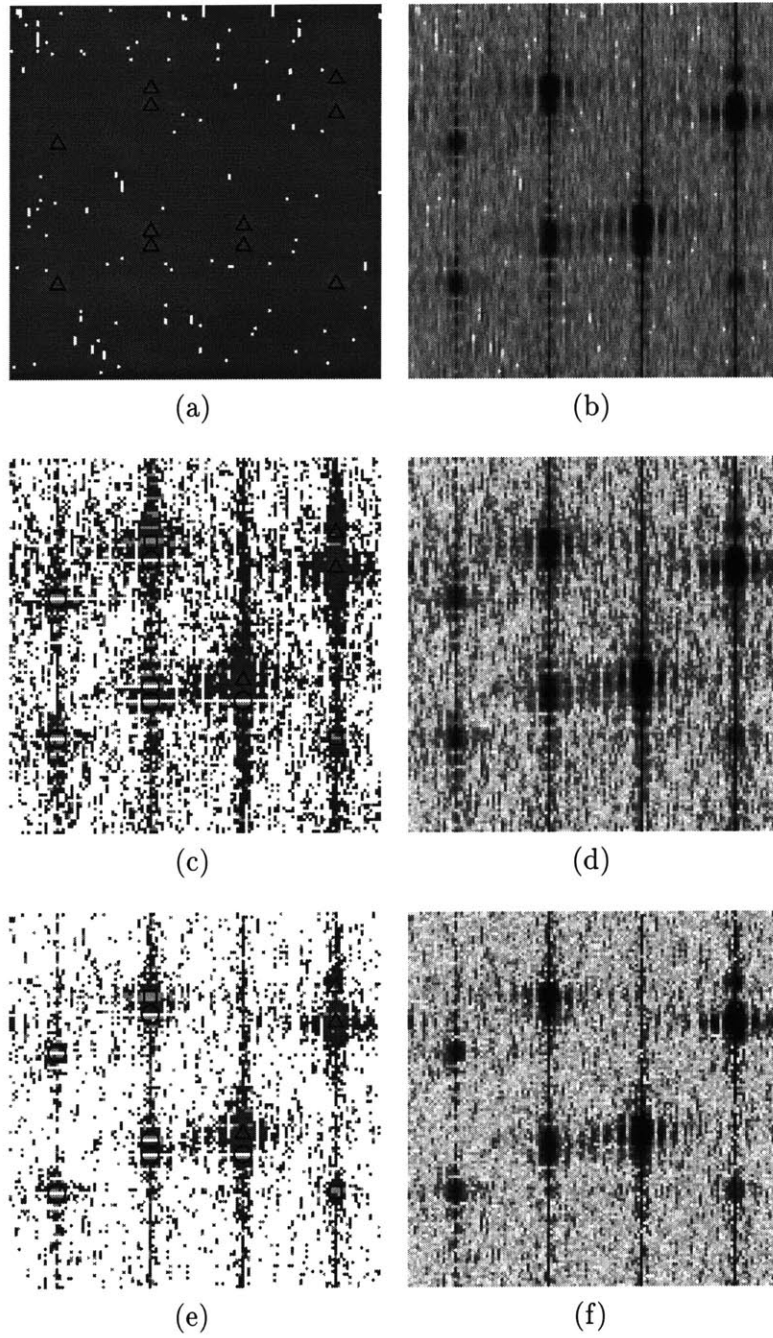
Figure 5.3. Results of using: (a,b) the reflectivity statistic; (c,d) the basic GLLR; and (e,f) the modified GLLR on the SAR data in Figure 5.2. Left column: images of degree of anisotropy. Right column: images of sub-aperture reflectivity.

image based algorithm where resolution is important. One approach to remedy these anisotropy misclassifications for *clutter pixels* is to require that a pixel's finest resolution reflectivity estimate exceed a particular threshold in order to be attributed. This condition can be interpreted as requiring that a pre-processing step, thresholding reflectivity in this case, detect pixels that contain a scatterer and thus need to be attributed with an anisotropy label. Figure 5.4 displays the same types of images as Figure 5.3 except that only pixels with a finest resolution reflectivity exceeding an SNR of 5dB are tested for anisotropy, while all other pixels are processed with the full-aperture. The vast majority of the clutter pixel misclassifications have been removed by this pre-processing step while still allowing the prominent scatterers to be attributed. The result is an image with much higher imaging resolution that still conveys anisotropy information from strong scatterers.

Unfortunately, the pre-screening step can do nothing to improved the attribution of prominent scatterers. In particular, the thresholding does nothing to correct the anisotropy attribution of the 20dB SNR scatterer labeled D in the upper-right of the images. This scatterer is misclassified due to the interference from the stronger 30dB SNR quarter-aperture scatterer. To improve the anisotropy attribution in the presence of neighboring scatterers, we need to explicitly account for the interference produced by the neighboring scatterers. The following sections in this chapter and Chapter 7 address the handling of this interference.

■ 5.2 Anisotropy Attribution – Multiple Scatterer Model

The results of the previous section demonstrate that although the use of the consistent (in terms of spatial coverage) reflectivity estimate $A_{m,i} = q_{0,0}/L_{m,i}$ in place of $A_{m,i} = q_{m,i}/L_{m,i}$ reduces the number of false anisotropy classifications in clutter, there are still a considerable number of misclassifications particularly surrounding a strong scatterer. These misclassifications lead to an overly coarse resolution in the image of sub-aperture reflectivities. Because the scatterers in the image have a high SNR, the attribution of many clutter pixels can be avoided with the use of a pre-screening threshold on the finest resolution reflectivity. However, such a pre-processing step might preclude the attribution of a weak scatterer and does not improve the classification of a strong scatterer whose observed response is distorted by interference from a neighboring scatterer. To address this problem, this section focuses on improving the attribution by generalizing the anisotropy observation model and hypothesis test account for the effects of interfering scatterers.

■ 5.2.1 Model and Test

To explicitly account for interfering scatterers, the sub-aperture scattering model in Eq. (5.5) must be generalized to account for the modulations produced by neighboring scatterers. To accomplish this task, we again start by assuming that the azimuthal response $a(\omega)$ is generated for a specified cross-range location y_0. The possibility of other

(a)   (b)

(c)   (d)

(e)   (f)
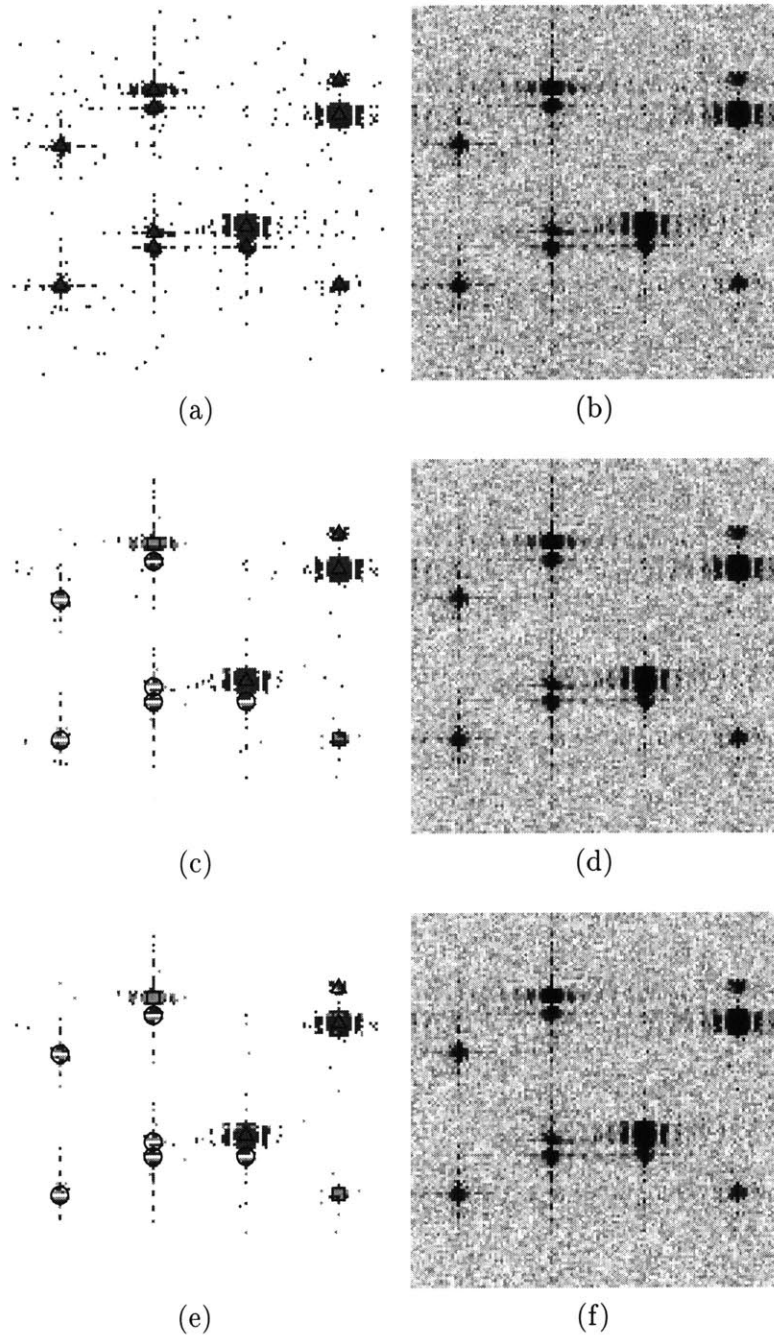
Figure 5.4. Results of using a pre-screening threshold with: (a,b) the reflectivity statistic; (c,d) the basic GLLR; and (e,f) the modified GLLR on the SAR data in Figure 5.2. Left column: images of degree of anisotropy. Right column: images of sub-aperture reflectivity.

scatterers are considered at discrete locations $y_0 + k\Delta_p$, where Δ_p specifies a cross-range sampling resolution that is taken to be less than the full-aperture image resolution Δ_r. Thus, the locations for neighboring scatterers are oversampled. Ideally, one would want to allow the neighboring scatterers to be represented at an arbitrary location instead of one of the pre-specified discrete locations. The former, however, significantly increases algorithmic complexity. The error incurred from modeling neighbors at $y_0 + k\Delta_p$ is expected to be small since the location error is less than one resolution cell (due to the oversampling).

To model the effect of a scatterer at location $y_0 + k\Delta_p$ on the measurements at location y_0, the azimuthal contribution from this neighboring scatterer is modulated to account for the shift in the image domain. The observed effect over the smallest sub-apertures is then given by

$$b^k(m_k, i_k)_j = \int_{S_{M,j}} e^{j2\pi \frac{k\Delta_p}{\Delta_r}\omega} 1_{H_{m_k, i_k}}(\omega) \ d\omega \tag{5.23}$$

where (m_k, i_k) is the sub-aperture anisotropy associated with the scatterer at $y_0 + k\Delta_p$. Incorporation of the neighboring scatterers is now modeled via superposition, i.e.

$$q_M = \sum_k A_k b^k(m_k, i_k) + \varepsilon \tag{5.24}$$

where A_k is the reflectivity associated with the scatterer at $y_0 + k\Delta_p$ and the noise model for ε is the same as in Eq. (5.6). In order to account for all scatterers that significantly affect the smallest sub-aperture measurements, the summation over k in Eq. (5.24) should be over a range large enough to include all scatterers contained in the largest resolution cell. Thus, the model should consider all $k \in \{-K, \ldots, K\}$ where K satisfies

$$K\Delta_p \geq \frac{\Delta_r}{2L_M} = \text{size of coarsest resolution cell}$$

$$K \geq \left\lceil \frac{\Delta_r}{2L_M \Delta_p} \right\rceil \tag{5.25}$$

and $k = 0$ corresponds to the focused resolution cell under investigation. Rewriting Eq. (5.24) as a matrix equation, it becomes

$$q_M = \left[b^{-K}(m_{-K}, i_{-K}) \ \mid \ \cdots \ \mid \ b^K(m_K, i_K) \right] \begin{bmatrix} A_{-K} \\ \vdots \\ A_K \end{bmatrix} + \varepsilon$$

$$= BA + \varepsilon \tag{5.26}$$

where B and A are appropriately defined.

Note that the hypotheses are now in a higher dimensional space. In particular, the hypothesis set is now $\mathcal{H}^{2K+1}$ since each location k has an associated scatterer. For the

initial work presented in this chapter, the hypotheses are restricted such that only the $k = 0$ (the focused resolution cell) hypothesis is allowed to vary, while the hypotheses for the other scatterers are fixed at the full-aperture hypothesis. Thus, the hypothesis space is still $\mathcal{H}$. This constraint on the anisotropy of neighbors is relaxed in Chapter 7 in order to more accurately account for the contributions from neighboring scatterers. The extension, however, requires addressing several complex issues and is thus left to another chapter.

The error $\boldsymbol{\varepsilon}$ in Eq. (5.26) can be used to compute the likelihood of each hypothesis. The resulting set of likelihoods for all the hypotheses can then be used in a decision rule. In particular, the error norm weighted by the inverse noise covariance, i.e. $\|\boldsymbol{\varepsilon}\|_{\Lambda^{-1}}$, is a sufficient statistic because the noise is additive and Gaussian. The value of the error $\boldsymbol{\varepsilon}$, however, is unknown and must be estimated. In the same spirit as the isolated scattering model, $\boldsymbol{\varepsilon}$ is estimated by first estimating $\boldsymbol{A}$ which is subsequently used to compute $\hat{\boldsymbol{\varepsilon}}$ from Eq. (5.26), i.e.

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{q}_M - B\hat{\boldsymbol{A}}$$

where $\hat{\boldsymbol{A}}$ is an estimate of $\boldsymbol{A}$ and $\hat{\boldsymbol{\varepsilon}}$ is the associated error estimate. One possibility for estimating $\boldsymbol{A}$ is ML estimation for the given hypothesis, i.e.

$$\hat{\boldsymbol{A}} = \arg\min_{\boldsymbol{A}} \left\{ \|\boldsymbol{\varepsilon}\|_{\Lambda^{-1}}^2 \right\} = (B^H \Lambda^{-1} B)^{-1} B^H \Lambda^{-1} \boldsymbol{q}_M \tag{5.27}$$

Taking $\boldsymbol{A}$ as the ML estimate, the weighted norm of the estimated error $\hat{\boldsymbol{\varepsilon}}$ becomes

$$\begin{aligned}
\|\hat{\boldsymbol{\varepsilon}}\|_{\Lambda^{-1}}^2 &= \|\boldsymbol{q}_M - B\hat{\boldsymbol{A}}\|_{\Lambda^{-1}}^2 \\
&= \frac{1}{2\sigma^2} \boldsymbol{q}_M^H \left[\Lambda^{-1} - \Lambda^{-1} B (B^H \Lambda^{-1} B)^{-1} B^H \Lambda^{-1} \right] \boldsymbol{q}_M
\end{aligned} \tag{5.28}$$

which simplifies to the basic GLLR in Eq. (5.16) for the isolated scatterer model in the case of limiting the model order to $K = 0$.

The model in Eq. (5.24) has too many parameters for the test given by the minimization of Eq. (5.28) to work well. In particular, if one chooses K sufficiently large to account for all the scatterers in the coarsest resolution cell, then the model order is greater than the number of sub-aperture measurements in $\boldsymbol{q}_M$, and thus, the estimated error $\hat{\boldsymbol{\varepsilon}}$ can be made zero for all hypotheses. Consequently, we regularize the optimization problem. In order for the estimated error $\hat{\boldsymbol{\varepsilon}}$ to be made small under the incorrect model, many of the values in $\boldsymbol{A}$ generally have to be made unreasonably large which results in an unrealistic set of reflectivities. Thus, a 2-norm regularization penalty is used in the estimation of $\boldsymbol{A}$. In particular, instead of minimizing the weighted squared error to estimate $\boldsymbol{A}$, $\boldsymbol{A}$ is estimated as

$$\begin{aligned}
\hat{\boldsymbol{A}} &= \arg\min_{\boldsymbol{A}} \left\{ \|\boldsymbol{\varepsilon}\|_{\Lambda^{-1}}^2 + \gamma \boldsymbol{A}^H R \boldsymbol{A} \right\} \\
&= (B^H \Lambda^{-1} B + \gamma R)^{-1} B^H \Lambda^{-1} \boldsymbol{q}_M \\
&= P \boldsymbol{q}_M
\end{aligned} \tag{5.29}$$

where P is defined accordingly, γ is a regularization parameter, and R is the regularization matrix that penalizes the energy in all the A_k other than $k = 0$, i.e.

$$R = I - e_K e_K^H = \text{diag}(1, \ldots, 1, 0, 1, \ldots, 1).$$

This regularization can be interpreted as a zero-mean Gaussian prior on the neighboring scatterer reflectivities. Use of the regularized value of A produces the following value for the weighted error norm

$$\begin{aligned}
\|\hat{\boldsymbol{\varepsilon}}\|_{\Lambda^{-1}}^2 &= \|\boldsymbol{q}_M - B\hat{A}\|_{\Lambda^{-1}}^2 \\
&= \frac{1}{2\sigma^2} \boldsymbol{q}_M^H \left[\Lambda^{-1} - 2\Lambda^{-1}BP + P^H B^H \Lambda^{-1} BP\right] \boldsymbol{q}_M
\end{aligned} \tag{5.30}$$

which can be used as the statistic in the anisotropy hypothesis test. For simplicity, the statistic in Eq. (5.30) will be called the *multiple scatterer model statistic* (or *MSM statistic*). Note that the weighting matrix in brackets of Eq. (5.30) is data independent. Thus, the weighting matrix can be pre-computed offline for each hypothesis thereby allowing for a computationally simple likelihood that corresponds to a norm on $\boldsymbol{q}_M$.

■ 5.2.2 Experimental Results

To demonstrate the performance of the anisotropy test based on the MSM statistic presented in this section, a GLRT based on this statistic is applied to the SAR data in Figure 5.2. The number of neighboring scatterers considered is set by $K = 6$ and their spacing is set by the image oversampling rate of $\Delta_r/\Delta_p = 1.25$. The value of the regularization parameter on neighboring reflectivities is set by $\gamma = 0.5$. All other parameters are set to the same values as in Section 5.1.7.

The results of the new test are displayed in Figures 5.5 and 5.6. Figure 5.5(a) shows the degree of anisotropy attributed, and Figure 5.5(b) shows the resulting sub-aperture reflectivity estimate. Comparing Figure 5.3 and Figure 5.5, neither of which use the pre-screening amplitude threshold, it is clear that the test using the MSM statistic better suppresses the number of false anisotropy classifications in clutter than the tests based on the isolated scatterer model. Furthermore, the test using the MSM statistic still correctly attributes all of the scattering centers correctly except for the weak scatterer labeled D in the upper-right portion of the image which is still attributed as quarter-aperture anisotropy. The reason this extended model does not correctly attribute this scatterer is the neighboring scatterers are modeled as isotropic with a reflectivity that is regularized towards zero. Consequently, the model does not fully account for the interference from strong neighboring scatterers, and the residual interference is still sufficient to cause a misclassification.

Results using the 5dB SNR pre-screening threshold with the new test are shown in Figure 5.6. As expected, the threshold further reduces the number of anisotropy classifications in clutter from that in Figure 5.4 for the modified GLLR test with a pre-screener and Figure 5.5 for the MSM statistic test without a pre-screener. Unfortunately, in all cases, there are still false anisotropy classifications for clutter pixels in

the vicinity of an actual scatterer resulting in a coarser image resolution. This coarse resolution is quite evident around the isotropic point scatterers whose energy is visible over several pixels in the sub-aperture imagery. The negative effects of such coarsening are also evident in Figure 5.5 where we see that without the use of a pre-screening threshold, scatterers in close proximity often appear as a single scatterer.

(a)   (b)
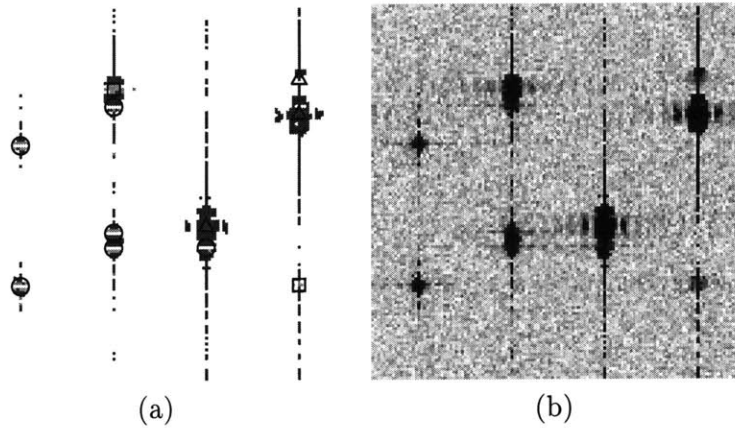
Figure 5.5. Results of using the anisotropy test based on the MSM statistic applied to the SAR data in Figure 5.2. Images of (a) degree of anisotropy and (b) sub-aperture reflectivity.

(a)   (b)
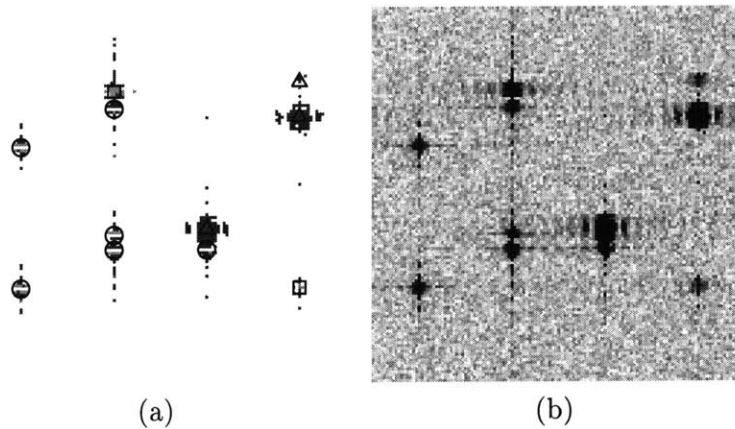
Figure 5.6. Results of using the anisotropy test based on the MSM statistic with a pre-screening threshold applied to the SAR data in Figure 5.2. Images of (a) degree of anisotropy and (b) sub-aperture reflectivity.

■ 5.3 Decision-Directed Telescopic Testing

The test based on the MSM statistic is, to a fair degree, able to suppress anisotropic classifications in clutter. Pixels in the immediate vicinity of a scatterer, however, are still

frequently classified as anisotropic leading to a coarsening of the image resolution around these scatterers, even when the pre-screening threshold is applied. To help address these misclassifications in clutter, this section presents a telescopic testing procedure for determining anisotropy attribution. The procedure tests anisotropy hypotheses in a decision-directed fashion making use of the knowledge that imaging resolution coarsens as sub-aperture size is decreased to produce anisotropy classifications that are robust to the presence of interfering scatterers. The improved anisotropy attribution thus produces sub-aperture reflectivity images that preserve high imaging resolution while also providing anisotropy information and improved sub-aperture reflectivity estimates.

■ 5.3.1 Telescopic Algorithm

The sub-aperture pyramid which is used to form sub-aperture measurements and define anisotropy hypotheses is convenient not only for revealing anisotropy information, but also for providing an efficient and robust means for performing the hypothesis tests. All of the tests presented in Sections 5.1 and 5.2 can be adapted so that instead of exhaustively computing and comparing the GLLR's for each of the sub-aperture hypotheses, the GLLR's need to be evaluated and compared for only a small subset of the candidate hypotheses. In particular, due to the nested structure of the sub-apertures composing a pyramid, given by condition (D) in Section 3.1, the tests can be performed in a telescopic fashion traversing down a branch in the pyramid of sub-apertures. This processing is depicted in the example in Figure 5.7 where the amount of shading conveys the value of the GLLR statistic with lighter shades indicating higher GLLR's. On the pyramid, one expects the likelihoods to increase as the pyramid is traversed downward and the hypothesized sub-apertures approach the correct sub-aperture, i.e. as the sub-aperture scattering model better fits the actual response. Continuing down the pyramid past the correct sub-aperture, one expects the likelihoods to then decrease as the hypothesized sub-apertures start to deviate from the correct sub-aperture, i.e. as the sub-aperture scattering model becomes excessively anisotropic. This intuition motivates performing the hypothesis test in the following manner:

Step 1: Start with the full-aperture at scale $m = 0$ and denote it as H_{0,i_0^*}.

Step 2: Consider those hypotheses at scale $m+1$ for which $S_{m+1,i_{m+1}} \subset S_{m,i_m^*}$, i.e. children of the current sub-aperture. Find the one which has the highest GLLR statistic and denote it as H_{m+1,i_{m+1}^*}.

Step 3: If the parent is sufficiently more likely (i.e. $\gamma_{m,i_m^*} > \gamma_{m+1,i_{m+1}^*} + T$, where T is the GLLR threshold computed from the cost structure), then stop and return H_{m,i_m^*} as the estimated hypothesis.

Step 4: If $m + 1 = M$, we are at the bottom of the tree, so stop and return H_{m+1,i_{m+1}^*} as the estimated hypothesis. Otherwise, increment m and goto Step 2.

This procedure provides an efficient approximation to selecting a hypothesis without having to evaluate all the sub-aperture statistics. In addition to reducing computational
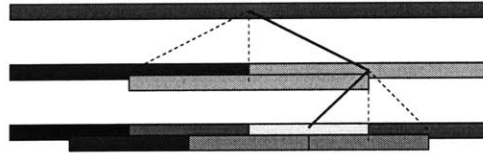
Figure 5.7. Illustration of how the anisotropy testing can be done in a decision-directed fashion by starting with the full-aperture, and at each scale, inspecting only the children of the most likely sub-aperture. The testing terminates as soon as the parent is sufficiently more likely than its children or the bottom of the pyramid is reached. Lighter shadings of the sub-apertures indicate higher likelihoods. Dashed lines denote which hypotheses are tested, and solid lines denote the branch traversed.

burden, performing the hypothesis test in a telescopic fashion also enforces a degree of consistency in the sub-apertures tested thereby increasing the robustness of the test. In particular, in order for a particular sub-aperture ultimately to be chosen, its ancestor hypotheses must also have been chosen in the test at the previous scales. While this behavior is reasonable for anisotropic scatterers, this pattern is not necessarily followed when there is interference from neighboring scatterers. In particular, the effect of a neighboring scatterer is to corrupt the statistics for smaller sub-aperture hypotheses since they have larger resolution cells. The statistics for larger sub-apertures, however, are unaffected since these sub-apertures have smaller resolutions cells. Thus, the decision-directed testing process frequently stops prior to reaching a scale where the interfering scatterer enters the coarser resolution cells and corrupts the anisotropy attribution.

■ 5.3.2 Results

The performance of the telescopic testing procedure is demonstrated using the modified GLLR and the MSM statistic. The SAR data and test parameters used here are the same as those used in Sections 5.1.7 and 5.2.2. The results of using the telescopic testing procedure with the modified GLLR are displayed in Figures 5.8(a) and (b) for the case of no pre-screening threshold and Figures 5.9(a) and (b) for the case of a pre-screening threshold of 5dB SNR. Similarly, the results of using the telescopic testing procedure with the MSM statistic are displayed in Figures 5.8(c) and (d) for the case of no pre-screening threshold and Figures 5.9(c) and (d) for the case of a pre-screening threshold of 5dB SNR.

As motivated, the use of the telescopic testing procedure has significantly reduced the number of anisotropy classifications around strong scatterers and correspondingly improved the resolution in these areas. In fact, with the telescopic testing method, the modified GLLR and MSM statistic tests now produce sub-aperture reflectivity images with sufficiently high resolution that individual scatterers can be identified. However, the scatterer in the upper-right of the image is still misclassified. As mentioned in the previous section, this misclassification is due to the corruption of the likelihood statistic by the interference from a strong neighboring scatterer. The purpose of the telescopic

(a)   (b)

(c)   (d)
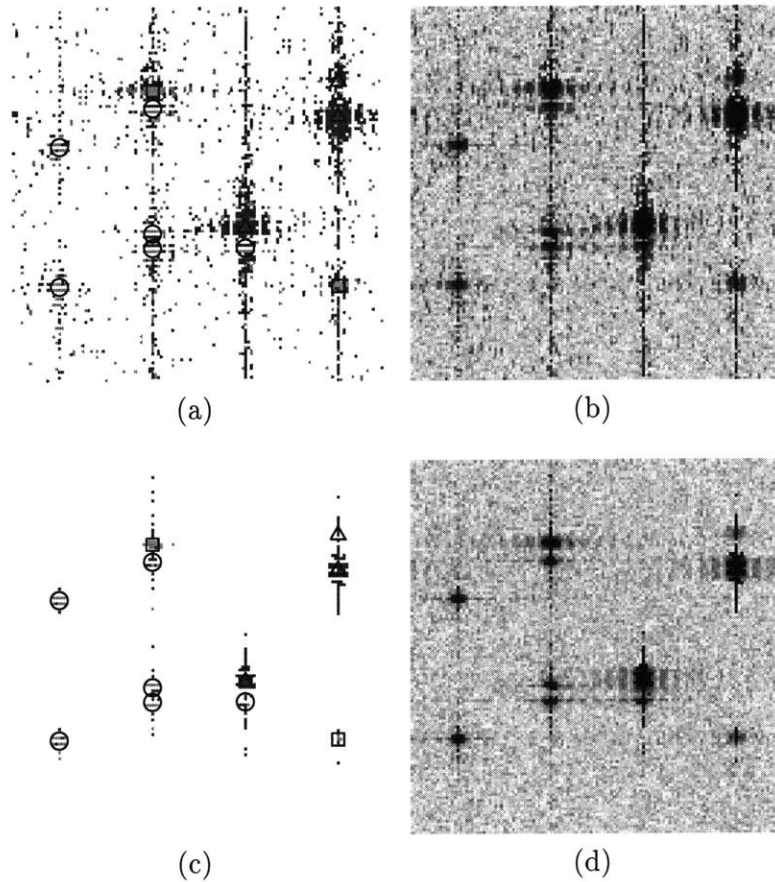
Figure 5.8. Results of using the telescopic testing procedure (a-b) with the modified GLLR and (c-d) with the MSM statistic. Images of (a,c) degree of anisotropy and (b,d) sub-aperture reflectivity.

testing procedure is to reduce anisotropy classifications in clutter where interference causes inconsistent patterns in the likelihoods on the pyramid as the resolution cell size varies with scale. The processing is not specifically designed to improve the classification on strong scatterers. Consequently, it is not surprising that the attribution of the misclassified scatterer is not corrected by the telescopic testing procedure. The accounting for neighboring scatterers must be improved in order to correct this problem. This task is left to Chapter 7.

■ 5.4 Results on MSTAR Data

This section presents results using the MSTAR public release data set. These images have a resolution of one foot in both range and cross-range that is achieved using a transmitted signal with a 590MHz bandwidth, 9.6GHz center frequency, and 2.8° azimuth. This data set is described in detail in Appendix B.

(a)   (b)

(c)   (d)
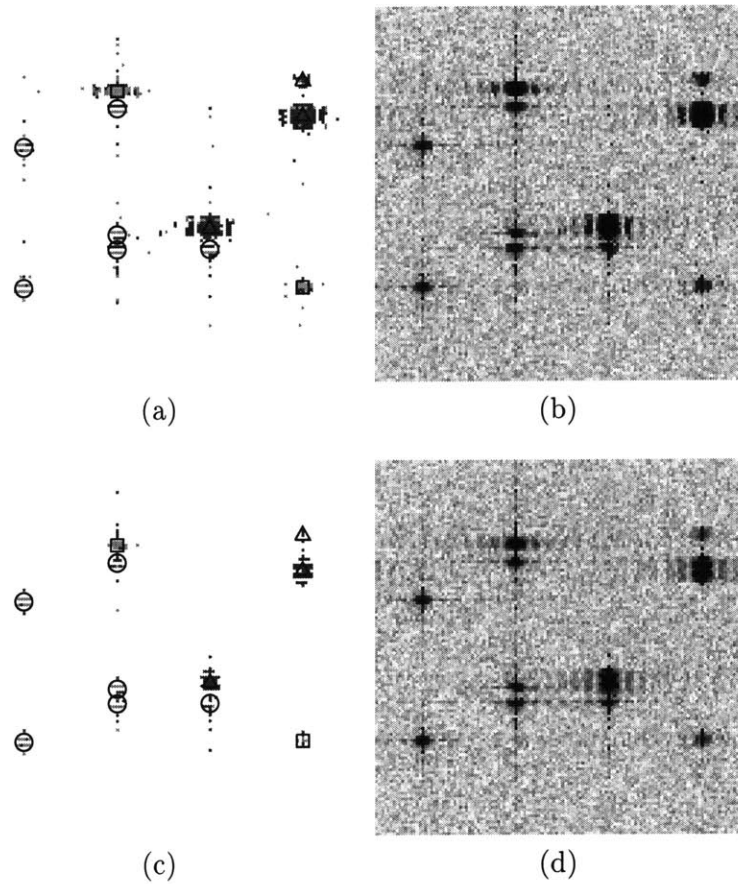
Figure 5.9. Results of using the telescopic testing procedure and a pre-screening threshold (a-b) with the modified GLLR and (c-d) with the MSM statistic. Images of (a,c) degree of anisotropy and (b,d) sub-aperture reflectivity.

All of the results in this section are based on the three-level, half-overlapping, half-aperture pyramid. The GLRT based on the MSM statistic using the telescopic testing procedure is used to characterize anisotropy. The parameters for the test are the same as those used in the previous sections. In particular, the number of neighboring scatterers is $K = 6$, the regularization parameter on neighboring reflectivities is $\gamma = 0.5$, and the cost matrix is the one given in Table 5.1. No pre-screening threshold is used for the MSTAR data as a significant threshold might preclude the attribution of real scatterers.

■ 5.4.1 Images of Anisotropy

To illustrate the anisotropy classifications produced for measured data, Figure 5.10 displays the results for a BMP-2 (c21) at $0°, 30°, 60°$, and $90°$ azimuths with a $17°$ depression. Each row of the figure displays three images for a given azimuthal orientation. The first image in each row is the standard full-aperture log-magnitude reflectivity im-

age. The second image is a display of the anisotropy attribution at each the pixel. The third image is the log-magnitude of the sub-aperture reflectivity image. Even though the aperture associated with this data set is relatively small at 2.8°, anisotropy is still detected. In particular, it usually appears to be associated with a part of the tank. Particular note is made of the exceptions at the 0° azimuth. Here, there are many clutter pixels classified as anisotropic. The reason that these pixels are declared as anisotropic is that the interference from neighboring large scatterer is not well modeled. Recall that the current formulation only considers the possibility of neighboring scatterers which are isotropic. However, the front edge of the tank generates a very strong anisotropic response which is not accurately captured in our current model. Extending the model to account for such scattering (as done in Chapter 7) should alleviate problems such as this.

■ 5.4.2 Confusion Matrices

The images in Figure 5.10 show that scatterers are being classified as anisotropic, however these images do not convey how useful that information is in characterizing targets. To address this issue, empirical confusion matrices for anisotropy classification are generated using the MSTAR data set. In particular, the confusion probabilities are estimated from a data set composed of the following vehicles: 2S1 (b01), BMP-2 (c21), BRDM-2 (E-71), D-7 (92v13015), T-72 (132), ZIL-131 (E12), and ZSU-23/4 (d08) where truth is taken as the empirical results from the 17° depression data and test data is taken from the 15° depression data. For each pair of training and testing images at the same azimuth, the 20 brightest peaks[6] are extracted from each image and a correspondence match, based on relative location, is performed and taken as truth. The empirical confusion matrices are then computed from the anisotropy attributions of these peaks. For the remainder of this section, the use the term "confusion matrix" refers to one computed in this fashion.

The confusion matrix for this data set is given in Table 5.2. Each row of this matrix represents the estimated probability distribution of the anisotropy attribution for a scatterer conditioned on the "truth" attribution of the sister vehicle. The most prominent aspect of this confusion matrix is the apparent independence of testing anisotropy and training anisotropy illustrated by the similarly of each row of the matrix, i.e. the near independence on the conditioning of the training hypothesis. This result is counter-intuitive. In particular, canonical scatterer models show a strong functional dependence of anisotropy on scatterer geometry. Consequently, we should expect the anisotropy attribution for these scatterers to be stable features. However, canonical scattering is not the only source of anisotropy. Anisotropy may also arise from the interference of multiple irresolvable scatterers. The conglomerate response from such a set of scatterers is a superposition of complex exponentials in azimuth. Thus, over a narrow aperture, such as the case in MSTAR data, the composite response may appear to be anisotropic.

---

[6]*Peaks* are taken as local maxima in the image with a minimum distance separating them.

The instability of such an instance of anisotropy is not surprising especially considering that the depression angle has changed.

To further evaluate the source and stability of anisotropic phenomena, the data set is partitioned into those images at a near-cardinal azimuth (within 2.5° of a cardinal angle) and those at off-cardinal azimuths. The motivation for this partitioning is that the influence of canonical scatterers is expected to be most pronounced at near-cardinal azimuths because of the natural rectangular shape of vehicles. Thus, near-cardinal azimuths should exhibit anisotropic scattering associated with flat plates and other large canonical scatterers oriented orthogonal to the impinging radar signal.

The confusion matrices for the near-cardinal and off-cardinal azimuths are given in Table 5.3 where there are significant differences between the two cases. In particular, there is now a noticeable presence along the diagonal of the confusion matrix based on the near-cardinal orientations, signaling that anisotropy is more stable at these angles as expected for canonical scatterers. That the diagonal entries are not larger is not surprising given that there are very few large canonical scatterers at these orientations, and several of the extracted peaks probably correspond to collections of irresolvable scatterers. The off-cardinal confusion matrix, however, shows that training anisotropy is independent of testing anisotropy. This result leads us to believe that there are at least two fundamental sources of anisotropy. One source is canonical scattering which is prominent at cardinal azimuths, but much less present at other azimuths in narrow-aperture data. The other is an unstable source of anisotropy which is commonly observed at off-cardinal azimuths. A likely candidate for this unstable anisotropy is the scintillation produced by the irresolvable interfering scatterers discussed earlier. Anisotropy arising from such interference is highly variable and changes unpredictably, which would account for the lack of correlation in the anisotropy classifications between the testing and training data in the off-cardinal confusion matrix.

Table 5.2. Anisotropy confusion matrix for 2S1, BMP-2, BRDM-2, D7, T-72, ZIL-131, and ZSU-23/4.

Training \ Testing	full-aperture	half-aperture	quarter-aperture
full-aperture	0.88	0.09	0.03
half-aperture	0.83	0.13	0.15
quarter-aperture	0.81	0.12	0.07

■ 5.4.3 Summary of MSTAR Results

The results on synthetic data in this chapter have shown that anisotropy information can be reliably extracted from image data. Care must be taken, however, to address the effects of interfering scatterers which corrupt the anisotropy attribution and lead to a coarse imaging resolution. Several anisotropy tests were presented in this chapter to address the effects of such interference. The GLRT based on the MSM statistic with the telescopic testing modification proved to be the most effective but this statistic does

Table 5.3. Anisotropy confusion matrix for vehicles at near-cardinal and off-cardinal angles.

Training \ Testing	Near-cardinal			Off-cardinal		
	full-ap.	half-ap.	qtr.-ap.	full-ap.	half-ap.	qtr.-ap.
full-ap.	0.82	0.11	0.07	0.89	0.08	0.03
half-ap.	0.72	0.26	0.02	0.83	0.12	0.05
qtr.-ap.	0.61	0.09	0.30	0.83	0.12	0.05

not adequately compensate for strong interfering scatterers. The issue of how to better deal with these neighboring scatterers is addressed in Chapter 7.

The results on the MSTAR data are not what one would expect given the results on the synthetic data. In particular, the anisotropy attribution of peaks in the MSTAR data set is not stable. The stability of anisotropy attributions are separately examined over near-cardinal and off-cardinal angles. This partitioning reveals distinct behavior over the two sets. In particular, the anisotropy attribution is very unstable at off-cardinal angles, but the attribution is relatively stable at near-cardinal angles. This behavior leads us to conjecture that there are two sources of anisotropy which are examined in detail in Chapter 6.
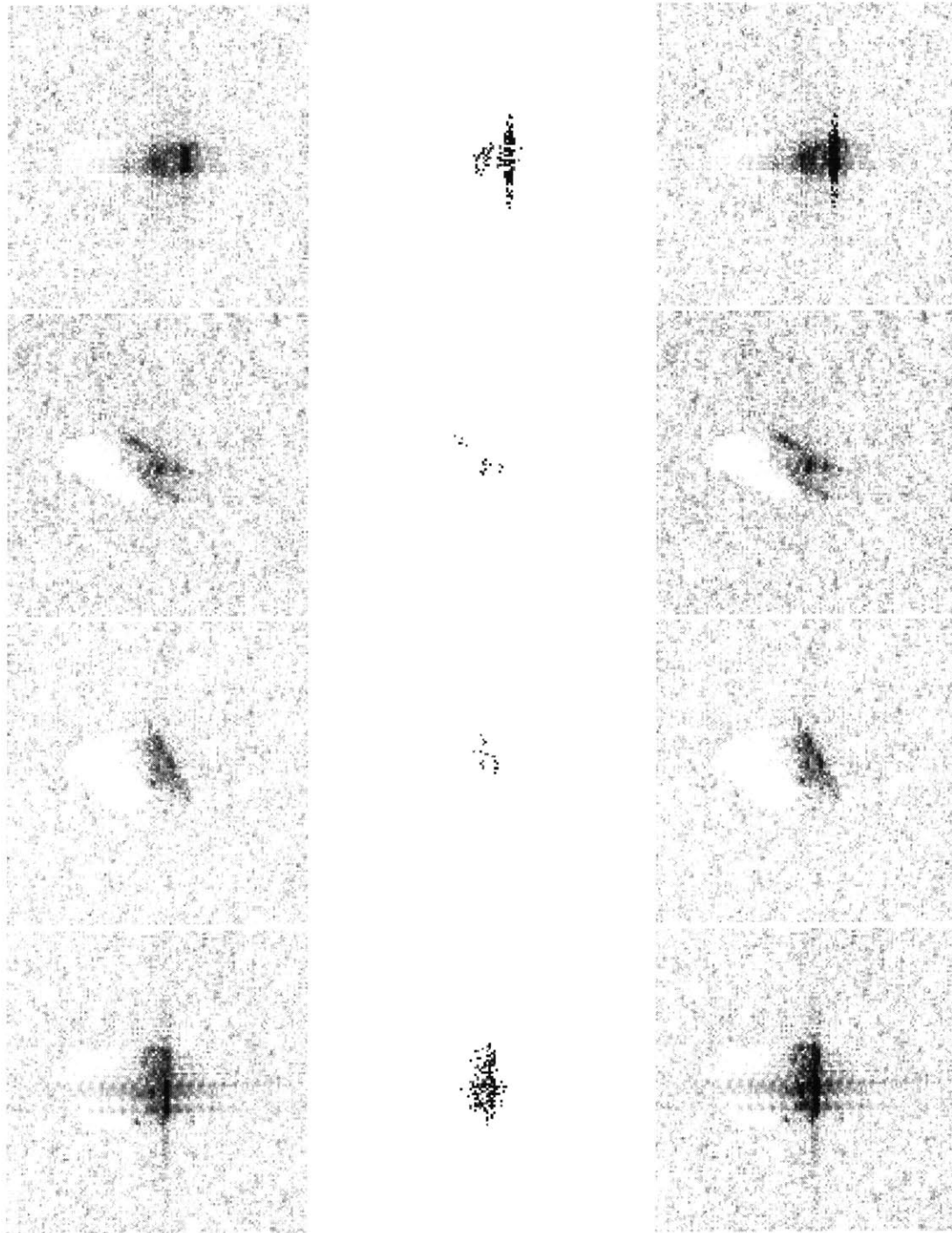
Figure 5.10. Anisotropy characterization of a BMP-2 at (first row) 0° azimuth, (second row) 30° azimuth, (third row) 60° azimuth, and (fourth row) 90° azimuth. First column: Images of log-magnitude reflectivities. Middle column: Image of anisotropy classifications. Last column: Image of sub-aperture reflectivities.

Chapter 6

Relation between Phenomenology and Anisotropy Attribution

In the previous chapter, we observed that anisotropy attribution is considerably less stable at off-cardinal angles than at near-cardinal angles in the MSTAR data set. This behavior supports the argument for separate sources of anisotropy. This chapter explores two such sources that are consistent with the observations in the MSTAR data. The two particular sources considered are geometric and volumetric scatterers. Scattering models, such as the PO and GTD models, give functional forms for radar scattering from canonical scatterers which clearly convey an azimuthal dependence on the shape, size, and orientation of the scatterer. These models, however, are based on very specific assumptions about the configuration of scatterer. For example, the model for a dihedral assumes two equal-sized rectangular plates conjoined to form a 90° angle. Our analysis of anisotropy, however, produces a more general characterization of anisotropy. This characterization allows for the consideration of a broader class of scatterers that we call *geometric scatterers*. This class includes not only canonical scatterers but also those scatterers which deviate from canonicity. The other source of anisotropy considered is what we call *volumetric scatterers*. These scatterers are actually a collection of multiple scatterers which are too closely spaced to resolve as a single scatterer and thus appear as a single scatterer with an erratic azimuthal response due to the chaotic interference among the constituent scatterers.

To demonstrate the scattering produced by these sources of anisotropy, Figure 6.1 shows the azimuthal response from a canonical flat plate measuring 1m in cross-range and a volumetric scatterer composed of 10 randomly located scatterers distributed over 1m in cross-range. The flat plate response is clearly anisotropic as the azimuthal energy is highly concentrated in the middle section of the aperture. The anisotropy characterization of the volumetric scatterer, however, is not as clear because the azimuthal response is less concentrated in azimuth, and the energy is erratically distributed across the aperture. Figure 6.1 illustrates an important point about the "stability" of anisotropy. In particular, over a relatively small aperture, such as the 2.8° aperture used in the MSTAR data, it is clear how the volumetric scatterer's response might be classified as anisotropic as only one of the localized fluctuations is observed. If
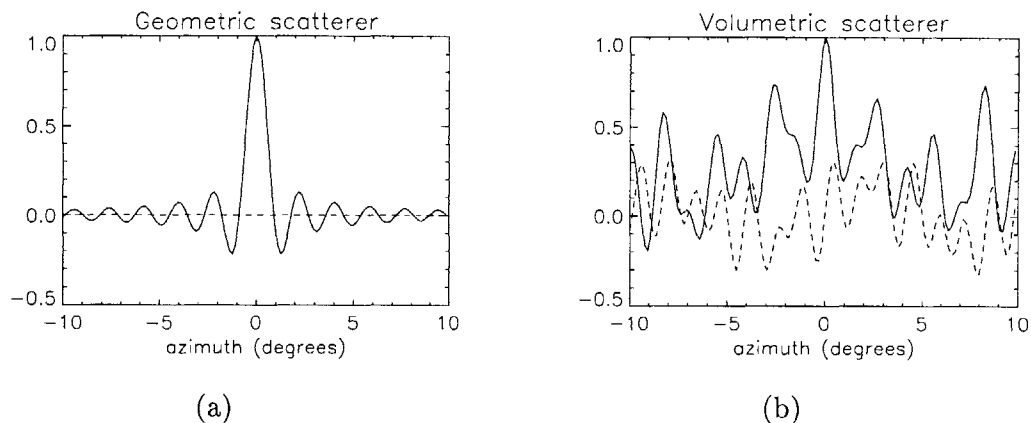
Figure 6.1. Instances of azimuthal scattering by a (a) geometric scatterer and (b) volumetric scatterer with the real and imaginary components of the response shown in the solid and dashed lines respectively.

the aperture is shifted or enlarged, however, the erratic distribution of energy over the aperture may lead to a different attribution for this scatterer because the sub-aperture scattering model poorly fits this azimuthal response. Thus, the anisotropy attribution for these volumetric scatterers is not expected to be stable. Geometric scatterers, in contrast, should produce a stable and informative anisotropy characterization as the unimodal azimuthal responses for these scatterers are relatively well modeled by the sub-aperture scattering model. Thus, the attribution for geometric scatterers should be robust to variations in the size and location of the aperture.

The remainder of this chapter elaborates on the ideas presented above. Section 6.1 discusses the behavior of geometric anisotropy[1] and the information that anisotropy attribution conveys for geometric scatterers. Section 6.2 discusses the behavior of volumetric anisotropy and how this behavior varies with the properties of the collection of scatterers, such as spatial density and coverage. Section 6.3 uses the results of the previous two sections to explain how the use of wide-aperture data results in a lower incidence of anisotropic attributions for volumetric scatterers while preserving the attributions for the more meaningful geometric scatterers. Section 6.4 describes an additional benefit of wide-aperture data which is the posing of additional, less anisotropic, sub-aperture hypotheses that allow for a more refined characterization of small scatterers. Section 6.5 concludes the chapter by demonstrating the argued benefits of a wide-aperture on the Grayling Ultra-wideband SAR[2] (GUS) data set.

---

[1]We will refer to anisotropy due to geometric scatterers as *geometric anisotropy* and anisotropy due to volumetric scatterers as *volumetric anisotropy*.

[2]The GUS data set is composed of extremely high resolution imagery achieved using large bandwidth of 4GHz and an aperture of 26°.

■ 6.1 Geometric Anisotropy

The behavior of geometric and volumetric anisotropy is studied in this chapter using the telescopic anisotropy test based on the multiple scatterer model developed in Chapter 5. The analyses presented in this and the following sections are based on narrow-aperture data because of the prevalence of such data in the SAR community. The data used in these sections are synthesized according to MSTAR operational parameters, i.e. one foot resolution in range and cross-range achieved with a 2.8° aperture, 9.6GHz center frequency, and 590MHz bandwidth. It should be emphasized that the use of MSTAR parameters is merely to aid intuitive understanding as the results can always be scaled with respect to aperture size. The three-level, half-overlapping, half-aperture pyramid depicted in Figure 3.3(b) is used for this narrow-aperture data. The number of neighboring scatterers considered is set by $K = 6$, and their spacing is set by the oversampling rate of $\Delta_r/\Delta_p = 1.25$ used in the MSTAR data set. The value of the regularization parameter on neighboring reflectivities is set by $\gamma = 0.5$. Deviations from the boxcar model are accounted for by setting $\rho = 0.1$.

To study the dependence of anisotropy classification on scatterer size, this thesis uses what will be called an *anisotropy plot*. This plot displays several curves, one for each degree of anisotropy, where each curve plots the probability of anisotropy classification conditioned on the dependent variable scatterer size. Thus, the plotted values for a fixed scatterer size, i.e. for any value of the independent coordinate (scatterer size), the values of the three curves correspond to the probability of choosing each of the three hypotheses. To generate the anisotropy plots, Monte Carlo probability estimates are used. In particular, for each scatterer size, $2^{13} = 8192$ Monte Carlo trials are run, and the probability of attribution for each degree of anisotropy is taken as the corresponding relative frequency of occurrence.

The flat rectangular plate scatterer is studied in this section to demonstrate the behavior of geometric anisotropy. The flat plate is only one instance of a geometric scatterer, but its unimodal azimuthal response is characteristic of the class of geometric scatterers considered. Note that it is the objective of this section to understand the stability of the anisotropy classification of these scatterers and not to specifically characterize parameters of the scatterers. Thus, the use of the flat plate alone is sufficient for our goals.

Anisotropy plots for the flat plate are shown in Figure 6.2 with each plot accounting for various uncertainties. Figure 6.2(a) is the simplest case where the scatterer is taken at broadside and the exact scattering center location is known. The synthesized azimuthal response is generated according to the PO model for the rectangular plate in Eq. (2.16). Measurement noise is incorporated in this data through the addition of white Gaussian noise across the aperture. The PSNR of the signal is kept constant at PSNR=20dB, thus noise variance scales with the maximum peak reflectivity in the image domain.

Contained in Figure 6.2(a) are three curves each of which represents the probability of a particular anisotropy attribution (full, half, or quarter-aperture) as a function
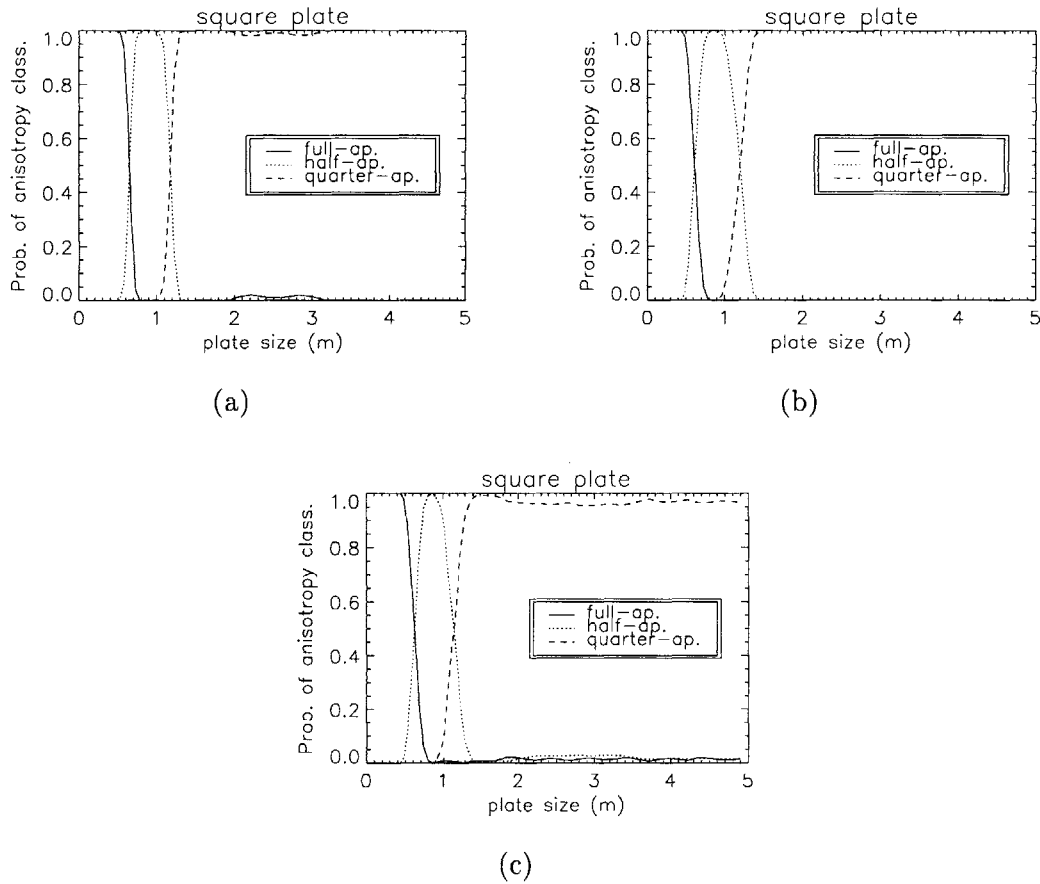
Figure 6.2. Anisotropy plots for a square plate in Gaussian noise at 20dB PSNR and (a) no azimuthal and location uncertainty; (b) maximal azimuthal uncertainty and no location uncertainty; (c) maximal azimuthal uncertainty and location uncertainty given by peak extraction.

of plate size. As expected, the full-aperture classification dominates for the smaller plate sizes which generate broader responses in azimuth. As plate size is increased, the half-aperture classification starts to dominate. Between 0.8m and 1m, the half-aperture classification probability is approximately one. The noise-free azimuthal response for a plate size in the middle of this region (0.9m) is shown in Figure 6.3(a) which is quite reasonably classified as half-aperture. The solid curve in Figure 6.3(a) represents the real component of the azimuthal response, and the dashed curve represents the imaginary component. The vertical lines are a visual aid dividing the aperture into eighths. As plate size is increased beyond 1m, the dominant classification in Figure 6.2(a) transitions from half-aperture to quarter-aperture. Between 1.3m and 1.9m, the quarter-aperture classification probability is approximately one. The azimuthal response for a

plate size in the middle of this region (1.6m) is shown in Figure 6.3(b) and is reasonably classified as quarter-aperture.
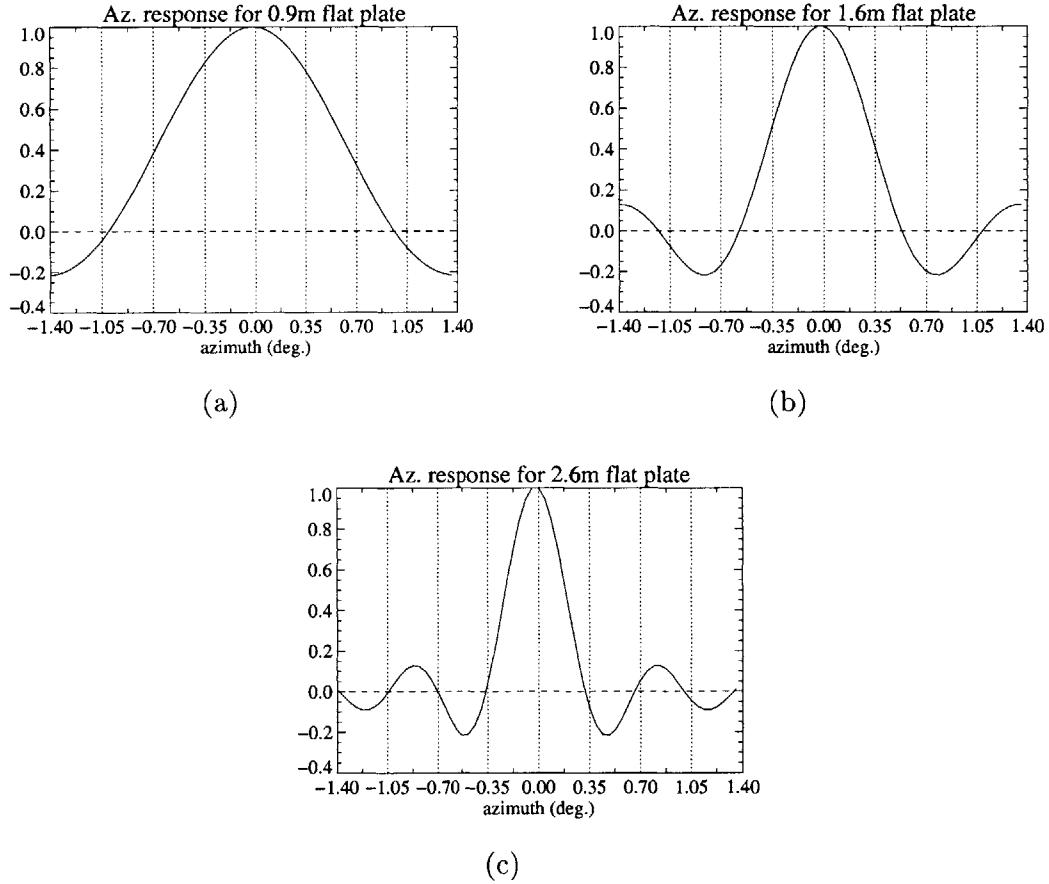
Figure 6.3. Azimuthal response for different plate sizes: (a) 0.9m, (b) 1.6m, and (c) 2.6m.

Intuition would suggest that the curves in the anisotropy plot would be unimodal and, in particular, that the probability of the quarter-aperture classification would monotonically increase with plate size since quarter-aperture anisotropy is the most anisotropic classification available. However, this behavior is not observed. In particular, starting at about 1.9m, the quarter-aperture classification probability begins to decrease slightly. The middle of the range over which the probability drops is about at 2.6m. The azimuthal response for this plate size is shown in Figure 6.3(c). One would expect this response to be classified as quarter-aperture anisotropy. In fact, the null-to-null bandwidth for this plate size is a quarter of an aperture. So, the question is why is this scatterer occasionally classified as full-aperture. This misclassification is an artifact of the telescopic hypothesis testing procedure. To see this fact, observe the GLLR's for the noise-free response given in the first column of Table 6.1. From

these statistics, one can see that the quarter-aperture hypothesis does indeed produce the highest GLLR, but because the full-aperture hypothesis is nearly as likely as the half-aperture hypothesis, the telescopic testing occasionally terminates prior to examining the statistics for the quarter-aperture hypotheses. In particular, susceptibility to noise can cause the full-aperture hypothesis to appear more likely than any of the half-aperture hypotheses, thus resulting in an immediate declaration of full-aperture scattering.

Table 6.1. GLLR's for noiseless 2.6m plate response (assuming an PSNR of 20dB).

	Without azimuthal uncertainty	With azimuthal uncertainty
full-aperture	0	0
half-aperture	0.60	1.0-2.3
quarter-aperture	8.4	4.2-7.5

To gain a better understanding of why the full-aperture hypothesis is nearly as likely as the half-aperture hypothesis for the 2.6m plate, consider the basic GLLR in Eq. (5.16). For the 2.6m plate, the zero-crossings of the sinc-like azimuthal response are located such that the sidelobes minimize the reflectivity in the center half-aperture, i.e. the first sidelobes maximally negate the influence of the mainlobe. Thus, the middle half-aperture reflectivity is reduced and produces a correspondingly lower GLLR for this hypothesis. Similarly, the quarter-aperture classification rate "dips" at plate sizes near multiples of 2.6m as the other "negative" sidelobes have a locally maximum negative effect on the middle half-aperture[3]. This problem could be alleviated by relaxing the hard decision made at each scale in the telescopic hypothesis testing procedure. However, such a modification is not used here because, as discussed next, this artifact is not an issue in practice due to the effect of azimuthal uncertainty. Furthermore, such a change in the telescopic testing procedure would reduce robustness to interfering scatterers as discussed in Section 5.3.

The anisotropy plot in Figure 6.2(a) corresponds to the case in which the plate is oriented exactly broadside to the center of the aperture. To incorporate the effects of azimuthal uncertainty, i.e. slight rotations of the scatterer, the experimental setup is modified as follows. The SAR data is synthesized with the same parameters as before except that the data is generated over an aperture three times as large, i.e. the azimuthal response is generated over 8.4°. From this extended aperture, a section measuring 2.8° is randomly chosen from a uniform distribution in a fashion that provides a maximal amount of uncertainty subject to the constraint that the mainlobe of the azimuthal response is always observed. In particular, the support of the uniform distribution is given according to the following cases:

---

[3]These other ripples are too small to be seen in Figure 6.2(a) but are apparent if the noise level is increased or the scaling of the vertical axis is changed.

- if the plate size is less than two wavelengths (i.e. the plate size is less than 6.25 cm), then the scatterer is considered isotropic and any 2.8° section of the extended aperture can be selected,

- if the plate size is larger than two wavelengths but still small enough such that the half-power bandwidth of the azimuthal response is larger than 2.8°, then any section of the extended aperture within the half-power band can be chosen, and

- if the plate size is sufficiently large such that the half-power bandwidth of the azimuthal response is smaller than 2.8°, then any section of the extended aperture containing the half-power band can be chosen.

The anisotropy plot incorporating this azimuthal uncertainty is shown in Figure 6.2(b). Interestingly, performance is improved for larger plate sizes by the addition of azimuthal uncertainty. In particular, the "dip" at 2.6m in Figure 6.2(a) is reduced. Looking at the GLLR's in the second column of Table 6.1 for the 2.6m plate (assuming 20dB PSNR), it is apparent that it is *not* that the quarter-aperture GLLR increases (in fact, the GLLR decreases slightly to typically vary between 4.2 and 7.5) but that the half-aperture GLLR increases to typically vary between 1.0 and 2.3. This increase occurs because the exact broadside orientation represents a worst-case scenario for the half-aperture hypothesis (relative to the full-aperture hypothesis) due to the maximal negative effect that the first (and other odd number) sidelobes have on the middle half-aperture reflectivity. Azimuthal uncertainty, however, perturbs this degenerate alignment.

In addition to azimuthal uncertainty, the effect of location uncertainty can also be incorporated into the anisotropy plot. In particular, location uncertainty is simulated by synthesizing a response for a random location and then using a peak extraction procedure to determine the location to be used in the anisotropy test. Modeling location uncertainty in this fashion closely mimics the location uncertainty which arises in peak-based ATR algorithms[25, 34, 13, 9, 23]. The anisotropy plot incorporating both azimuthal and location uncertainty is shown in Figure 6.2(c). Not surprisingly, there is a slight degradation in performance, particularly for larger plate sizes. This degradation is due to the fact that large plates span several resolution cells over which the reflectivity is approximately constant. Thus, the peak extraction is susceptible to noise over these resolution cells and frequently reports a location that, while on the scatterer, does not correspond to the scattering center. The resulting peak location error introduces a modulation on the observed azimuthal response which distorts the anisotropy classification[4]. In particular, for larger apertures, the linear phase on the azimuthal response prevents the sub-aperture reflectivity estimates from reaching their true value due to the destructive interference that occurs when integrating over the sub-apertures. Thus, the energy from larger sub-apertures is biased towards smaller

---

[4]One might choose to model this location error as a nuisance parameter and address the error with a statistical approach. We do not, however, take this course due to the difficulties involved with the nonlinear effects the error has on the observed response. We thus view location error as an artifact of the peak-extraction algorithm and leave the problem for the peak-extraction to address.

values. Smaller sub-apertures, on the other hand, are less susceptible to this phase variation due to the shorter integration interval and thus give stronger reflectivity estimates. This interference is the reason that errors in scatterer location tend to bias the anisotropy classification towards higher degrees of anisotropy.

The anisotropy plots in Figure 6.2 show that even in the presence of azimuthal and location uncertainty, the anisotropy attribution exhibits the strong dependence of anisotropy on plate size that is predicted by canonical scattering models. Furthermore, the anisotropy classifications are consistent with these models as illustrated by the azimuthal responses in Figure 6.3. One would naturally expect the dependencies in the anisotropy plot to be preserved as the aperture size is increased. As discussed in Section 6.3, this is the case, but this seemingly trivial property turns out to be quite significant when contrasted to the behavior of volumetric anisotropy which is discussed next.

■ 6.2 Volumetric Anisotropy

Similar to the anisotropy plots for the flat plate scatterer, a notion of scatterer "size" is used as the dependent variable in the anisotropy plots for a volumetric scatterer. The "size" in this case is the size of the support for the pdf generating the locations of individual isotropic scatterers composing the volumetric scatterer. The radar and anisotropy test parameters are the same as those used for the geometric anisotropy plots in Figure 6.2. For each Monte Carlo sample, a volumetric scatterer is taken as a collection of random point scatterers generated according to the following parameters:

- The density of the scatterers in terms of the average number of scatterers per resolution cell.

- The support for the uniform distribution used to produce the down-range and cross-range location of the scatterers. The support of the pdf producing the down-range location of the scatterers is taken as one resolution cell. The support of the pdf producing the cross-range location is the dependent variable in the anisotropy plot.

- The azimuthal uncertainty which is taken to be one quarter of the aperture, i.e. $0.7°$.

- The distribution producing the real-valued reflectivity[5] of the scatterers. The particular distribution used is a Gaussian, with a mean and variance of 1, truncated to prevent negative values.

- The PSNR of the collective scattering which is taken to be 20dB.

---

[5]The nonzero phase on the measured reflectivity comes from the displacement of the scatterers in range and cross-range.

Figure 6.4 shows the anisotropy plots for scattering densities of 1, 10, and 100 scatterers per resolution cell or equivalently 1, 10, and 100 scatterers per foot in cross-range under the MSTAR parameters. From the differences among these plots, it is immediately clear that the behavior of volumetric anisotropy is heavily dependent on the spatial density of the constituent scatterers. In particular, the plots go from saying that anisotropy is independent of scatterer size for low scattering densities to saying that anisotropy is highly dependent on scatterer size at densities sufficiently high that the volumetric scatterer resembles a flat plate. This behavior is quite reasonable. For an average of one scatterer per resolution cell, one should expect the multiple scatterer model to be able to account for the responses from the individual isotropic scatterers, thus producing the full-aperture classification for the focused location regardless of the scatterer size. However, these low-density volumetric scatterers are not always declared full-aperture because constituent scatterers may be considerably closer than one resolution cell apart or the peak extraction may cause the algorithm to focus to the wrong location resulting in an anisotropic attribution. For a high scattering density of 100 scatterers per resolution cell, there are so many scatterers, whose spatial locations are drawn from the uniform distribution, that the composite response very closely resembles that from a flat plate with a size given by the spatial support for the volumetric scatterer. Deviations from the anisotropy plot for the flat plate can be attributed to the fact that

1. the scatterers are not regularly spaced since their locations are random and

2. the amplitudes on the scatterers are random.

The anisotropy plot for a density of 10 scatterers per resolution cell is a transitional phase between the two extreme cases of 1 and 100 scatterers per resolution cell where behaviors from both regimes are evident.

From the anisotropy plots in Figure 6.4, one can conclude that low density volumetric scattering is more useful than high density scattering because the former minimizes the classification rate of volumetric anisotropy and more accurately represents the response as arising from a collection of point scatterers instead of a single anisotropic scatterer. The reduction in volumetric anisotropy attributions allows for scatterers that are declared anisotropic to be more confidently associated with geometric scatterers from which reliable physical information can be inferred. One may at first be inclined to believe that we have to accept a given incidence of volumetric anisotropy as volumetric scatterer density is a fixed parameter, but this is not the case. The scattering density in our analysis is measured with respect to the number of scatterers per *resolution cell*. The size of the resolution cell is inversely related to the size of the aperture. Thus, by using wide-aperture data, one should be able to reduce volumetric scattering density and hence the detection rate for volumetric anisotropy as discussed in the next section.
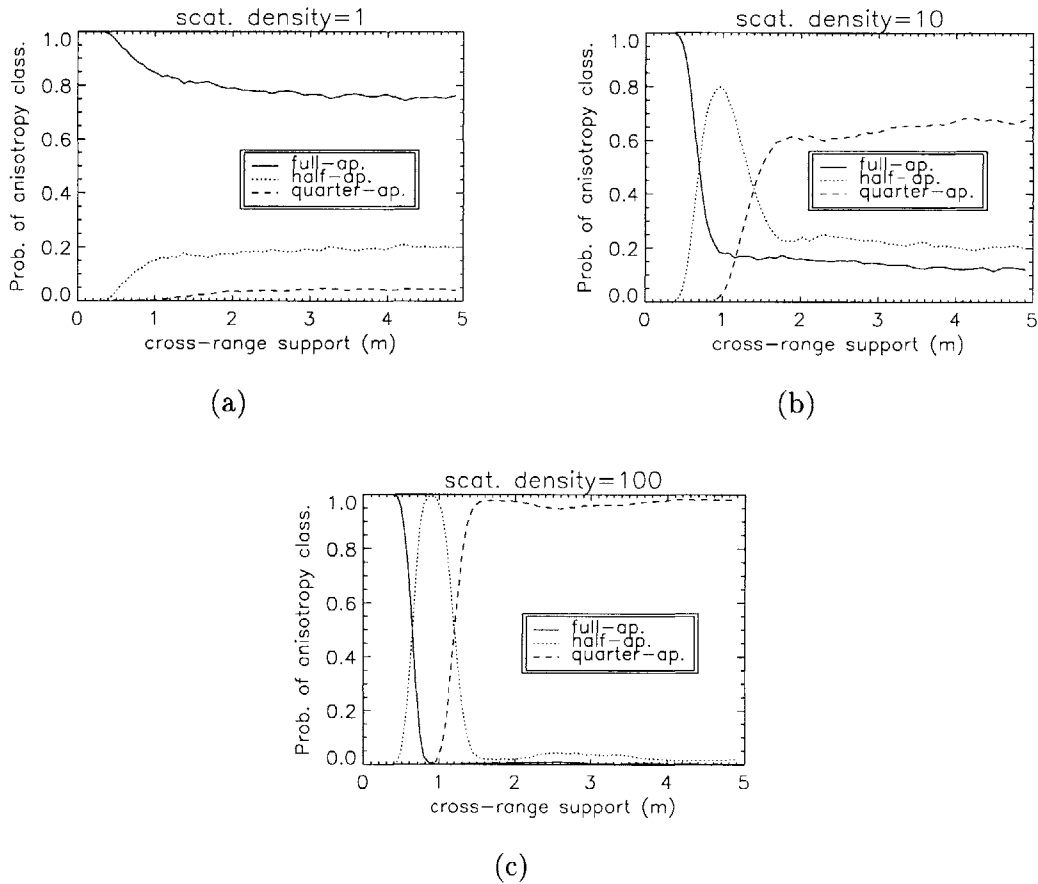
Figure 6.4. Anisotropy plots for volumetric scatterers with a scattering density of (a) 1, (b) 10, and (c) 100 scatterers per resolution cell.

■ 6.3 Extensions to Wide-Aperture Data

The results for geometric anisotropy attribution in Figures 6.2 and volumetric anisotropy attribution in Figure 6.4 are promising in that they imply that by increasing the size of the aperture, one can preserve the detection rate of geometric anisotropy classifications while reducing the rate of volumetric anisotropy classifications. This effect would then allow for a more thorough analysis of predictable anisotropic geometric scatterers in an ATR algorithm using wide-aperture data.

To see why increasing the aperture length decreases the incidence of volumetric (but not geometric) anisotropy, note that for low-density volumetric scattering (per resolution cell), Figure 6.4(a) shows that the anisotropy classification is usually full-aperture regardless of the cross-range support. However, at medium to high scattering densities, the number of anisotropy classifications increase. Wide-aperture data should,

thus, reduce the number of anisotropy classifications because the larger aperture reduces the scattering density in terms of the number of scatterers per resolution cell. For example, increasing the aperture size by a factor of ten decreases the resolution cell size and hence scattering density by the same factor. From the anisotropy plots in Figure 6.4, it is apparent that such a factor should have a significant impact on the rate of volumetric anisotropy classifications. Geometric scatterers, on the other hand, can be considered to be composed of a continuum of infinitely small scatterers (with respect to practical imaging resolutions), so increasing the resolution by an order of magnitude, or more, does not detract from the high scatterer density thus preserving the structure of the anisotropy plot.

The above is an image (or spatial) domain explanation for why wide-aperture data would reduce the rate of volumetric anisotropy while preserving that of geometric anisotropy. There is an additional azimuthal domain interpretation. Consider a scatterer that is declared anisotropic based on narrow-aperture data. This classification is made because most of the energy is concentrated in one region of the aperture. Increasing the aperture length, however, increases the range over which the azimuthal response is observed. For a volumetric scatterer, there is likely to be a considerable energy contribution in the newly added sections of the aperture because the underlying response is not truly unimodal as depicted in Figure 6.1(b). However, a geometric scatterer with a unimodal azimuthal response should have little energy in the newly added aperture sections due to the these measurements capturing the tails of the azimuthal response. Furthermore, many anisotropic geometric scatterers exhibit sinc-like oscillations in the tails of their response (which nearly integrate to zero in the sub-apertures) while many volumetric scatterers do not. Thus, for volumetric scatterers, there is likely to be considerable energy in newly appended aperture sections which may change the anisotropy classification, but for geometric scatterers, the newly appended aperture sections will contain little energy thereby leaving the anisotropy attribution unchanged.

To verify this conjecture empirically, anisotropy plots for the geometric and volumetric scatterers are generated in the same fashion as before except that the aperture size is increased by a factor of eight. This modification naturally changes the sub-aperture pyramid used. To be consistent with the narrow-aperture results, we use a four-level, half-overlapping, half-aperture pyramid which consists of

- the 22.4° full-aperture at scale 0,

- 2.8° $1/8^{th}$-apertures at scale 1,

- 1.4° $1/16^{th}$-apertures at scale 2, and

- 0.7° $1/32^{th}$-apertures at scale 3.

Furthermore, again for consistency, the 22.4° full-aperture classification is not allowed to be declared in the anisotropy attribution process as this degree of anisotropy is not

available in the narrow-aperture setting[6]. The latter three hypotheses correspond to the full, half, and quarter-apertures in the narrow-aperture case and are thus consistent in terms of the physical characterization of the underlying scatterer. To prevent any confusion between the narrow and wide-aperture cases, these hypotheses will be referred to with their measurements in degrees (i.e. 2.8°, 1.4°, and 0.7°) instead of normalized sub-aperture length. To be consistent in the geometric and volumetric settings, azimuthal uncertainty in both cases is modeled with a uniform distribution over 1/4 of the narrow-aperture, i.e. 0.7°. For consistency in the noise level between the narrow and wide-aperture settings, the same noise spectral density (in the azimuthal domain) is used for the wide-aperture data as in the narrow-aperture case. Thus, for isotropic scatterers, the wide-aperture PSNR is actually higher than in the narrow-aperture case, and for anisotropic scatterers, the wide-aperture PSNR is lower than in the narrow-aperture case due to the coherent averaging involved[7].

The wide-aperture geometric anisotropy plot incorporating azimuthal and location uncertainty is displayed in Figure 6.5. The two plots in Figure 6.5 correspond to two different peak extraction methods. Figure 6.5(a) is for a coarse-resolution peak extraction which is done using the center 2.8° narrow-aperture. Figure 6.5(b) is for a fine-resolution peak extraction which is done using the full-aperture. As predicted, Figure 6.5(a) looks quite similar to the narrow-aperture anisotropy plot in Figure 6.2(c) showing a strong dependence on the underlying plate size. Figure 6.5(b) also looks quite similar to Figure 6.2(c), but the classification rates for 1.4° and 0.7° anisotropy have dropped slightly for larger plate sizes. One can also see that the "transition band" between anisotropy classifications is wider for the fine-resolution peak extraction. This degradation in performance is due to the fact that in the high resolution setting, the plate sizes corresponding to the degrees of anisotropy being tested span several resolution cells. Over these resolution cells, the reflectivity is nearly constant which results in a peak location that comes from a uniform distribution over the scatterer. This location error results in a modulation of the observed azimuthal response which corrupts the anisotropy attribution as discussed at the end of Section 6.1 for large plate sizes. This location error is not as significant for the coarse-resolution peak extraction method because the resolution cell size is the same order of magnitude as the plate sizes which correspond to our tested degrees of anisotropy. Thus, the coarse spatial averaging usually covers a significant portion of the scatterer and is thus more likely to return a peak location near the center of the flat plate.

Now, consider the behavior of volumetric scatterers in the wide-aperture setting. The nominal scattering density examined is 10 scatterers per coarse resolution cell or equivalently 1.25 scatterers per fine-resolution cell in the fine-resolution setting. Figure 6.6 displays the wide-aperture anisotropy plots. Figure 6.6(a) is for the coarse-

---

[6]We note that the omission of the full-aperture classification is solely to aid our comparison of narrow and wide-aperture results. In applied settings, we recommend that the full-aperture always be used to make maximum use of the aperture and provide maximum imaging resolution.

[7]The difference in PSNR is due to the signal processing gain associated with a larger aperture. The underlying measurement conditions are matched in the narrow-aperture and wide-aperture scenarios.
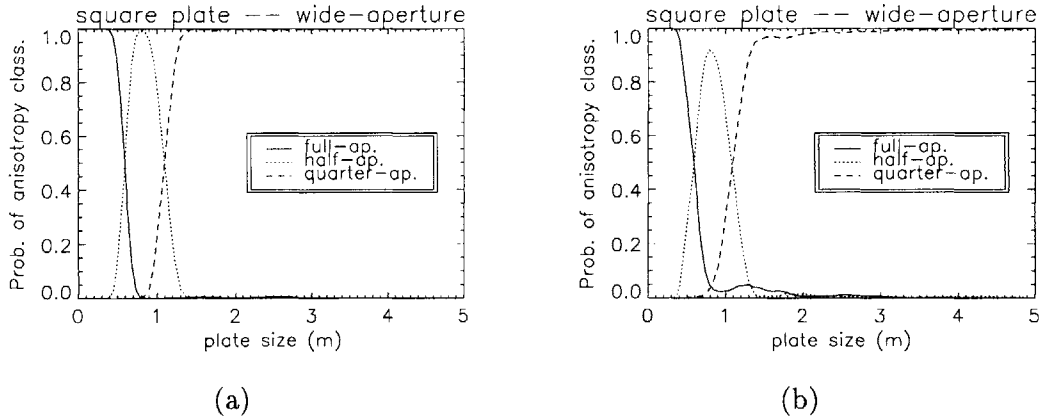
Figure 6.5. Anisotropy plots based on wide-aperture data for the canonical flat plate scatterer using (a) the coarse-resolution peak extraction and (b) the fine-resolution peak extraction.

resolution peak extraction using the center 2.8° narrow-aperture, and Figure 6.6(b) is for the fine-resolution peak extraction using the full-aperture. The volumetric anisotropy plot in Figure 6.6(b) behaves as conjectured, reducing the rate of volumetric anisotropy attributions and bearing a close resemblance to the anisotropy plot for a scattering density of one scatterer per resolution cell in the narrow-aperture setting. However, Figure 6.6(a) shows virtually no reduction in the rate of anisotropy classifications. Recalling the image domain rationale that wide-aperture data reduces volumetric anisotropy because the use of the larger aperture allows one to better resolve constituent scatterers, this result is not surprising. When the coarse-resolution peak extraction is used, the algorithm is not focusing on peaks that could be resolved in the high-resolution regime thus disregarding that information which is beneficial.

The dependence on the peak extraction process can be summarized by saying that the coarse-resolution peak extraction preserves the anisotropy classification rates for both geometric and volumetric scatterers, while the fine-resolution peak extraction slightly decreases the anisotropy classification rate for geometric scatterers but significantly reduces the rate for volumetric scatterers. However, by utilizing both the coarse and fine-resolution peak extractions, we should be able to obtain the desired behavior for both geometric and volumetric anisotropy. In particular, one can select which peak extraction method to use based on scatterer size inferred from the anisotropy attribution. Recall that the fine-resolution peak extraction degrades the attribution for geometric scatterers because the finer resolution provides a worse scatterer location estimate for large scatterers. This location error results in over-estimation of the degree of anisotropy. However, this type of location error occurs only for large scatterers that span several pixels and have high degrees of anisotropy. Thus, for these scatterers it can only be beneficial to repeat the anisotropy attribution using the coarse-resolution peak location which gives a more accurate estimate of the scattering center. Consequently,
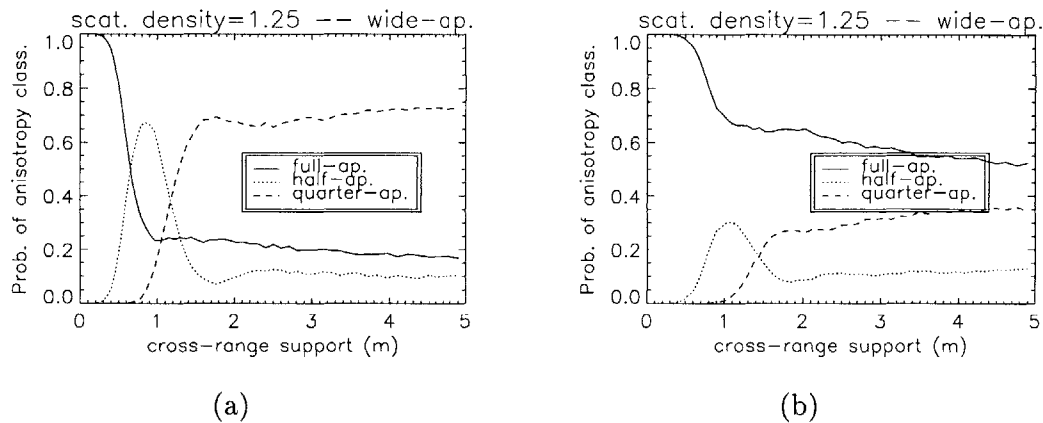
Figure 6.6. Anisotropy plots based on wide-aperture data for volumetric scatterers with scattering density of 10 scatterers per coarse-resolution cell using (a) the coarse-resolution peak extraction and (b) the fine-resolution peak extraction.

a decision-directed procedure can be used to determine which scale peak extraction should be used in the anisotropy attribution. In particular, one can first use the peak location from the fine-resolution peak extraction, and if the extraction results in an anisotropic declaration (conveying that there is likely to be a large geometric scatterer there), then one can defer to the peak location from the coarse-resolution peak extraction[8]. Note that this approach preserves the high rate of the minimally anisotropic $2.8°$ classification for volumetric scatterers since these classifications do not defer to the coarse-resolution peak extraction. The high rates for the anisotropic $1.4°$ and $0.7°$ classifications on large geometric scatterers, however, are also preserved since their initial anisotropy classification is usually over-estimated and thus defers to the location estimate given by the coarse-resolution peak extraction which is more accurate for these larger scatterers.

The anisotropy plots using this hybrid coarse and fine-resolution peak extraction are given in Figure 6.7 which display the predicted behavior. In particular, the hybrid peak extraction offers a unified approach which preserves the high rate of geometric anisotropy attributions but reduces the rate of volumetric anisotropy attributions. Thus, as postulated at the beginning of this section, the detection rate of volumetric anisotropy can be reduced while preserving the rate of the more meaningful geometric anisotropy by increasing the length of the aperture.

---

[8]Note that there is also the issue of consistency of the peak extractions, i.e. the peaks extracted from the coarse and fine-resolution methods may correspond to different scatterers that are not even in the same resolution cell. In order to overcome this, we first estimate the coarse-resolution peak location and then limit the fine-resolution peak location to reside within the resolution cell corresponding to the coarse-resolution peak extraction. Thus, both the coarse and fine-resolution extractions are guaranteed to come from the same coarse resolution cell.
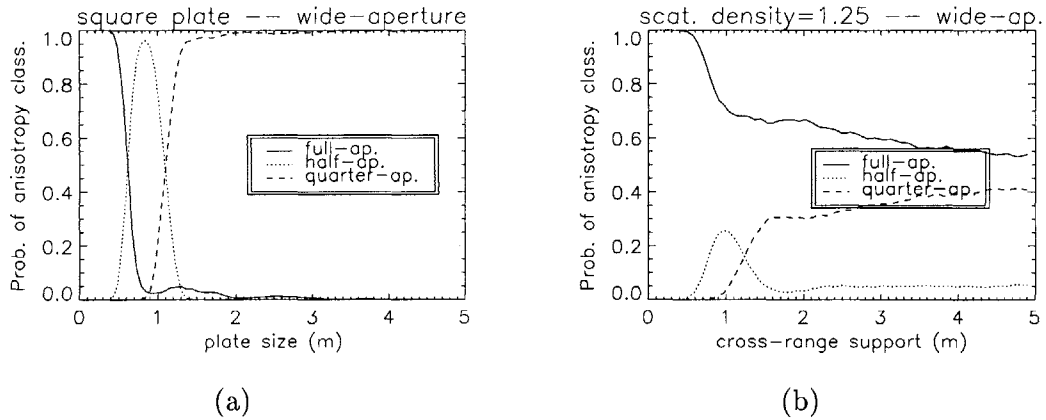
Figure 6.7. Anisotropy plots using the hybrid-resolution peak extraction for (a) the canonical flat plate scatterer and (b) the volumetric scatterer with 10 scatterers per coarse resolution cell.

■ 6.4 Additional Wide-aperture Hypotheses

Up to this point, for clarity in the comparison of narrow and wide-aperture anisotropy attributions, the same hypothesis set consisting of $2.8°, 1.4°$, and $0.7°$ sub-apertures has been used for both the narrow and wide-aperture data. However, in addition to reducing the rate of volumetric anisotropy while preserving that of geometric anisotropy, wide-aperture data allows for additional anisotropy hypotheses, up to the size of the larger aperture, to be defined. The inclusion of these less anisotropic hypotheses does not affect the argument for reducing the rate of volumetric anisotropy, although these additional hypotheses do allow for a more accurate anisotropy characterization of smaller scatterers.

To demonstrate this improvement, Figure 6.8 displays the geometric anisotropy plots for the flat plate scatterer using the *augmented hypothesis set* consisting of $22.4°$, $11.2°$, $5.6°$, $2.8°$, $1.4°$, and $0.7°$ sub-apertures instead of the *restricted hypothesis set* (composed of $2.8°$, $1.4°$, and $0.7°$ sub-apertures) that has been used up to this point. The range of plate sizes used here is reduced to 0m-2.5m to better illustrate the behavior for small scatterers. The behavior observed in Figure 6.8 is a natural extension of the behavior in the geometric anisotropy plots in Figures 6.5 and 6.7(a). In particular, the only significant difference is related to the $2.8°$ classification in the restricted hypothesis set. Under the augmented hypothesis set, scatterers which were previously characterized as $2.8°$ (the minimum degree of anisotropy in the restricted hypothesis set) are now classified as $22.4°$, $11.2°$, $5.6°$, or $2.8°$. The $22.4°$ classification dominates at the smallest plate sizes, and the dominant attribution transitions to $11.2°$, $5.6°$, and $2.8°$ as plate size is increased as one would expect. Interestingly, the $22.4°$ classification probability starts out noticeably lower when using the coarse-resolution peak extraction in Figure 6.8(a) than when using the fine-resolution peak extraction or hybrid peak ex-

traction in Figures 6.8(b) and (c). The reason for this behavior is that for the smaller plate sizes, the coarse-resolution peak extraction introduces location errors on the order of the size of the coarse resolution cell, i.e. one foot. As previously noted, location error introduces a modulation on the response which biases the classification result towards higher degrees of anisotropy. However, since this location error is small, the only effect is to misclassify some of the small plates as 11.2° instead of 22.4°. Note that in the hybrid peak extraction approach, the fine-resolution peak extraction is analyzed first and returns a 22.4° attribution indicating the presence of a small scatterer. Thus, the hybrid peak extraction does not defer to the coarse-resolution extraction resulting in the correct attribution.

(a)

(b)

(c)
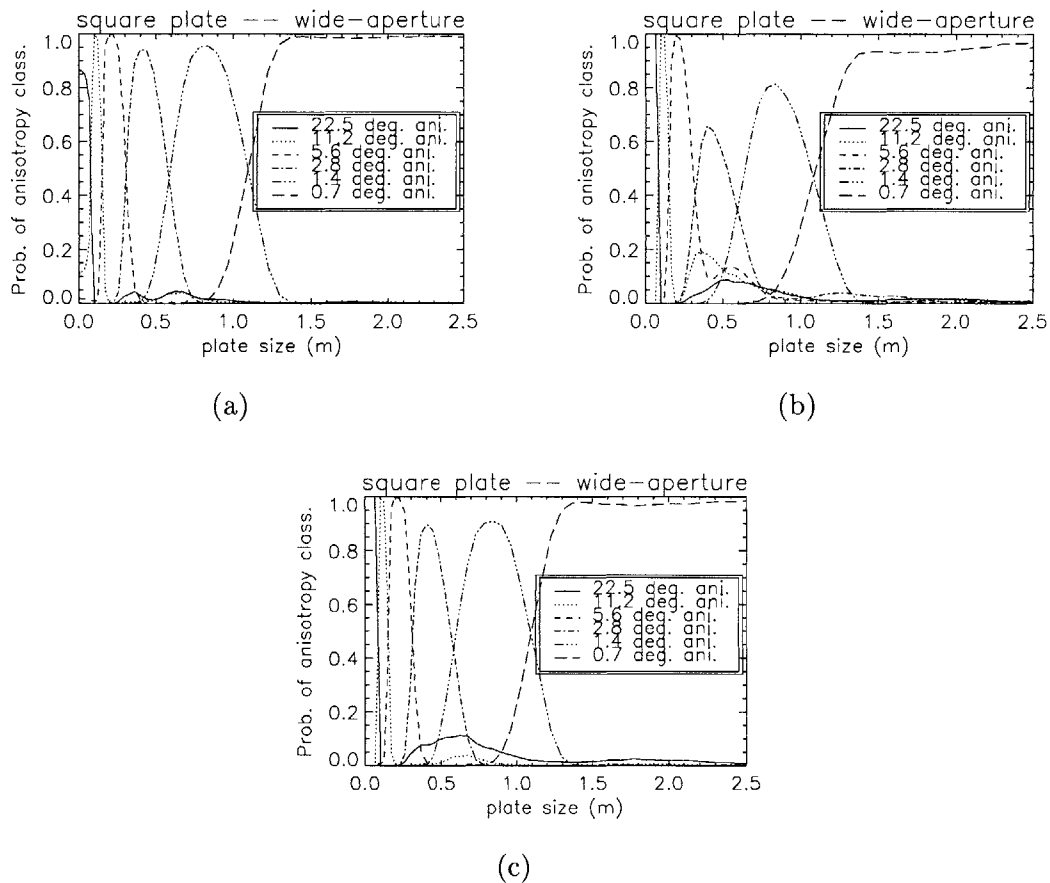
Figure 6.8. Anisotropy plots for the canonical flat plate scatterer using the augmented hypothesis set and (a) the coarse-resolution peak extraction, (b) the fine-resolution peak extraction, and (c) the hybrid-resolution peak extraction.

Figure 6.9 displays the volumetric anisotropy plots using the augmented hypothesis set. Again, the behavior of each plot is a natural extension of Figures 6.6 and

6.7(b). In particular, the coarse-resolution peak extraction in Figure 6.9(a) shows a high rate of anisotropic classifications, but the use of the fine-resolution peak extraction in Figure 6.9(b) allows the anisotropy test to account for and resolve constituent isotropic scatterers. The hybrid peak extraction utilizes the first stage fine-resolution peak extraction to preserve the low rate of volumetric anisotropy as demonstrated in Figure 6.9(c). Thus, as argued earlier, the use of the augmented hypothesis set with wide-aperture data allows for a more accurate characterization of anisotropy for smaller geometric scatterers while maintaining a low incidence of volumetric anisotropy.

(a)

(b)

(c)
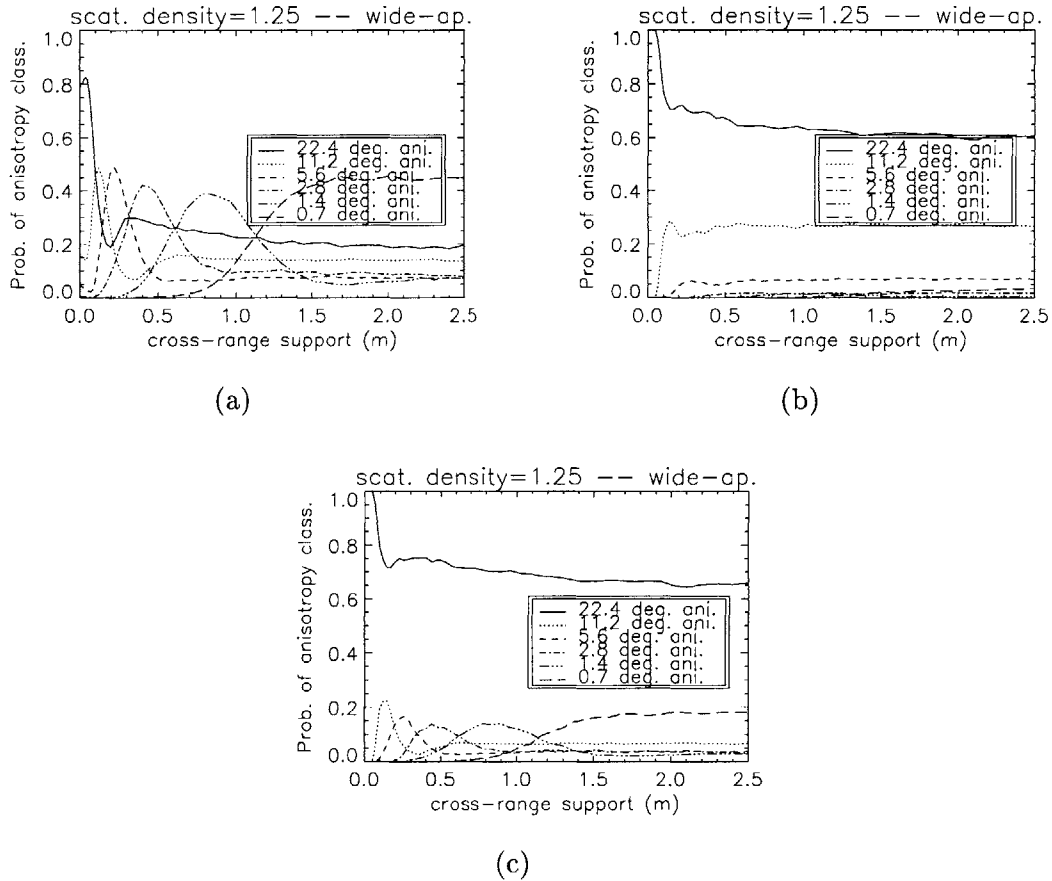
Figure 6.9. Anisotropy plots for volumetric scatterers with a scattering density of 10 scatterers per coarse resolution cell using the augmented hypothesis set and (a) the coarse-resolution peak extraction, (b) the fine-resolution peak extraction, and (c) the hybrid-resolution peak extraction.

■ 6.5 Wide-aperture GUS Data

This section applies the anisotropy attribution to peak scatterers extracted from the measured wide-aperture Grayling Ultra-wideband SAR (GUS) data set. The data has a resolution of 1.9 inches in down and cross-range achieved using a 4GHz bandwidth, a 10GHz center frequency, and a 26° aperture. For narrow-aperture comparisons, the middle 3.3° section of this aperture (i.e. the middle eighth-aperture) is taken as the narrow-aperture data. For each narrow-aperture peak extraction, the anisotropy test is applied with the restricted hypothesis set composed of 3.3°, 1.6°, and 0.8° sub-apertures which correspond to the full, half, and quarter narrow-apertures. For the wide-aperture data, the hybrid-resolution peak extraction method described in the previous section is used to determine peak locations. Because the wider aperture allows for more anisotropy hypotheses to be defined, there are several choices for the hypothesis set in this regime. Two particular sets are examined. One choice is the restricted hypothesis set used in the narrow-aperture case. An alternative is the augmented hypothesis set obtained by taking all the sub-apertures in the half-overlapping, half-aperture pyramid for the wide-aperture, i.e. the degrees of anisotropy in the augmented hypothesis set are {26°, 13°, 6.5°, 3.3°, 1.6°, 0.8°}. Figures 6.10 and 6.11 show the extracted peaks and their associated anisotropy attribution using the narrow and wide-aperture data for two different vehicles. Figures 6.10 and Figures 6.11 show the standard log-magnitude reflectivity image using a coarse imaging resolution[9]. The symbol overlays in Figures 6.10(a) and 6.11(a) specify the location and anisotropy of each peak extraction based on the narrow-aperture data. The symbol overlays in Figures 6.10(b) and 6.11(b) specify the location and anisotropy of each peak extraction based on the wide-aperture data. For both figures (a) and (b), a black circle denotes 3.3° anisotropy, a square denotes 1.6° anisotropy, and a triangle denotes 0.8° anisotropy. To denote the additional wide-aperture 26°, 13°, and 6.5° attributions in figure (b), we again use the circle, square, and triangle, respectively, but display them in white instead of black.

In each of the images, the peak extractions that are labeled as anisotropic are frequently associated with an apparent geometric scatterer. In particular, the collections labeled "A" and "B" in Figure 6.10(b) are large geometric scatterers[10] and are appropriately characterized as anisotropic. Note that based on the narrow-aperture data, the two scatterers labeled "A" in Figure 6.10(a) are attributed as 3.3°, but there are no means of telling if the underlying scatterers are actually 3.3° or perhaps less anisotropic since 3.3° is the lowest degree of anisotropy available in the narrow-aperture setting. However, using the augmented hypothesis set afforded by the wide-aperture, it is apparent that the degree of anisotropy for both scatterers is in fact approximately 3.3°, and not much more. This example demonstrates how the additional hypotheses af-

---

[9]The coarse resolution is used for the wide-aperture case (even though a finer imaging resolution can be achieved) due to restrictions on publishing the GUS data.

[10]That the scatterers in "A" and "B" of Figure 6.10(b) and "A" through "C" of Figure 6.11(b) are geometric in nature is evident from the omitted (due to publishing restrictions) high-resolution imagery.
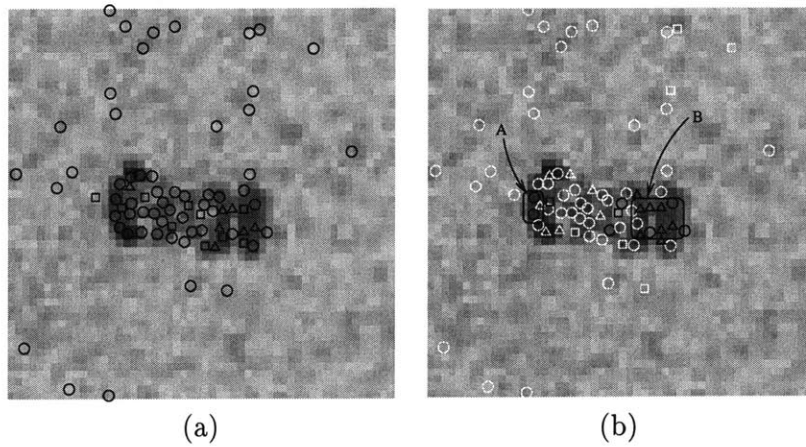
Figure 6.10. Anisotropy characterization of peaks from wide-aperture data for vehicle 1. Standard log-magnitude image with peak anisotropy attributions based on (a) narrow-aperture data and (b) wide-aperture data.
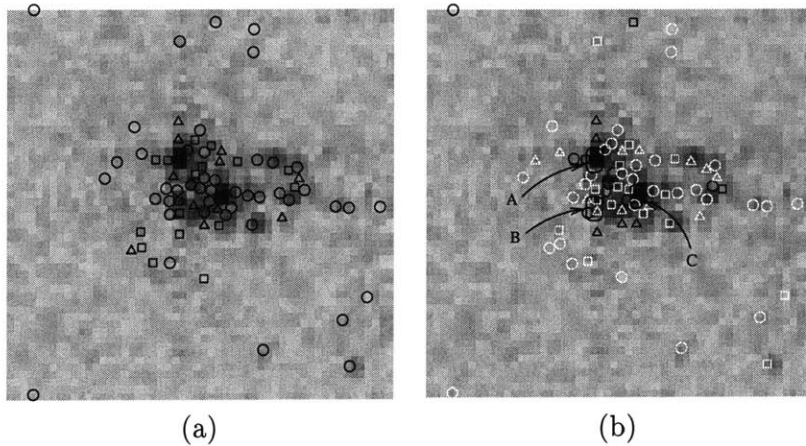
Figure 6.11. Anisotropy characterization of peaks from wide-aperture data for vehicle 2. Standard log-magnitude image with peak anisotropy attributions based on (a) narrow-aperture data and (b) wide-aperture data.

forded by the larger aperture allow more information about the scatterer anisotropy to be conveyed.

Another instance of large geometric scattering is the collection labeled "B" in Figure 6.10. These peak extractions appear to be associated with large flat geometric scatterers, which are reflected in their anisotropy attributions most of which are 0.8°. Collections "A"-"C" in Figure 6.11(b) highlight several other prominent scatterers which appear to be geometric in nature. It is also interesting to note that in Figures 6.10 and 6.11, many scatterers (both on and off the vehicle) are attributed as

26° in the wide-aperture data. The vast majority of these attributions are reasonable since they appear to be small point scatterers or clutter in the high-resolution image.

So far, we have observed the conjectured benefit of wide-aperture data over narrow-aperture data on the anisotropy characterization of geometric scatterers, i.e. the preservation of higher degrees of anisotropy observed in the narrow-aperture data and the elaboration of the least anisotropic narrow-aperture hypothesis by augmenting the hypothesis set with larger sub-aperture hypotheses. This result is predicted by the analysis presented earlier. That analysis also predicts a reduction in the rate of volumetric anisotropy as aperture size is increased. To test this behavior with the GUS data, scenes of imaged clutter are used with the same 26° and 3.3° apertures as before. In particular, the three clutter scenes shown in Figure 6.12 are examined. For these scenes, the narrow-aperture anisotropy characterization is displayed in the left-hand side and the wide-aperture characterization is displayed in the right-hand side. The set of hypotheses and respective graphical labels are the same as for the geometric scatterer case in Figures 6.10 and 6.11. It is immediately clear that there is a significant reduction in the rate of higher degrees of anisotropy using the wide-aperture data. To explicitly see that this reduction is due to these scatterers being resolved into individual isotropic scatterers (in contrast to their simply being uniformly distributed over the larger hypothesis set), Tables 6.2 and 6.3 show the tabulated anisotropy correspondences between the two aperture settings. Table 6.2 shows the correspondences between the narrow and wide-aperture anisotropy attributions when both data sets are restricted to use the same set of narrow-aperture hypotheses. The empirical results here strongly argue that the use of the wider aperture has resolved the scatterers which are attributed as 1.6° and 0.8° under the narrow-aperture into individual minimally anisotropic 3.3° scatterers. In particular, of the 49 scatterers that are attributed as anisotropic when using the narrow-aperture data, only one is attributed as anisotropic when using the wide-aperture data.

Table 6.2. Anisotropy classification correspondences for clutter with narrow and wide-aperture data using the same anisotropy hypotheses.

Narrow-ap. \ Wide-ap.	3.3°	1.6°	0.8°
3.3°	176	0	0
1.6°	45	0	1
0.8°	3	0	0

Table 6.3 shows the same type of correspondence as Table 6.2 except that the augmented hypothesis set is used for the wide-aperture data. The tendency to resolve anisotropic scatterers into isotropic scatterers is again observed in this table. In addition, many of the 3.3° anisotropy classifications are further resolved into the less anisotropic classifications with the majority of them assigned to the 26° minimal de-
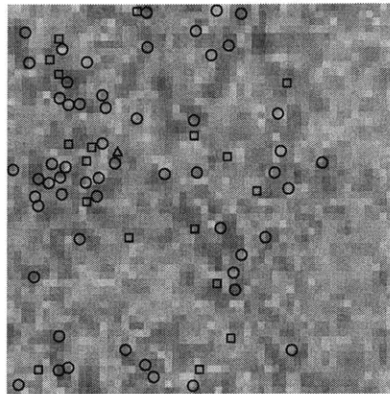
Clutter Scene 1 – narrow-aperture   Clutter Scene 1 – wide-aperture
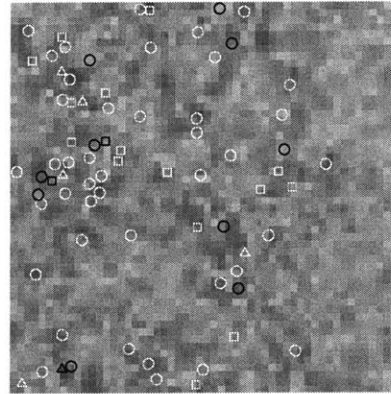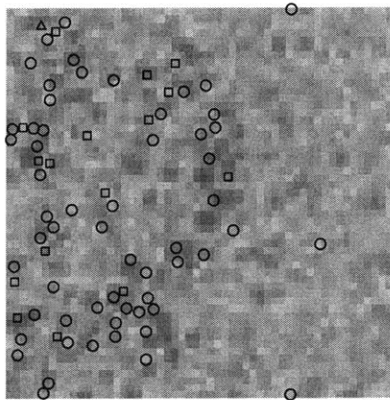
Clutter Scene 2 – narrow-aperture   Clutter Scene 2 – wide-aperture
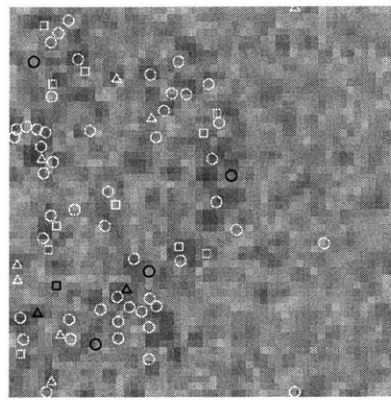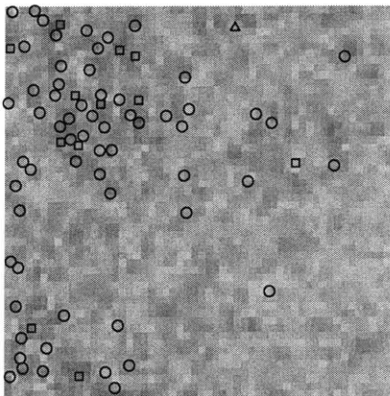
Clutter Scene 3 – narrow-aperture   Clutter Scene 3 – wide-aperture
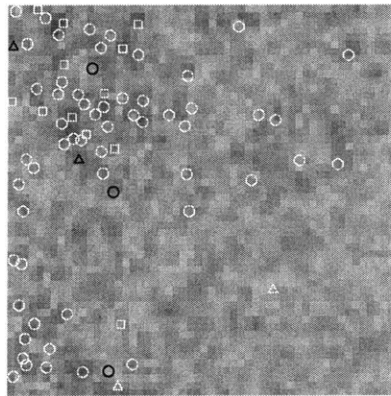
Figure 6.12. Anisotropy characterization of peaks from 3 instances of clutter in the GUS data. Left: Narrow-aperture anisotropy attributions. Right: Wide-aperture anisotropy attributions.

gree of anisotropy. Thus, the predicted reduction in the rate of unstable volumetric anisotropy is observed in real wide-aperture data.

Table 6.3. Anisotropy classification correspondences for clutter with narrow and wide-aperture data using the augmented hypothesis set for wide-aperture data.

Narrow-ap. \ Wide-ap.	26°	13°	6.5°	3.3°	1.6°	0.8°
3.3°	117	28	10	15	3	3
1.6°	29	8	5	2	0	2
0.8°	1	2	0	0	0	0

Chapter 7

Iterative Anisotropy Attribution

In this chapter, we develop an enhanced, iterative algorithm for attributing anisotropy and estimating sub-aperture reflectivity. The idea behind this algorithm begins with a simple examination of the structure and limitations of the anisotropy attribution procedure developed in Chapters 5 and used in Chapter 6. Specifically, that algorithm produces an anisotropy classification and a reflectivity estimate for each pixel by using a pyramid of sub-aperture reflectivity measurements. However, because these sub-aperture measurements use only a fraction of the full-aperture, they have correspondingly coarser resolutions. These larger resolution cells allow the sub-aperture measurements to be corrupted by scatterers that are located within these larger cells but outside the finest resolution cell. This corruption from neighboring scatterers results in both a coarse imaging resolution and incorrect anisotropy attributions.

In Chapter 5, we described an approach to deal with this "nuisance" effect of interference from scatterers in neighboring finest-resolution pixels. In particular, we considered the estimation of anisotropy and reflectivity at a particular pixel (which we will refer to as the *focused pixel*). In processing the sub-aperture measurements corresponding to that focused pixel, we also included a set of additional unknown parameters corresponding to the reflectivities in an entire symmetric window of neighboring cross-range (finest-resolution) pixels. We then performed a generalized likelihood computation. This computation first estimates the reflectivities of the neighboring pixels in an MAP fashion. These estimates are then used to subtract out the effects of the interfering pixels in the anisotropy attribution of the focused pixel. As we will see, the approach described in this chapter also performs a GLLR-based attribution and estimation but with some differences to improve the modeling of neighboring scatterers.

The algorithm developed in Chapter 5, while producing very useful results, suffers from two potential inconsistencies. The first concerns the physical size of scatterers with highly anisotropic attributions. In particular, suppose that a given focused pixel has a highly anisotropic scatterer, i.e. the level of anisotropy indicates that the physical extent of the scatterer corresponds to several pixels (which we will call *on-scatterer pixels*) in the finest-scale image. In this case, the reflectivities in these neighboring pixels are not nuisances since these pixels correspond to the same physical scatterer as the focused pixel. Consequently, the GLLR approach of Chapter 5 erroneously subtracts out the effects of these "neighbors". The over-compensation for this single, physically large

scatterer distorts the estimation and attribution of the focused pixel. However, suppose that we have a preliminary estimate of attribution at all of these locations, e.g. by applying the algorithm of Chapter 5. Perhaps the attributions of these neighboring pixels can be used to group the pixels into sets corresponding to individual scatterers in a manner that reflects the relationship between the degree of anisotropy and spatial extent of the scatterer. With such groupings, we can re-perform the anisotropy attribution in a manner that reduces this over-compensation, resulting in more accurate estimates of reflectivity and anisotropy.

The other inconsistency in the algorithm used in Chapter 5 is the assumption that the "nuisance" contributions from neighboring pixels come from weak isotropic scatterers. Clearly, strong neighboring scatterers violate this assumption, but the assumption is also violated by spatially extended, anisotropic scatterers that span several pixels. Contrary to the case of neighboring on-scatterer pixels, we *do* want to remove the effects from these other locations, since these pixels correspond to a different scatterer than the focused pixel (and are hence called *off-scatterer pixels*). One way to do this, would be to take each grouping of off-scatterer pixels (corresponding to individual large scatterers) and subtract off their aggregate effect in a single step. Note that if we were to subtract off the anisotropic effect for each individual pixel in this group, we would again be over-compensating, since the anisotropy already captures the fact that the influence is from an extended scatterer spanning this group. However, an alternative that does not over-compensate is to subtract the effect of each of the pixels in that grouping individually assuming that each of these neighbors is isotropic. While not rigorously the correct thing to do (since this approach neglects the phase consistency across the pixels due to the common underlying scatterer), this approach works well in practice, as we will see.

The remainder of this chapter develops the iterative attribution algorithm and is organized as follows. Section 7.1 provides a high-level description of the algorithm which is based on an aggregation/re-attribution iteration step. Section 7.2 addresses the pixel aggregation task by posing it in a maximum likelihood framework. Section 7.3 then extends the anisotropy test from Chapter 5 to utilize the information from the scatterer aggregation to remove the effects of neighboring scatterers. Section 7.4 provides results on synthesized and MSTAR data to demonstrate the improvement over the methods presented in Chapter 5.

■ 7.1 Algorithm Overview

Figure 7.1 describes the structure of the algorithm developed in this chapter. The algorithm is broken down into three processing stages. Inputs to each processing block are denoted by inward pointing arrows. Arrows coming from the top designate measurement data that do not change as the algorithm progresses. Arrows coming from the left are inputs from another block's output. Thus, these data vary as the iteration progresses. Processing outputs are denoted by arrows leaving a block. Arrows pointing

to the right represent estimated quantities that are to be used by a subsequent processing stage. Arrows pointing downwards denote outputs to the user, i.e. an image of anisotropy attributions or reflectivity estimates.
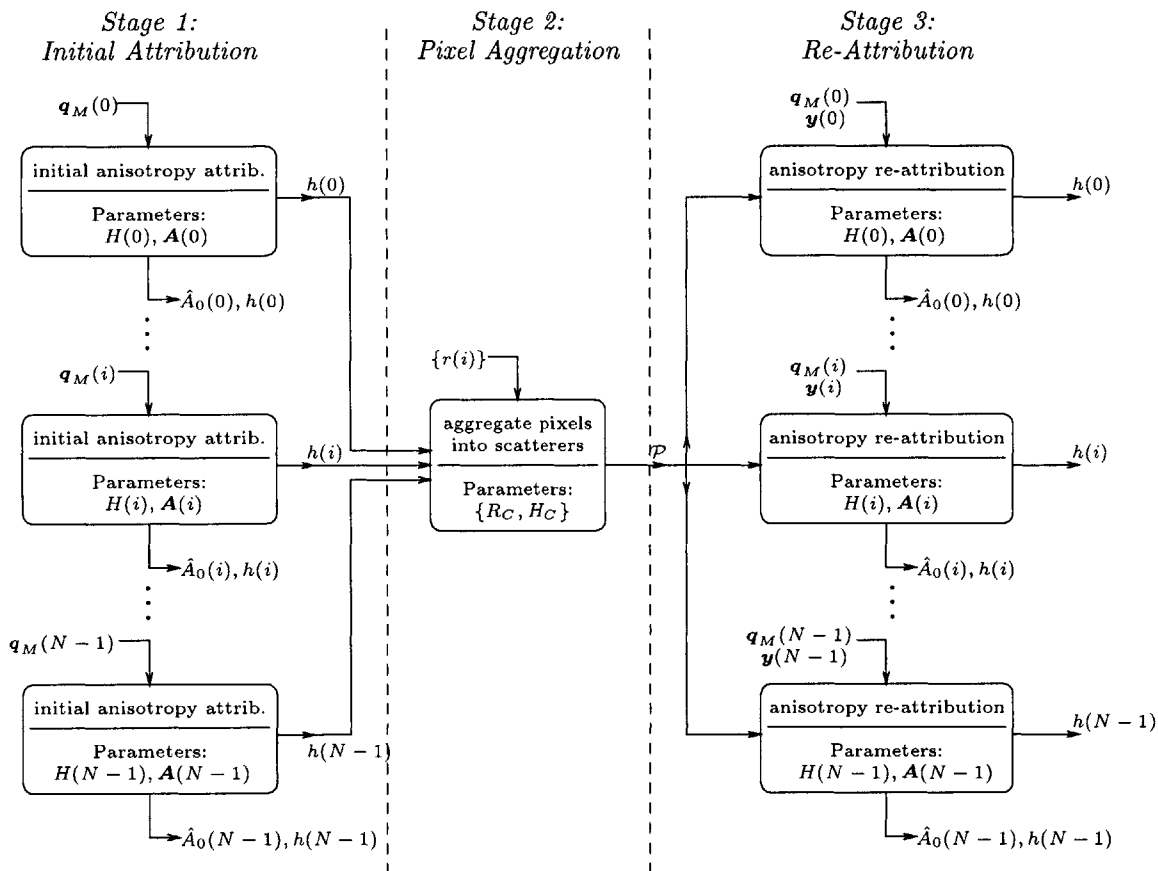
Figure 7.1. The iterative attribution algorithm used to classify anisotropy.

The first stage of the algorithm generates outputs on a per pixel basis. In particular, for each pixel i, a processing block in this stage generates an estimate[1] $h(i)$ of the true anisotropy $H(i)$ and an estimate $\hat{A}_0(i)$ of the reflectivity (where $A_0(i)$ denotes the true reflectivity for the focused pixel i). The method described in Chapter 5 is used to obtain these initial estimates from the sub-aperture measurements $q_M(i)$. As described in Chapter 5, estimation for this initial block is based on a pixel-by-pixel measurement model for $p(q_M(i)|H(i), A(i))$ where $A(i)$ is the vector of reflectivities of the focused pixel (denoted by $A_0(i)$) and its neighbors (denoted by $A_k(i)$, $k \in \{-K, \ldots, -1, 1, \ldots, K\}$).

---

[1] Because the estimated anisotropy is treated as a random measurement in a subsequent processing block, the estimate is denoted as a random variable $h(i)$ instead of the usual notation for an estimate, namely $\hat{H}(i)$.

The remaining two stages in Figure 7.1 correspond to an iterative refinement loop to be developed in the following two sections. The first of these two stages (i.e. the pixel aggregation step) takes the standard full-aperture estimates of log-magnitude reflectivity (denoted by $r(i)$) and current anisotropy estimates (denoted by $h(i)$) for every pixel in the image and groups pixels so that each group corresponds to an estimate of a single, possibly physically-extended, scatterer. The result is a partition $\mathcal{P}$ grouping the image pixels according to their underlying scatterers. The method we develop for this aggregation step involves a Bayesian formulation that captures the fact that for each cluster C of pixels corresponding to a single scatterer:

1. The log-magnitude full-aperture reflectivity estimates for each of the pixels in the grouping should be statistically consistent with each other because these reflectivities are all measurements of the single, physically-extended scatterer's log-magnitude reflectivity R_C.

2. The anisotropy classification for each of the pixels should be statistically consistent with each other as these attributions are all measurements of the scatterer anisotropy H_C.

3. The length of the grouping should be statistically consistent with the scatterer anisotropy H_C.

The Bayesian formulation we describe specifies a model $p(\{r(i), h(i)\}_{i \in C} | R_C, H_C)$ capturing the characteristics for any hypothesized cluster C. Using a GLR philosophy, this allows one to compute the most likely values of R_C and H_C for any hypothesized cluster and then to use these values in order to determine an approximation to the most likely partition $\mathcal{P}$ of clusters.

The last stage in Figure 7.1 (i.e. the scatterer re-attribution step) refines the estimates of anisotropy and reflectivity for each pixel i on a pixel-by-pixel basis. Each processing block in this stage, takes as inputs:

- the groupings $\mathcal{P}$ just estimated in the scatterer aggregation step,

- the sub-aperture measurements $q_M(i)$, and

- the vector $y(i)$ of full-aperture complex reflectivity measurements for the focused pixel (denoted by $y_0(i)$) and its neighbors at the finest resolution (denoted by $y_k(i)$, $k \in \{-K, \ldots, -1, 1, \ldots, K\}$).

These quantities are used to re-estimate scatterer anisotropy and reflectivity using a modified version of the algorithm in Chapter 5. The purpose of the modification is to avoid over-compensating for neighboring pixels that are part of the same scatterer grouping. In particular, in addition to using a measurement model for the observed sub-aperture reflectivities $q_M(i)$, we pose a measurement model for the observed finest-resolution complex reflectivities $y(i)$. The underlying parameters (which we wish to

estimate) in both of these models are the sub-aperture anisotropy $H(i)$ and vector of reflectivities $A(i)$ as was the case in the initial attribution. The new observation $y(i)$ replaces the zero-mean Gaussian prior on neighboring reflectivity estimates used in Chapter 5 and leads to more accurate reflectivity estimates of the reflectivities in $A(i)$. The segmentation is also used to aid the re-attribution by identifying on-scatterer pixels and allowing their effects to be preserved, while allowing for the compensation of effects from interfering off-scatterer pixels. The anisotropy attributions that are produced by this stage can be provided to the pixel aggregation block in the second stage of Figure 7.1, thus starting another aggregation/re-attribution iteration.

■ 7.2 Pixel Aggregation

The aggregation of image pixels into scatterers is the first phase of the two-stage iteration. This section presents a maximum likelihood approach based on the input measurements of pixel log-magnitude reflectivities[2] and pixel anisotropy attributions. The goal of this stage is to produce a partition $\mathcal{P}$ of the pixels into a set of groupings corresponding to scatterers. To achieve this goal, a measurement model is posed for the cluster measurements $\{r(i)\}_{i \in C}$ and $\{h(i)\}_{i \in C}$ in any hypothesized cluster C conditioned on the unknown parameters R_C and H_C. It is then shown that the posed model produces a likelihood function that corresponds to an additive cost function. Consequently, the problem of finding the maximum likelihood partition becomes an optimization problem on an additive cost function. An efficient, greedy algorithm is then presented.

■ 7.2.1 Problem Setup

We start by characterizing the relationship between (i) the anisotropy attributions $h(i)$ and reflectivities $r(i)$ that serve as the input data for the aggregation procedure and (ii) the unknown parameters H_C and R_C for any hypothesized cluster C. In particular, a Gaussian noise model is used for the reflectivities, i.e.

$$r(i) = R_C + u_i \tag{7.1}$$

where $u_i \sim \mathcal{N}(0, \rho_r^2)$ and ρ_r^2 is the variance on the reflectivities.

The pixel-wise anisotropy attributions $h(i)$ are modeled as identically distributed samples given the underlying scatterer attribution. However, posing a reliable model for the pixel anisotropy attributions is a difficult task. In particular, the confusion probabilities $p(h(i)|H_C)$ cannot be represented in closed form due to the manner in which the anisotropy attribution test partitions the observation space. Thus, we estimate these probabilities using Monte Carlo simulations. In particular, data are generated using

---

[2]For brevity, the term *reflectivity* will refer to the full-aperture log-magnitude reflectivity unless otherwise specified for the remainder of this section.

the sub-aperture scattering model as described in Eq. (5.22), i.e.

$$q_M = Ab(m, i) + \tilde{w}$$

where $\tilde{w} \sim \mathcal{N}(0, 2(|A|^2 \rho^2 + \sigma^2)\Lambda)$. We will assume high SNR scattering which leads to the approximation

$$\tilde{w} \sim \mathcal{N}(0, 2|A|^2 \rho^2 \Lambda).$$

This approximation allows us to consider only the normalized case of $A = 1$ since the noise standard deviation scales with the complex reflectivity A. The GLRT using the basic GLLR statistic is then used to attribute each MC trial. The confusion probabilities $p(h(i)|H_C)$ are then taken as the relative frequency of occurrence in the MC trials and stored as a look-up table.

In addition to the pixel reflectivities $r(i)$ and anisotropy attributions $h(i)$, there is an inherent scatterer size conveyed by the grouping C. In particular, the cardinality $\#(C)$ of C specifies the number of pixels (and hence the spatial extent) for the scatterer. Recall from Section 2.1.3 that the image and azimuthal domains are related by the Fourier transform. Thus, the cross-range extent of the scatterer is related to the degree of anisotropy. In particular, because spatial and azimuthal concentration are inversely related, the model for the observed scatterer size is defined in the azimuthal domain as

$$a(\#(C)) = d(H_C)(1 + v) \tag{7.2}$$

where $a(\#(C))$ is the 3dB width in the azimuthal domain of a flat plate of size $\#(C)$, $d(H_C)$ is the degree of anisotropy associated with H_C, and v is noise with a double-exponential distribution. That Eq. (7.2) is not a deterministic relationship between size and anisotropy is due to the rudimentary nature of our characterization of anisotropy (e.g. many non-canonic responses can lead to the same level of anisotropy in our model). The multiplicative form of the model (leading to an error term proportional to the degree of anisotropy) is consistent with the geometrically related lengths of the sub-apertures in our model. The use of the double exponential model for v is for computational convenience. For reference, the model in Eq. (7.2) produces the likelihood

$$p\left(\#(C)|d(H_C)\right) = \frac{1}{2\rho_l} \exp\left[\frac{-|a(\#(C)) - d(H_C)|}{\rho_l \, d(H_C)}\right]. \tag{7.3}$$

■ 7.2.2 Aggregation Likelihood

In formulating the likelihood of an entire aggregation of pixels, we start by considering a single cluster C. The likelihood of the data in this grouping is

$$p\left(\{r(i)\}_{i \in C}, \{h(i)\}_{i \in C} \big| R_C, H_C\right) \tag{7.4}$$

The models posed in the previous sub-section are used to decouple and simplify this likelihood expression. First, note that although the *individual* pixel reflectivities $r(i)$

are characterized by R_C and the *individual* pixel anisotropy attributions $h(i)$ are characterized by H_C, the likelihood in Eq. (7.4) cannot immediately be written as a product of the terms $p(r(i)|R_C)$ and $p(h(i)|H_C)$. This cannot be done because the cardinality of the sets $\{r(i)\}_{i \in C}$ and $\{h(i)\}_{i \in C}$ are related to scatterer size. Consequently, consideration of the grouping introduces an additional dependency on the underlying anisotropy. However, by rewriting Eq. (7.4) to explicitly include the cardinality of C, we can obtain the desired separation of terms. In particular, we can write

$$
p\left(\{r(i)\}_{i \in C}, \{h(i)\}_{i \in C} \middle| R_C, H_C\right)
$$

$$
= p\left(\#(C), \{r(i)\}_{i \in C}, \{h(i)\}_{i \in C} \middle| R_C, H_C\right) \tag{7.5a}
$$

$$
= p\left(\#(C) \middle| R_C, H_C\right) \, p\left(\{r(i)\}_{i \in C}, \{h(i)\}_{i \in C} \middle| \#(C), R_C, H_C\right) \tag{7.5b}
$$

$$
= p\left(\#(C) \middle| R_C, H_C\right) \prod_{i \in C} p(r(i)|R_C) \prod_{i \in C} p(h(i)|H_C) \tag{7.5c}
$$

$$
= p\left(\#(C) \middle| d(H_C)\right) \prod_{i \in C} p(r(i)|R_C) \prod_{i \in C} p(h(i)|H_C). \tag{7.5d}
$$

Eq. (7.5a) follows from the fact that $\#(C)$ is redundant in that can be determined from $\{r(i)\}_{i \in C}$ and $\{h(i)\}_{i \in C}$. Eq. (7.5b) is an application of Bayes rule. Eq. (7.5c) follows from the fact that given R_C, H_C, *and the size of the set* C, we can model the $r(i)$ and $h(i)$ as i.i.d. Eq. (7.5d) follows from the degree of anisotropy characterizing the spatial extent of the scatterer.

With the likelihood specification from Eq. (7.5) in hand, we now proceed to develop the aggregation problem. We start with the simple case in which there are two pixel groupings X and Y that neighbor each other, i.e. $X \cup Y$ is a connected set. For each grouping (X or Y), it is known that all the pixels in that group correspond to the same scatterer, i.e. X and Y each span part of a scatterer but not necessarily the entire scatterer. Based on the data $r(i)$ and $h(i)$, we wish to test whether $X \cup Y$ corresponds to a single scatterer (hypothesis G_1) or separate scatterers (hypothesis G_2). A maximum likelihood test is used for this task.

Because the likelihood in Eq. (7.5) depends on the unknown parameters scatterer reflectivity R_C and anisotropy H_C, these parameters are estimated and used to produce a generalized likelihood. In particular, Eq. (7.1) is used to generate the ML estimate of R_C, which is the sample mean of the reflectivities in group C. This estimate will be denoted by $\hat{R}_C$. The ML estimate of H_C is given by

$$
\hat{H}_C = \arg\max_{(n,j)} \left\{ \prod_{i \in C} p\left(h(i)|H = (n,j)\right) \right\}. \tag{7.6}
$$

The probability models for the pixel reflectivity, pixel anisotropy, and cluster size defined in Section 7.2.1 can be inserted into Eq. (7.5) to determine the generalized likelihood of X and Y corresponding to two separate scatterers (G_2) or $X \cup Y$ corresponding to a single scatterer (G_1). In particular, the generalized likelihood of a group C (with

C representing X, Y, or $X \cup Y$) is

$$\ell_C = p\left(\#(C)\Big|\hat{H}_C\right) \prod_{i \in C} p\left(r_i\Big|\hat{R}_C\right) \prod_{i \in C} p\left(h(i)\Big|\hat{H}_C\right)$$

$$= \frac{1}{2\rho_l} \exp\left[-\frac{|a(\#(C)) - d(\hat{H}_C)|}{\rho_l \, d(\hat{H}_C)}\right] (2\pi\rho_r^2)^{-\#(C)/2} \prod_{i \in C} \exp\left[\frac{(r(i) - \hat{R}_C)^2}{2\rho_r^2}\right] \prod_{i \in C} p\left(h(i)\Big|\hat{H}_C\right).$$

Thus, the GLLR between the two hypotheses G_2 and G_1 can be written as

$$\gamma = \log\left(\frac{\ell_X \ell_Y}{\ell_{X \cup Y}}\right) = f(X \cup Y) - f(X) - f(Y) \tag{7.7}$$

where $f(C)$ is the negative log-likelihood for the data in set C. Note that the density normalization factor for the pixel reflectivities cancel in the likelihood ratio and can thus be omitted from the cost function, i.e.

$$f(C) = \log(2\rho_l) + \frac{|a(\#(C)) - d(\hat{H}_C)|}{\rho_l d(\hat{H}_C)} + \frac{1}{2\rho_r^2} \sum_{i \in C}(r(i) - \hat{R}_C)^2 - \sum_{i \in C} \log\left(p\left(h(i)|\hat{H}_C\right)\right). \tag{7.8}$$

Consequently, the decision on whether to merge the groupings X and Y is determined by whether the costs of X and Y are lower individually (i.e. $f(X) + f(Y) < f(X \cup Y)$) or the cost is lower by combining the two (i.e. $f(X) + f(Y) > f(X \cup Y)$).

The additive nature of the GLLR in Eq. (7.7) allows the aggregation rule easily to be extended to consider an entire partition. In particular, assume that we are presented with two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ of all the cross-range pixels and that $\mathcal{P}_2$ is a refinement of $\mathcal{P}_1$, i.e. every $C \in \mathcal{P}_1$ can be partitioned by elements of $\mathcal{P}_2$. The task is then to determine which partition has the higher generalized likelihood. The GLLR of these two partitions can be written as

$$\log\left(\frac{\prod_{C \in \mathcal{P}_2} \ell_C}{\prod_{C \in \mathcal{P}_1} \ell_C}\right) = F(\mathcal{P}_1) - F(\mathcal{P}_2) \tag{7.9}$$

where

$$F(\mathcal{P}) = \sum_{C \in \mathcal{P}} f(C). \tag{7.10}$$

Thus, the ML aggregation problem amounts to finding the partition with the minimum cost given by Eqs. (7.8) and (7.10)

■ 7.2.3 Greedy Aggregation Algorithm

It is computationally prohibitive to evaluate the cost in Eq. (7.10) for every valid partition. Motivated by the additive structure of the cost in Eq. (7.10), we propose the

following greedy algorithm that, at each iteration, merges the two groupings which would produce the greatest increase in the likelihood. This merging process is repeatedly performed until further mergers no longer increase the likelihood.

Step 1: Initialize the partition $\mathcal{P}$ as the set of all singletons, i.e. each grouping contains only one pixel.

Step 2: Consider the set $\mathcal{M}$ of merger candidates obtained from merging two *neighboring* groupings $X, Y \in \mathcal{P}$. Compute the sub-cost reduction $[f(X) + f(Y) - f(X \cup Y)]$ produced by each merger.

Step 3: Find the merger in $\mathcal{M}$ which provides the largest sub-cost reduction. Denote this merger as $M^* = X^* \cup Y^*$. If the sub-cost reduction is negative, then we are at a local minimum, i.e. there is no pairwise merger that will increase the likelihood of the resulting partition, so the algorithm terminates. If the maximum sub-cost reduction is positive, then the merger is accepted and $\mathcal{P}$ is updated to reflect the merger, i.e. M^* replaces X^* and Y^* in $\mathcal{P}$.

Step 4: Update $\mathcal{M}$ to account for the new merger. That is, if $[W, X, Y, Z]$ forms a connected set, then the update is to remove $M^*, W \cup X$, and $Y \cup Z$ from $\mathcal{M}$ and insert the new merger candidates $W \cup M^*$ and $M^* \cup Z$. Compute sub-cost reduction of the two new candidates, i.e. $[f(W) + f(M^*) - f(W \cup M^*)]$ and $[f(M^*) + f(Z) - f(M^* \cup Z)]$.

Step 5: Goto step 3 to find the best merger.

■ 7.2.4 Aggregation Results

To demonstrate the performance of the proposed aggregation algorithm, we use the synthesized data from Chapter 5 which is displayed again in Figure 7.2. Figure 7.2(a) shows the standard log-magnitude reflectivity image. Figure 7.2(b) shows this same image but includes annotations to describe the scatterers. The geometric symbols in this figure are used to denote the underlying true anisotropy. In particular,

- a circle represents a flat plate scatterer whose cross-range extent is half of a resolution cell and gives a full-aperture response,

- a square represents a flat plate scatterer whose cross-range extent is two resolution cells and gives a half-aperture response,

- a triangle represents a flat plate scatterer whose cross-range extent is four resolution cells and gives a quarter-aperture response,

Furthermore, each scatterer is given an identifier to facilitate the analysis of the segmentation in this section. In particular,

- the label A_i ($i \in \{1, \ldots, 6\}$) refers to an isotropic scatterer with a 30dB PSNR,

- the label B_1 refers to the half-aperture scatterer with a 30dB PSNR,

- the label C_i ($i \in \{1, 2\}$) refers to a quarter-aperture scatterer with a 30dB PSNR, and

- the label D_i ($i \in \{1, 2\}$) refers to a half-aperture scatterer with a 20dB PSNR.

(a)   (b)
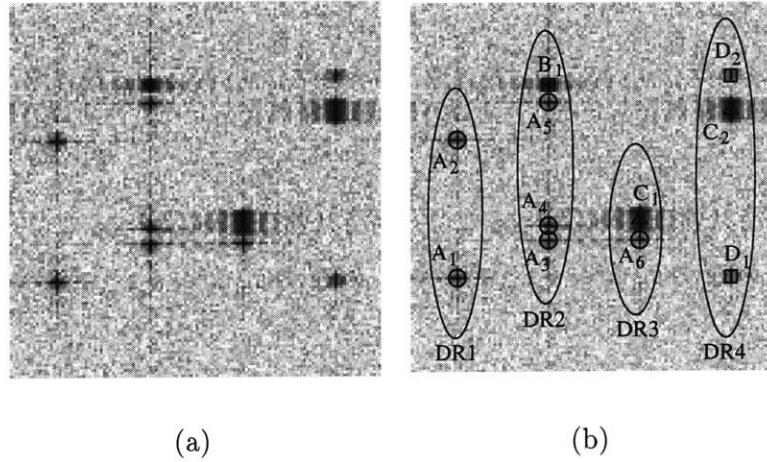
Figure 7.2. Standard log-magnitude image of the synthesized SAR data used in Chapter 5 (a) without annotations and (b) with annotations describing the individual scatterers. A circle denotes an isotropic scatterer; a square denotes a half-aperture scatterer; and a triangle denotes a quarter-aperture scatterer. The large ellipses denote which scatterers are used to define a particular down-range value (i.e. DR1, DR2, DR3, and DR4).

The aggregation procedure operates on all the cross-range pixels for a given down-range value. Thus, each down-range is examined individually. For the analysis presented in this section, attention is concentrated on the down-range (referred to as DR3) containing the scatterers A_6 and C_1 and the down-range (referred to as DR4) containing the scatterers D_1, C_2, and D_2. Later in Section 7.4, the down-range (referred to as DR1) containing scatterers A_1 and A_2 and the down-range (referred to as DR2) containing scatterers A_3, A_4, A_5, and B_1 will also be examined.

The results of the anisotropy attribution developed in Chapter 5 are used for the attribution inputs to the segmentation. In particular, the attributions used are the ones displayed in Figure 5.8(d) from Chapter 5. The segmentations resulting from the approach developed in this section are shown in Figures 7.3 and 7.4 for DR3 and DR4, respectively. In each display, figure (a) plots the standard log-magnitude reflectivities to illustrate the location and spatial extent of the scatterers, and figure (b) plots the segmentation. In order to clearly demarcate the many groupings involved, the segmentation is conveyed in the following manner. At every boundary between two groups, the value of the plot changes sign, i.e. from positive-to-negative or negative-to-positive. Furthermore, the magnitude of the plotted value is set to the number of pixels in the

grouping to help convey the size. A dashed vertical line is used to denote the true locations of the scattering centers in both figures. For reference, these dashed lines are labeled with the identifier (e.g. A_6) for each scatterer in Figures 7.3(a) and 7.4(a).
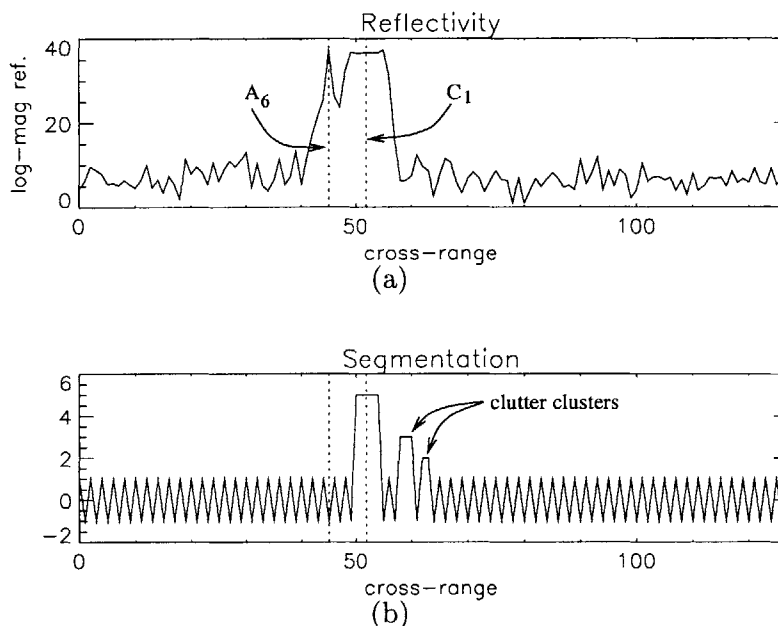
Figure 7.3. (a) Reflectivity of all cross-range pixels at down-range DR3 containing the isotropic scatterer A_6 and quarter-aperture scatterer C_1. (b) Segmentation for pixels at down-range DR3.

The results of the segmentation using the initial anisotropy attribution are quite accurate. In particular, both scatterers in Figure 7.3 are reasonably characterized. Note that physical extent of scatterers A_6 and C_1 are half a resolution cell and four resolution cells, respectively. Consequently, given the image oversampling ratio of 1.25 pixels per resolution cell, characterizing A_6 as one pixel and C_1 as five pixels is accurate. Note that the segmentation in Figure 7.3(b) also reports two other non-singleton groupings (labeled as "clutter clusters") in clutter. These groupings are associated with erroneous anisotropy attributions due to interference from the larger scatterer C_1. However, these large clutter groupings should not significantly effect the anisotropy re-attribution process described in the next section. In particular, the pixels in these clutter groupings are marked as belonging to a different cluster than the true scatterers, and thus, the re-attribution process should compensate for the interference (as we see in Section 7.4).

The results of the segmentation in Figure 7.4 for down-range DR4 are also fairly accurate. Recall from Section 5.1 that the spatial extent of the three scatterers in this image are two resolution cells, four resolution cells, and two resolution cells, respectively, in cross-range. Accounting for the image oversampling, the characterizations of three pixels and five pixels for the first two scatterers are accurate and are consistent with
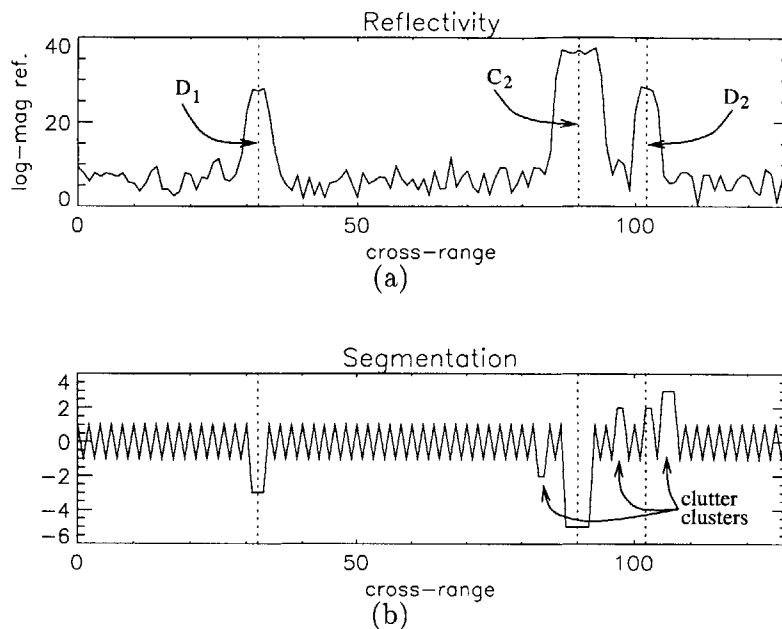
Figure 7.4. (a) Reflectivity of all cross-range pixels at down-range DR4 containing the half-aperture scatterers D_1 and D_2 and the quarter-aperture scatterer C_2. (b) Segmentation for pixels at down-range DR4.

the spatial extent implied by the plotted reflectivity values. The third scatterer is characterized as two pixels. The plotted reflectivity values indicate that the grouping for this scatterer should be extended to the left by one more pixel. The grouping is still acceptable, however, considering that the underlying scatterer spans 2.5 pixels.

■ 7.3 Anisotropy Re-attribution

The aggregation of pixels produced by the algorithm in the previous section allows us to account for the effects of neighboring scatterers when re-attributing the anisotropy of each pixel. In particular, as illustrated in Figure 7.1, the re-attribution processing stage takes as inputs: the aggregation $\mathcal{P}$ of pixels, the sub-aperture measurements $q_M(i)$, and finest-resolution complex reflectivity[3] measurements $y(i)$, where $y(i)$ is the same dimension as $A(i)$ and the elements of $y(i)$ correspond to the same pixel locations as the elements of $A(i)$. The idea of the attribution algorithm follows the same principles as the attribution developed in Chapter 5, except that the aggregation $\mathcal{P}$ and reflectivities in $y(i)$ are used to better estimate the unknown parameter of reflectivities A (and thus H). This improvement is achieved by (i) incorporating a model for the data $y(i)$ and

---

[3]Note that in this section we are using the complex reflectivities in contrast to the log-magnitude reflectivities used in Section 7.2.

(ii) enhancing the model for the data $q_M(i)$. The result of these improvements is an estimation procedure for $A(i)$ that

1. regularizes the off-scatterer components of $A(i)$ towards the corresponding measured values in $y(i)$, thus improving the modeling of neighboring scatterers, and

2. explicitly excludes neighboring on-scatterer pixels (i.e. neighbors of the focused pixel included in the same single scatterer grouping) from the re-estimation/re-attribution process, thereby avoiding over-compensation for the influence of this single scatterer on the observed sub-aperture measurement vector.

Note that the analysis presented in this section is performed for each focused pixel i. Thus, we will describe the methodology for a single focused pixel and drop the notational dependence on i.

■ 7.3.1 Re-attribution Algorithm

Examination of Measurement Model from Chapter 5

Prior to posing the new measurement models for q_M and y, we first re-examine the measurement model used for the attribution in Chapter 5, namely that given in Eq. (5.26) of Section 5.2 which models neighboring scatterers at regularly spaced locations corresponding to image pixels, i.e.

$$q_M = \begin{bmatrix} b^{-K}(m_{-K}, i_{-K}) & | & \cdots & | & b^K(m_K, i_K) \end{bmatrix} \begin{bmatrix} A_{-K} \\ \vdots \\ A_K \end{bmatrix} + \varepsilon$$

$$= BA + \varepsilon. \tag{7.11}$$

The matrix B is formed from the basis functions $b^k(m_k, i_k)$ as indicated in Eq. (7.11). Each $b^k(m_k, i_k)$ models the interference from a neighboring scatterer whose anisotropy is given by the sub-aperture scattering hypothesis H_{m_k, i_k}. For the approach taken in Chapter 5, H_{m_k, i_k} is taken as the full-aperture hypothesis for all neighboring pixels, i.e. $k \neq 0$. Thus, only the hypothesis H_{m_0, i_0} effects the contents of the matrix B and is simply denoted as H.

The estimate used for A in Section 5.2 can be interpreted as the MAP estimate of A assuming q_M is the measurement and A has the Gaussian prior

$$A \sim \mathcal{N}\left(0, \frac{2\sigma^2 L_M}{\gamma} R_\delta^{-1}\right) \tag{7.12}$$

where A and ε are independent, L_M is the length of the smallest sub-apertures,

$$R_\delta = \text{diag}[1, \ldots, 1, \delta, 1, \ldots, 1] = I - (1 - \delta)e_K e_K^H,$$

and δ is an infinitesimally small value. The estimate of A under anisotropy hypothesis H can be viewed as the MAP estimate

$$
\begin{aligned}
\hat{A} &= \arg\max_{A} \{p(A|q_M)\} \\
&= \arg\max_{A} \left\{ \frac{p(q_M|A)\, p(A)}{p(q_M)} \right\} \\
&= \arg\max_{A} \{p(q_M|A)\, p(A)\} \\
&= \arg\min_{A} \left\{ \|q_M - BA\|^2_{\Lambda^{-1}} + \gamma A^H R A \right\}
\end{aligned}
\tag{7.13}
$$

where $R = \mathrm{diag}[1, \ldots, 1, 0, 1, \ldots, 1] = I - e_K e_K^H$, and the anisotropy hypothesis enters into the equation through B. The matrix R effectively drives all the reflectivities (except the focused pixel's reflectivity) towards zero, while the focused pixel's reflectivity A_0 is left unregularized.

Enhanced Measurement Model

Except for the case of an isolated scatterer in clutter, Eq. (7.12) is a primitive model for neighboring scatterers. In fact, the inadequacies of this model are responsible for many of the misclassifications from the GLRT based on the MSM statistic when dealing with interfering scatterers. To address this problem, the model for A is improved by taking advantage of the segmentation $\mathcal{P}$ and the full-resolution reflectivity estimates in y.

In order to clearly motivate the modeling choices that are made in this section, we first present a simple measurement model and develop this model into the one that we use. To start, we define a reduced order reflectivity parameter $\tilde{A}$ which is a sub-vector of A. In particular, $\tilde{A}$ is formed by taking A and removing those elements that correspond to neighboring, on-scatterer pixels (i.e. pixels in the same grouping as the focused pixel but not the focused pixel itself). This procedure is illustrated in the example in Figure 7.5 where $[A_{-1}, \ldots, A_2]$ constitute a single grouping in $\mathcal{P}$. $\tilde{A}$ is then obtained by removing A_{-1}, A_1 and A_2 from the vector. Using this reduced order parameter reflects the fact that there is a single, physically-distributed scatterer spanning those pixels, and thus prevents the over-compensation of the focused pixel by neighboring, on-scatterer pixels. To account for the reduced model order, B is appropriately adjusted by removing those columns corresponding to the discarded components of A. Denoting this modified matrix as $\tilde{B}$, the sub-aperture measurement model becomes

$$
q_M = \tilde{B}\tilde{A} + \varepsilon.
\tag{7.14}
$$

The use of the reduced vector $\tilde{A}$ addresses the problem of over-compensation from neighboring on-scatterer pixels but not does address the interference from off-scatterer pixels. To deal with the interference, the additional measurement y is used to estimate $\tilde{A}$ instead of using the prior in Eq. (7.12). A natural model for these finest resolution

$$A = \begin{bmatrix} A_{-K} \\ \vdots \\ A_{-2} \\ \hline A_{-1} \\ A_0 \\ A_1 \\ A_2 \\ \hline A_3 \\ \vdots \\ A_K \end{bmatrix} \left.\vphantom{\begin{matrix} A_{-1} \\ A_0 \\ A_1 \\ A_2 \end{matrix}}\right\} \begin{matrix} \text{single} \\ \text{grouping} \\ \text{in } \mathcal{P} \end{matrix} \qquad \widetilde{A} = \begin{bmatrix} A_{-K} \\ \vdots \\ A_{-2} \\ A_0 \\ A_3 \\ \vdots \\ A_K \end{bmatrix}$$
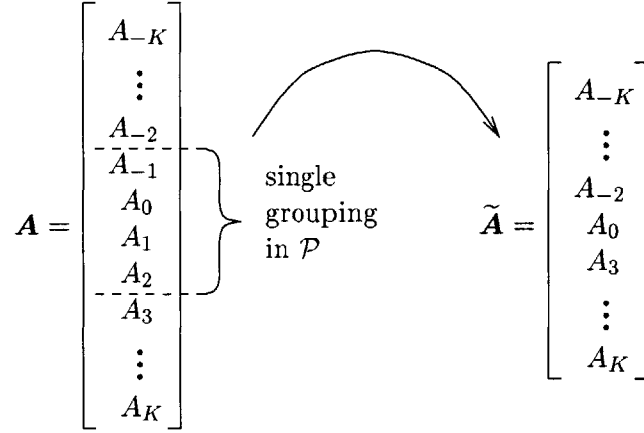
Figure 7.5. Illustration of reducing the model order of A by removing neighboring, on-scatterer components to produce $\widetilde{A}$.

reflectivity measurements is

$$y = P\widetilde{A} + v \tag{7.15}$$

where $v \sim \mathcal{N}(0, \frac{2\sigma^2 L_M}{\gamma} I)$ is noise that is independent of ε, and P describes how a scatterer in $\widetilde{A}$ is observed in y. In particular, the jth column of P represents the impulse response associated with the scatterer corresponding to the jth element of $\widetilde{A}$. To be consistent with $\widetilde{B}$ in the sub-aperture scattering model of Eq. (7.14), all pixels are associated with the full-aperture scattering hypothesis except the focused pixel which is assumed to obey anisotropy hypothesis H. The impulse response in the column of P corresponding to the focused pixel is the sinc response associated with anisotropy hypothesis H. Furthermore, the response for this focused pixel is shifted in order to align the mainlobe of the response with the scattering center (in contrast to the focused pixel) where the scattering center is taken as the centroid of the cluster in $\mathcal{P}$ containing the focused pixel. For example, the estimated scattering center of the grouping shown in Figure 7.5 is taken as the location halfway between the pixels corresponding to A_0 and A_1. The impulse response in any other column of P is the narrow sinc function corresponding to the over-sampled[4] response from a full-aperture scatterer. Note that, in contrast to $\widetilde{A}$, the dimensionality of the vector y is not reduced since the response associated with A_0 should model the reflectivities over *all* the on-scatterer pixels. From the models in Eqs. (7.14) and (7.15), one could[5] estimate $\widetilde{A}$ as

$$\widehat{\widetilde{A}} = \arg\max_{\widetilde{A}} \left\{ p(q_M, y | \widetilde{A}) \right\}$$

---

[4]In practice, SAR imagery is usually over-sampled by zero-padding the aperture.

[5]For reasons explained shortly, we do not actually use the model in Eq. (7.15) and the test defined by Eqs. (7.16) and (7.17).

$$= \arg\max_{\tilde{A}} \left\{ p(\boldsymbol{q}_M | \tilde{A}) p(\boldsymbol{y} | \tilde{A}) \right\}$$

$$= \arg\min_{\tilde{A}} \left\{ \|\boldsymbol{q}_M - \tilde{B}\tilde{A}\|^2_{\Lambda^{-1}} + \gamma \|\boldsymbol{y} - P\tilde{A}\|^2 \right\}$$

$$= \left[\tilde{B}^H \Lambda^{-1} \tilde{B} + \gamma P^H P \right]^{-1} \left(\tilde{B}^H \Lambda^{-1} \boldsymbol{q}_M + \gamma P^H \boldsymbol{y} \right). \tag{7.16}$$

This estimate of $\tilde{A}$ could then be used to compute the error-norm

$$\|\boldsymbol{\varepsilon}\|^2_{\Lambda^{-1}} = \|\boldsymbol{q}_M - \tilde{B}\hat{\tilde{A}}\|^2_{\Lambda^{-1}} \tag{7.17}$$

to determine a generalized likelihood of each sub-aperture hypothesis. These generalized likelihoods can then be used to perform a GLRT to characterize anisotropy.

The model in Eq. (7.15) accurately models the reflectivities in $\tilde{A}$ under the correct hypothesis, but not under the incorrect hypothesis. Such behavior is not surprising and is not a fundamental drawback. However, this behavior *does* severely impact the performance of the anisotropy test if a telescopic hypothesis testing procedure is used. In particular, for a highly anisotropic scatterer, decision errors are made in the early stages of the telescopic testing sequence that prevent the correct anisotropy hypothesis from being tested. This problem is best illustrated with an example. Consider the situation illustrated in Figure 7.6 where a large, distributed scatterer is being modeled with an insufficiently anisotropic response (i.e. one using too large of a sub-aperture). The location of the focused pixel in this figure is labeled with a "×" on the horizontal axis. In order to match the finest resolution reflectivity for that location with an ill-fitting sinc function, the estimate $\hat{A}_0$ is regularized towards a large negative value since we are focused on a negative sidelobe. Using such a poorly regularized value of $\tilde{A}$ in Eq. (7.17) produces large erratic error values for $\boldsymbol{\varepsilon}$ under incorrect sub-aperture hypotheses. Thus, in the early stages of the telescopic test (when none of the examined hypotheses are correct) the telescopic testing procedure frequently makes an "incorrect" decision in the sense that the decision either terminates the testing (having decided that the correct sub-aperture has been reached) or leads the telescopic testing procedure down an incorrect path on the pyramid where the correct hypothesis is never evaluated.

We mention two methods to address this problem. One approach would be to omit the telescopic testing procedure and evaluate all the sub-aperture hypotheses since the problem only arises when the incorrect sub-aperture hypothesis is evaluated. This course is not taken in this thesis due to the computational savings the telescopic testing procedure provides. Instead, the approach taken in this section is along the same lines as the approach in Section 5.2 where the reflectivity for the scatterer under examination is left unregularized. The approach taken here uses the measurement $\boldsymbol{y}$ only to estimate the reflectivity of neighboring scatterers and leaves the estimation of the focused pixel's reflectivity to be determined by the sub-aperture measurements $\boldsymbol{q}_M$. In particular, the dimensionality of the measurement $\boldsymbol{y}$ is reduced by discarding *all* measurements
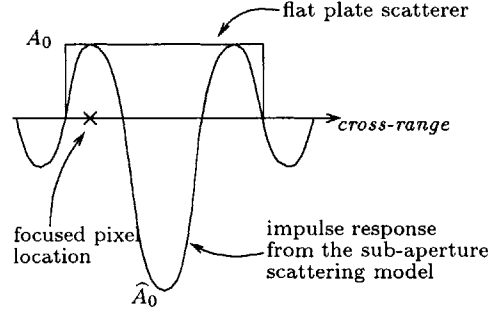
Figure 7.6. Illustration of why the reflectivity is incorrectly estimated in the early stages of the telescopic test on a spatially distributed scatterer. The sinc function is centered at the scattering center, but the sinc associated with the hypothesis being tested is too narrow. The point we are trying to fit (labeled with a cross-hatch) corresponds to a sidelobe of the sinc, but the observation model in Eq. (7.15) assumes that the sinc mainlobe spans the entire scatterer. Thus, the estimated reflectivity is in gross error and severely distorts the anisotropy characterization.

(including y_0) of on-scatterer pixels. This reduced vector is denoted as y_{off} and is modeled as

$$y_{off} = P_{off}\widetilde{A} + \eta \tag{7.18}$$

where $\eta \sim \mathcal{N}\left(0, \frac{2\sigma^2 L_M}{\gamma}I\right)$ and P_{off} is the matrix obtained from P by removing those rows associated with the discarded elements of y. Using the new model in Eq. (7.18), the value of $\widetilde{A}$ is estimated as

$$\widehat{\widetilde{A}} = \arg\max_{\widetilde{A}}\left\{p(q_M, y_{off}|\widetilde{A})\right\}$$

$$= \arg\max_{\widetilde{A}}\left\{p(q_M|\widetilde{A})p(y_{off}|\widetilde{A})\right\}$$

$$= \arg\min_{\widetilde{A}}\left\{\|q_M - \widetilde{B}\widetilde{A}\|^2_{\Lambda^{-1}} + \gamma\|y_{off} - P_{off}\widetilde{A}\|^2\right\}$$

$$= \left[\widetilde{B}^H\Lambda^{-1}\widetilde{B} + \gamma P^H_{off}P_{off}\right]^{-1}\left(\widetilde{B}^H\Lambda^{-1}q_M + \gamma P^H_{off}y_{off}\right). \tag{7.19}$$

This estimate of $\widetilde{A}$ is then used to compute the error-norm

$$\|\varepsilon\|^2_{\Lambda^{-1}} = \|q_M - \widetilde{B}\widehat{\widetilde{A}}\|^2_{\Lambda^{-1}} \tag{7.20}$$

that determines the generalized likelihood of each sub-aperture hypothesis and is used to perform a GLRT to characterize anisotropy. This test using Eqs. (7.19) and (7.20) is the one we use for the results in Section 7.4.

■ 7.3.2 Comments about the Algorithm

Prior to demonstrating the performance of the aggregation/re-attribution process, this section makes a few comments about the re-attribution step.

Modeling Neighbors as Isotropic

The first point we address is whether modeling neighboring scatterers as isotropic captures the interference from anisotropic interfering scatterers. In particular, a single, physically-distributed scatterer exhibits a consistent phase on the reflectivity over the on-scatterer pixels. Although the approach presented in this section does not explicitly model this phase consistency, the estimate in Eq. (7.19) is still able to capture this phase consistency to a large degree. In particular, this phase consistency is captured by the measured complex reflectivities, albeit corrupted by measurement noise. Thus, for scatterers with a high SNR, there should be a strong (although not perfect) consistency in the phase on the estimates of $\widetilde{A}$ corresponding to a single scatterer. The success of this approach in capturing the behavior of interfering scatterers is demonstrated in Section 7.4.

Omission of the Previous Iteration's Anisotropy Estimates

The re-attribution statistic obtained from Eqs. (7.19) and (7.20) uses only the aggregated groupings, the full-aperture reflectivity measurements, and the sub-aperture reflectivity measurements to estimate the anisotropy of a given pixel. Each neighboring scatterer is modeled as a series of isotropic scatterers instead of a single anisotropic scatterer because this pixel-wise approach is robust to attribution errors from earlier iterations. If a previous iteration's anisotropy attribution and sub-aperture reflectivity estimate are used to model neighboring scatterers, then initial attribution errors could be catastrophic. In particular, consider the false anisotropy attributions in clutter around a strong scatterer. We will call these misclassifications *phantom attributions* as the anisotropy attribution is due to the interference from a neighboring scatterer and not the contents of the focused pixel's resolution cell. Many of the sub-aperture reflectivity estimates for these phantom classifications are large because the associated coarse resolution cell contains part of the interfering scatterer. If these erroneous attributions and reflectivity estimates are used in a re-attribution step, then the algorithm would compensate for a non-existent scatterer which would almost surely lead to other errors. The re-attribution approach proposed in this chapter, however, is robust to such errors since the finest resolution reflectivities are used and the response from neighboring scatterers are modeled in a pixel-wise fashion. Because the finest resolution reflectivities are used, one does not suffer from the coarser resolution associated with small sub-apertures. Furthermore, because the scatterers are modeled pixel-by-pixel, the re-attribution step need only be able to distinguish on-scatterer from off-scatterer pixels, i.e. it is not necessary to determine the spatial extent of each interfering scatterer.

Skipping the Re-attribution Step

We also point out that the re-attribution for a pixel needs to be performed only if the grouping containing that pixel has changed from the previous aggregation/re-attribution iteration. In particular, if the set of on-scatterer pixels for a given location

is unchanged from the previous iteration, then all the quantities in Eqs. (7.19) and (7.20) are unchanged and the re-attribution can be skipped. We have observed that in practice, the vast majority of pixels need to be re-attributed only once, i.e. after the initial aggregation step. This is especially the case for clutter pixels which are usually grouped as singletons.

Independence of ε and η

The observation models in Eqs. (7.14) and (7.18) assume that the noises ε and v on the sub-aperture and finest-resolution reflectivity measurements are independent. This assumption is not precisely true since the sub-aperture measurements q_M and reflectivities y_{off} are obtained from the same set of azimuthal data. However, we believe that any existing dependencies are weak due to the differences in the way the measurements are obtained. In particular,

- For the sub-aperture measurements, the azimuthal data is demodulated to focus on a particular pixel location. Then, the components of q_M are obtained by integrating this *single azimuthal response* over each individual *sub-aperture*.

- For the finest-resolution reflectivities, an azimuthal response is calculated for each pixel location where the demodulation of the response varies according to the focused location. Each of these *multiple azimuthal responses* are then integrated over the *entire aperture* to produce the set of reflectivities y_{off}.

Thus, the dependence (conditioned on A) between the two measurements should be small because of the different modulations used prior to the integration and the different integration intervals used.

■ 7.4 Experimental Results

This section demonstrates the results of the iterative anisotropy attribution. The algorithm is applied to the synthesized data used in Chapter 5 and the MSTAR data imaged at cardinal angles. The results on synthesized data are presented first to present a detailed analysis of the workings of the algorithm. The results on MSTAR data are then presented to demonstrate that the algorithm performs well in real-world settings.

For all the results in this section, the initial anisotropy attribution in stage 1 of Figure 7.1 uses the GLRT based on the MSM statistic of Section 5.2 with the telescopic testing procedure and the same testing parameters as in Chapter 5. The segmentation parameters are taken as $\rho_l = 0.5$ and $\rho_r = 0.5$. The re-attribution parameters are taken to be the same as in the initial attribution except for the regularization parameter which is now taken to be $\gamma = 0.25$. The value of γ is reduced from 0.5 for the initial anisotropy attribution to 0.25 for the re-attribution in order to produce a consistent model for the sub-aperture and finest resolution reflectivity measurements. In particular, recall that the error η for the finest resolution reflectivities in Eq. (7.18) specifies the noise variance

as $\frac{2\sigma^2 L_M}{\gamma} Q_\delta^{-1}$ in order to produce the objective function in Eq. (7.19). Furthermore, recall from Eq. (5.5) that the noise variance on the full-aperture reflectivity estimate is $2\sigma^2$. Thus, in order for these two models to be consistent, it must follow that

$$\frac{2\sigma^2 L_M}{\gamma} = 2\sigma^2$$

$$\gamma = \frac{1}{L_M} = 0.25$$

for the three-level, half-overlapping, half-aperture pyramid. The larger value of $\gamma = 0.5$ is necessary for the initial anisotropy attribution in order to suppress the contribution from other on-scatterer pixels. The re-attribution approach, however, already explicitly removes these redundant parameters from the model, and thus, the regularization parameter need not be set artificially high.

Figures 7.7 through 7.11 display the results of the iterative anisotropy attribution for the synthetic image data used in Chapter 5 which is repeated in Figure 7.2. Figure 7.7 displays the anisotropy attribution and sub-aperture reflectivity images at each iteration. Figures 7.7(a) and (b) are the result of the initial attribution using the GLRT based on the MSM statistic of Section 5.2. Figures 7.7(c) through (f) are the results of the first and second re-attribution iterations. Subsequent aggregation/re-attribution iterations do not result in any changes from the attributions shown in figures (e) and (f). The improvement of the anisotropy re-attribution over the initial attribution is clear. The most obvious result is that all of the phantom attributions in clutter have been removed[6] in Figures 7.7(c) and (e). In particular, the initial attribution in figure (a) produces several spurious anisotropic classifications in the vicinity of the annotated scatterers. However, the aggregation/re-attribution corrects for all of these phantom attributions. The attribution of these clutter pixels as full-aperture by the re-attribution procedure is achieved through the accurate modeling of the interference from the neighboring scatterers. The only criticism of the results in Figure 7.7 is that the attributions of the pixels on the half-aperture scatterers (scatterers B_1, D_1, and D_2 in Figure 7.2) are not all correct. In particular, while the scattering center for each scatterer is correctly attributed as half-aperture (as marked by the square annotation in Figures 7.7(c) and (e), some of the pixels neighboring the scattering center are attributed incorrectly as quarter-aperture (as indicated by their darker shading). These misclassifications are due to location error. In particular, as described in Section 6.1, the displacement from the scattering center produces a modulation on the observed azimuthal response which biases the measured energy of larger sub-apertures towards zero, and thus, biases the anisotropy characterization towards more highly anisotropic attributions.

To examine the workings of the re-attribution in further detail, Figures 7.8 through 7.11 display the segmentations produced for down-range locations DR1 through DR4 in Figure 7.2. For each of the displays, figure (a) shows the standard, full-aperture, log-magnitude reflectivity. Figures (b) through (d) show the segmentation based on

---

[6]Note that a pre-screening threshold is *not* used for these results.

(a)   (b)

(c)   (d)

(e)   (f)
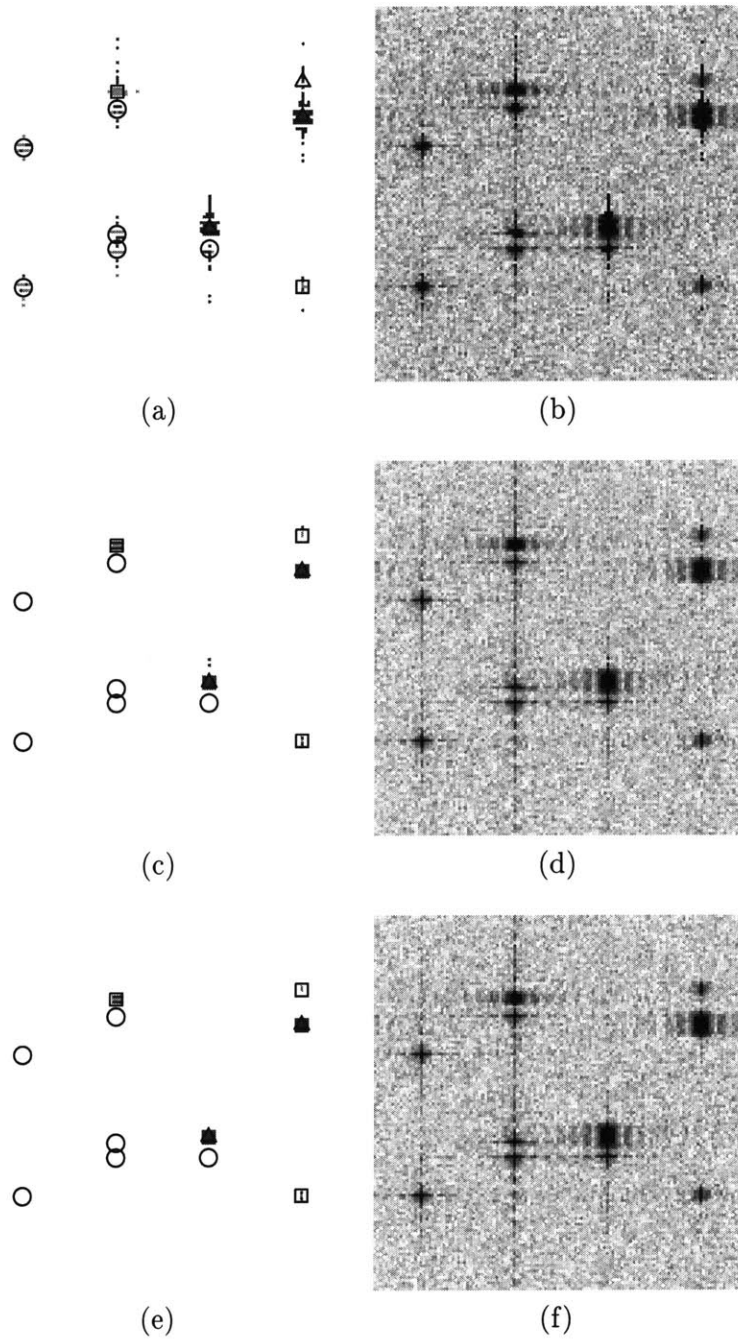
Figure 7.7. Results of the iterative anisotropy attribution on the synthesized data in Figure 7.2. First column: images of the degree of anisotropy. Second column: images of the resulting sub-aperture reflectivities. First row: initial attribution using the MSM statistic from Chapter 5. Second row: first anisotropy re-attribution. Third row: second anisotropy re-attribution.

the initial attribution, first re-attribution, and second re-attribution. The structure and information conveyed by these plots are the same as for the segmentation plots in Section 7.2.

The first characteristic of the segmentation plots that we point out is that each of the prominent isotropic scatterers (denoted as A_i in Figure 7.2(b)) is segmented into a singleton group in the first segmentation iteration. Consequently, when re-attributing clutter pixels in the vicinity of a scatterer A_i, A_i is identified as an off-scatterer pixel by the segmentation, and thus, the re-attribution procedure compensates for the interference produced by A_i resulting in the clutter being correctly attributed as full-aperture.

Another detail of the segmentation plots is that in the initial segmentation, there are a few instances where clutter pixels are grouped together in Figures 7.9(b) through Figures 7.11(b). These groupings are associated with a series of phantom anisotropy attributions in clutter. Because these phantom attributions have a consistent initial anisotropy attribution (resembling the anisotropy of the neighboring anisotropic scatterer), the segmentation interprets the pixels as a weak anisotropic scatterer. It is important to note that while these clutter pixels are grouped into the same cluster, the cluster does not include the underlying scatterer generating the interference. This separation between the true scatterer and the phantoms allows for the first re-attribution iteration to recognize and compensate for the scatterer's interference. While this first re-attribution step removes the large majority of the phantom anisotropy attributions, a few remain. We believe these remnants are due to volumetric anisotropy. In particular, all the clutter in the cluster coherently combine to produce a chaotic response in azimuth that is occasionally characterized as anisotropic. The subsequent segmentation step, however, recognizes that these are anomalous anisotropy attributions. In particular, because the anisotropy observed over the pixels is not consistent, the likelihood of all the pixels covering the same scatterer is very low, and thus, the pixels are not merged. The result is that these clutter pixels are grouped as singletons as displayed in Figures 7.9(c) through Figures 7.11(c). Consequently, the second anisotropy re-attribution characterizes these pixels as full-aperture.

As a final comment, note that the segmentation for figure (d), based on the second re-attribution, is the same as in (c) in Figures 7.8 through Figures 7.11. Because the segmentations are the same, there is no purpose in performing a third re-attribution step because the results would be the same as the second re-attribution. In fact, the second re-attribution only has to re-examine the small handful of pixels whose grouping in figure (b) is different than in figure (c).

Examining the reflectivity of the pixels in Figures 7.8 through 7.11, a skeptic might argue that the segmentation could be performed based on the log-magnitude data alone, i.e. that the anisotropy attribution is unnecessary in the segmentation step and thus the algorithm as a whole. To demonstrate that this is not the case, we modify the synthetic data by moving some of the scatterers closer together to the point that they cannot easily be resolved. In particular, the four scatterer pairs in Figure 7.2 which have been moved closer together are: A_3/A_4, A_5/B_1, A_6/C_1, and C_2/D_2 as displayed
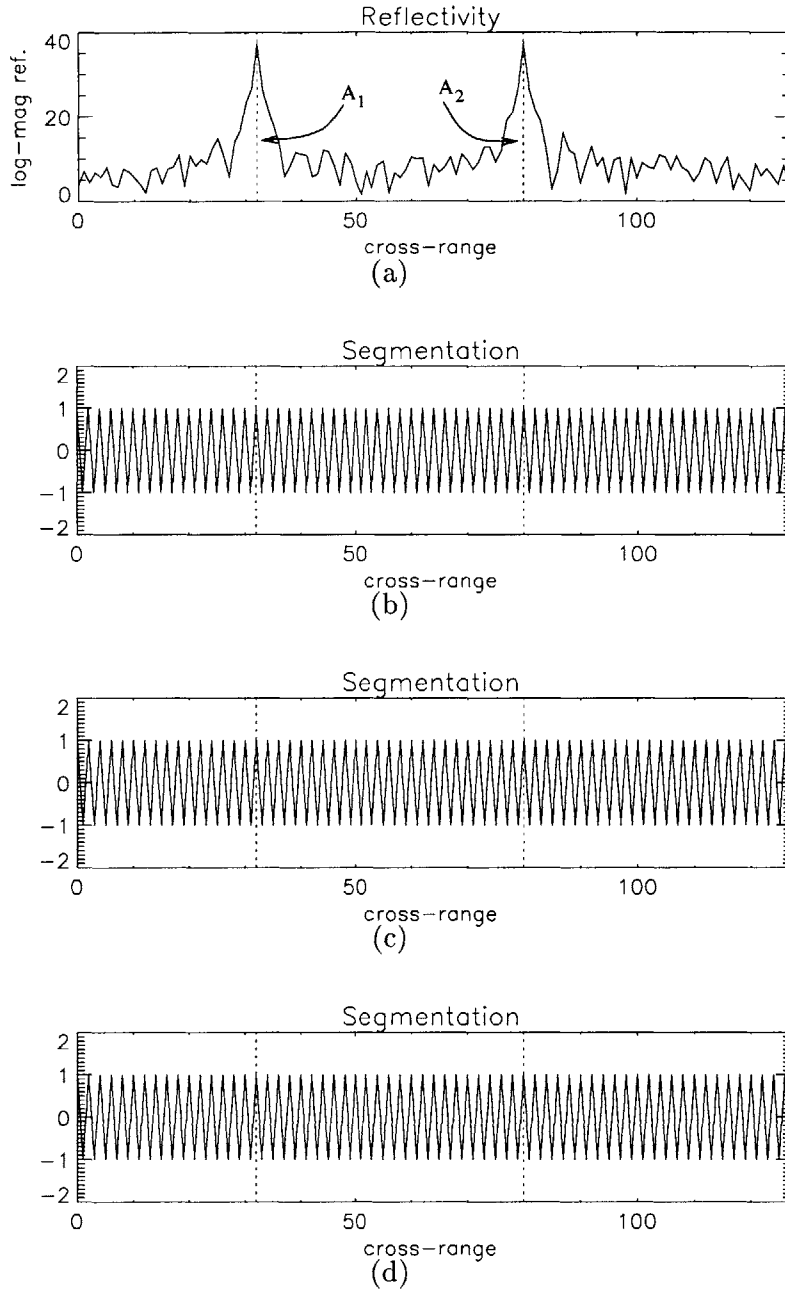
Figure 7.8. Segmentation of pixels at down-range DR1 in Figure 7.2. (a) Full-aperture reflectivity. (b) Segmentation based on the initial anisotropy attribution. (c) Segmentation based on the first anisotropy re-attribution. (d) Segmentation based on the second anisotropy re-attribution.
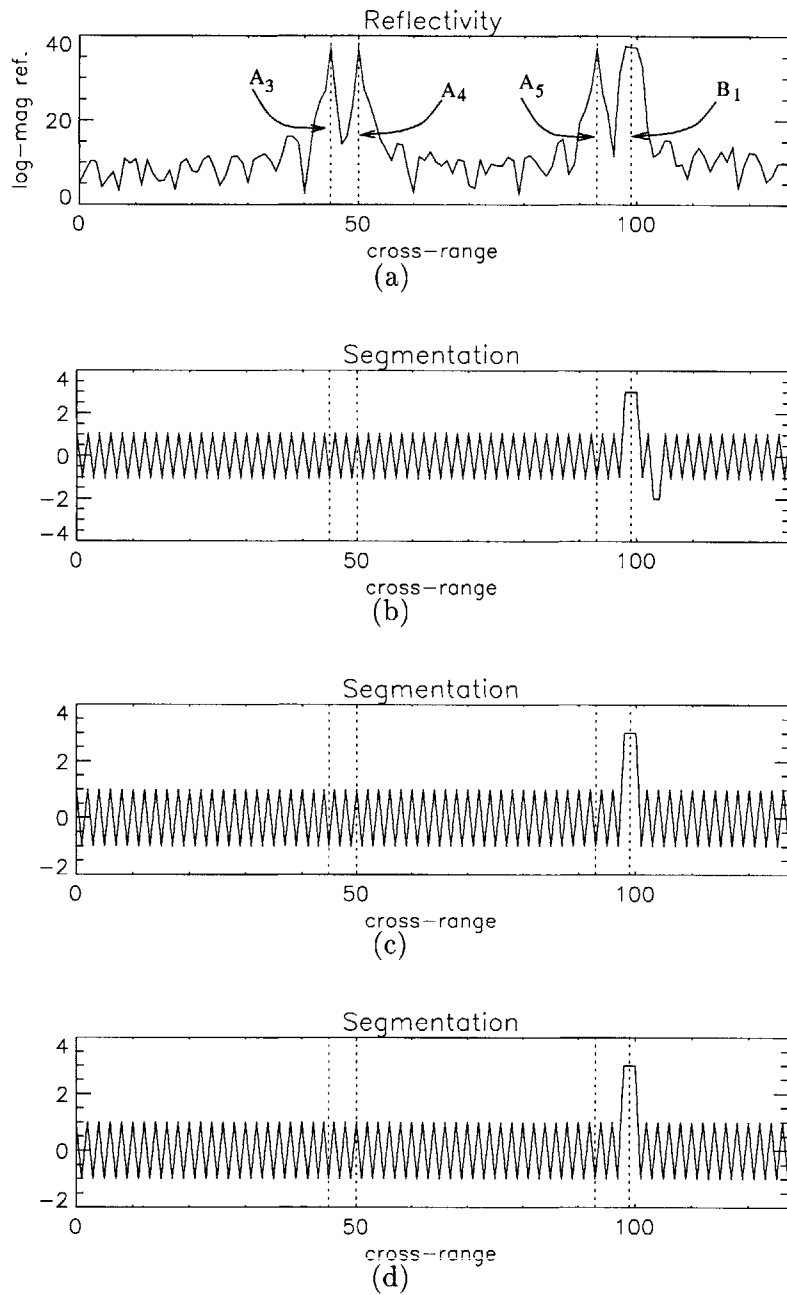
Figure 7.9. Segmentation of pixels at down-range DR2 in Figure 7.2. (a) Full-aperture reflectivity. (b) Segmentation based on the initial anisotropy attribution. (c) Segmentation based on the first anisotropy re-attribution. (d) Segmentation based on the second anisotropy re-attribution.
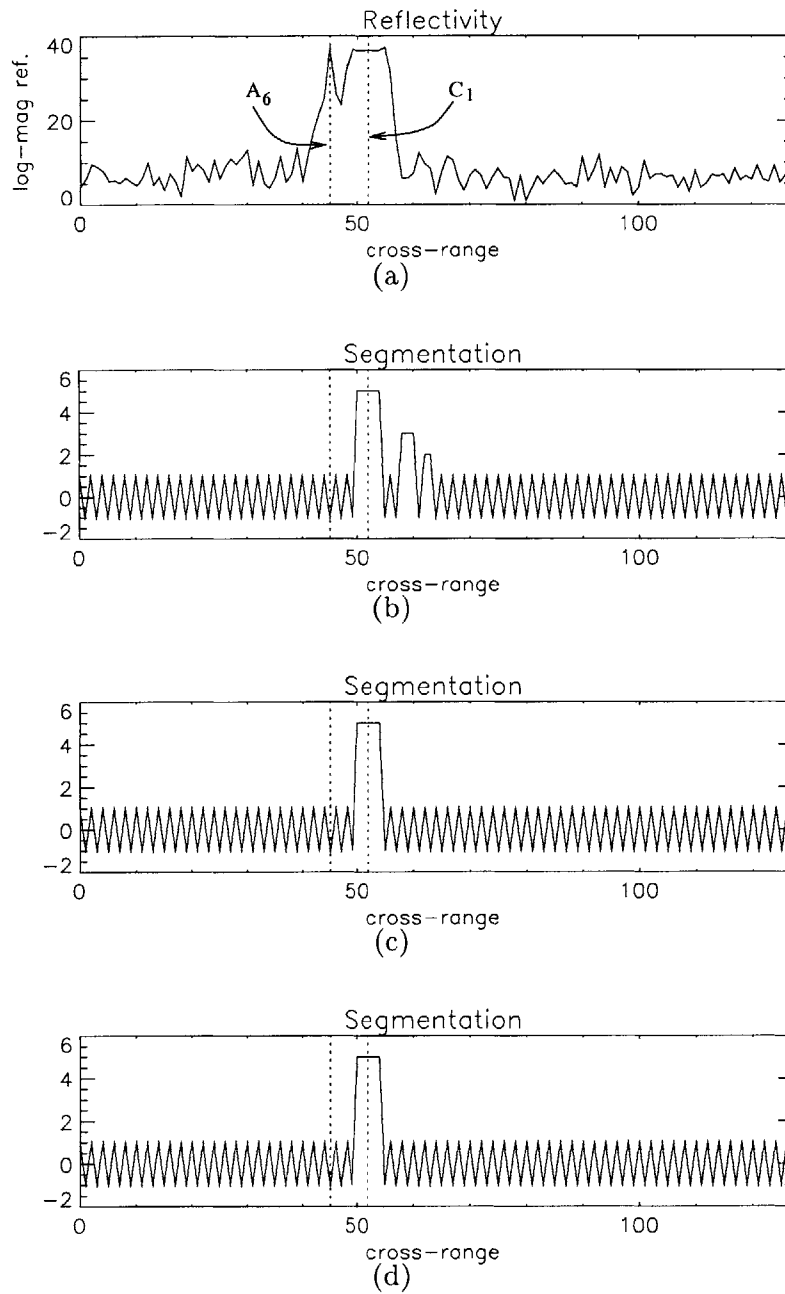
Figure 7.10. Segmentation of pixels at down-range DR3 in Figure 7.2. (a) Full-aperture reflectivity. (b) Segmentation based on the initial anisotropy attribution. (c) Segmentation based on the first anisotropy re-attribution. (d) Segmentation based on the second anisotropy re-attribution.
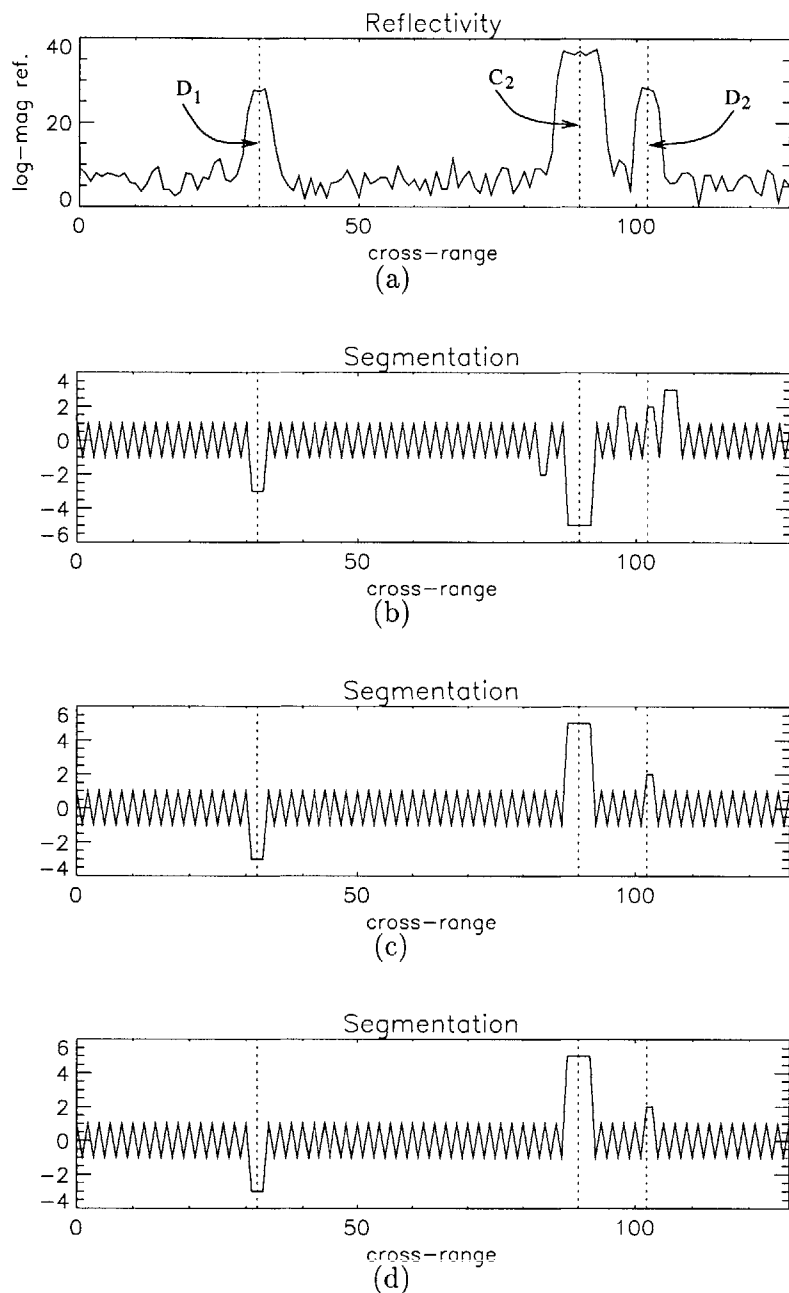
Figure 7.11. Segmentation of pixels at down-range DR4 in Figure 7.2. (a) Full-aperture reflectivity. (b) Segmentation based on the initial anisotropy attribution. (c) Segmentation based on the first anisotropy re-attribution. (d) Segmentation based on the second anisotropy re-attribution.

in Figure 7.12. The scatterers used here are the same as before since only their locations have changed. The resulting anisotropy attributions and sub-aperture reflectivities are shown in Figure 7.13. The results of the algorithm on the data in Figure 7.12 are impressive since most (10 out of 11) of the scattering centers are all correctly attributed even though they are in very close proximity to each other and difficult to resolve.

(a)   (b)
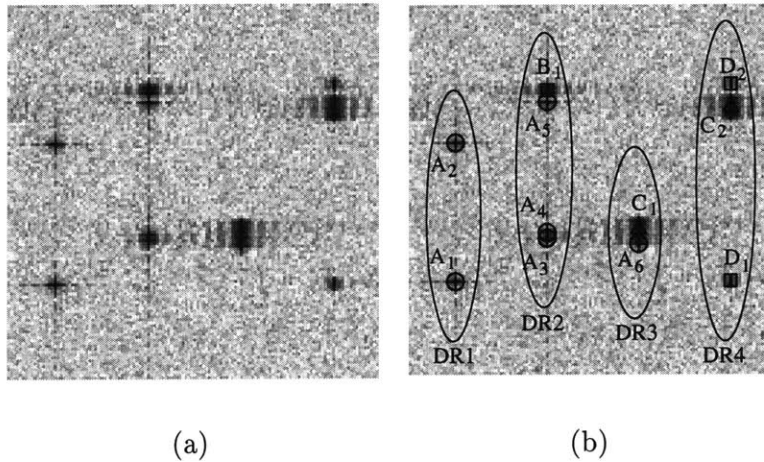
Figure 7.12. Log-magnitude image of the SAR data with closely spaced scatterers (a) without annotations and (b) with annotations.

The success in correctly attributing the scatterers lies in the use of the clustering information. The segmentation plots for the scatterers are given[7] in Figures 7.14 through 7.16. From the reflectivity plots in Figures 7.14(a) through 7.16(a), it is clear that a segmentation algorithm that uses only the log-magnitude reflectivity will fail to separate the scatterers in the new data. The aggregation procedure proposed in Section 7.2, however, makes use of the additional anisotropy information to separate the scatterers and produce the correct anisotropy attribution. The aggregation algorithm has difficulty correctly assessing the spatial extent of the scatterer, but the resulting segmentation is close enough to produce accurate anisotropy attributions.

To conclude this chapter, we demonstrate the performance of the anisotropy attribution on the BMP-2 from the MSTAR data set. In particular, the BMP-2 is taken at the cardinal angles of 0° and 90° since there is a known canonical response observed at each of these orientations due to the front and side of the vehicle. The standard log-magnitude reflectivity images for these orientations are shown in Figure 7.17. The results for the 0° azimuthal orientation are shown in Figure 7.18. The layout of this figure is the same as in Figure 7.7 except that the anisotropy attributions from only the initial and fourth iterations are shown. Again, the initial attribution is obtained using the approach described in Section 5.2. From Figure 7.18, it is apparent that the

---

[7]The segmentation is not displayed for down-range DR1 containing scatterers A_1 and A_2 as their scattering locations were not altered.

(a)   (b)

(c)   (d)

(e)   (f)
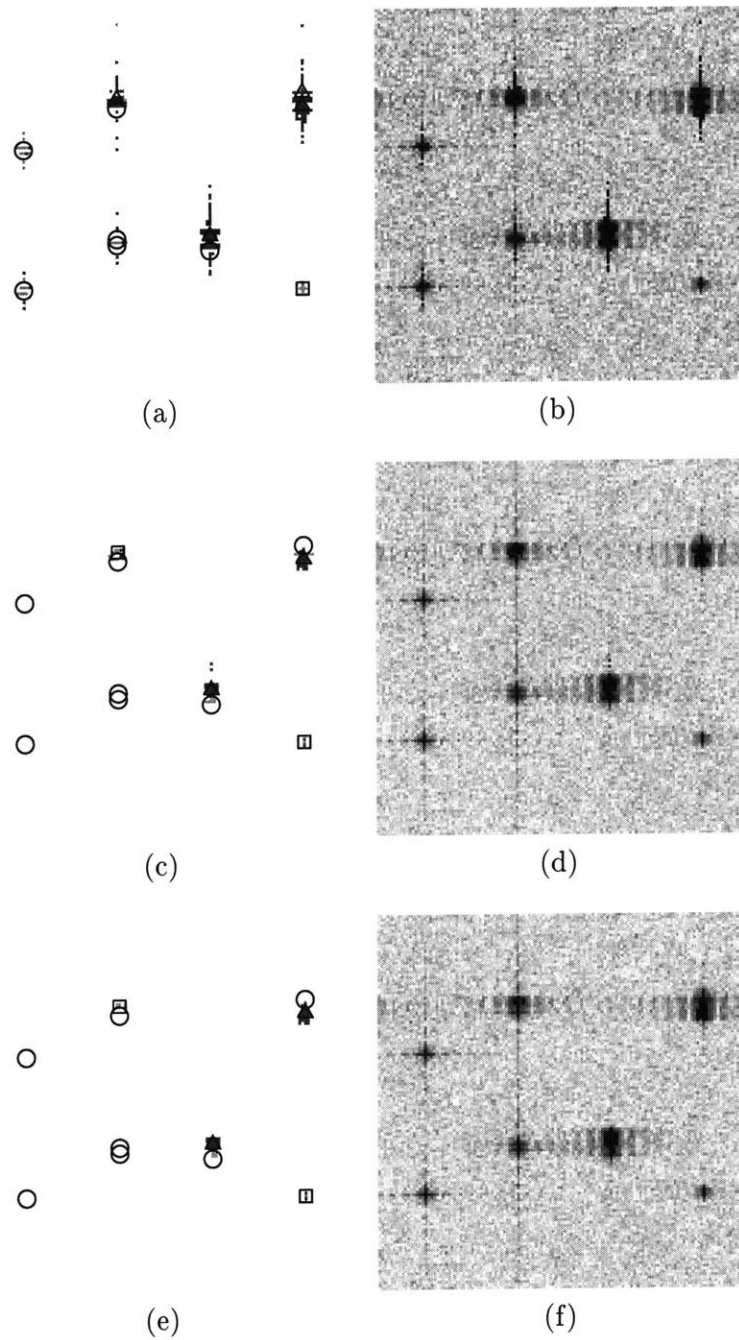
Figure 7.13. Results of the iterative anisotropy attribution on the synthesized data in Figure 7.12. First column: images of the degree of anisotropy. Second column: images of the resulting sub-aperture reflectivities. First row: initial attribution using the MSM statistic from Chapter 5. Second row: first anisotropy re-attribution. Third row: second anisotropy re-attribution.
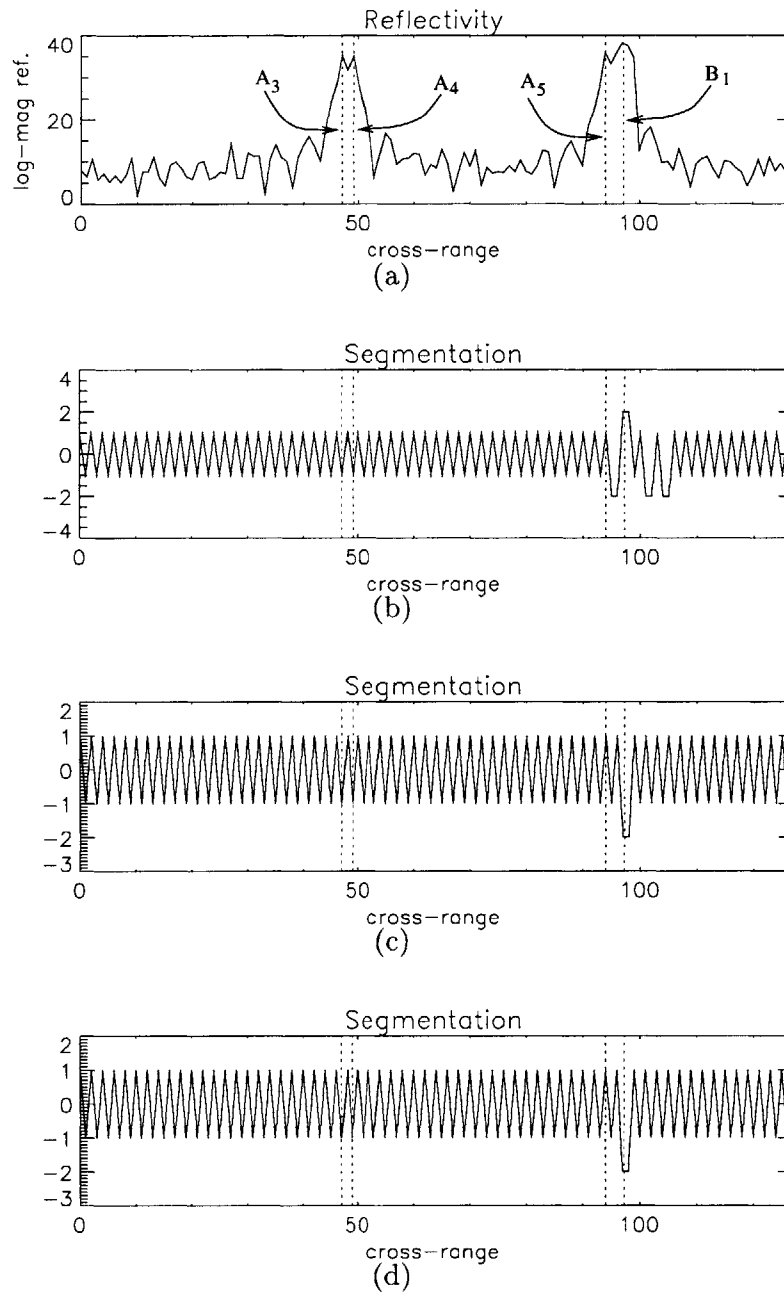
Figure 7.14. Segmentation of pixels at down-range DR2 in Figure 7.12. (a) Full-aperture reflectivity. (b) Segmentation based on the initial anisotropy attribution. (c) Segmentation based on the first anisotropy re-attribution. (d) Segmentation based on the second anisotropy re-attribution.
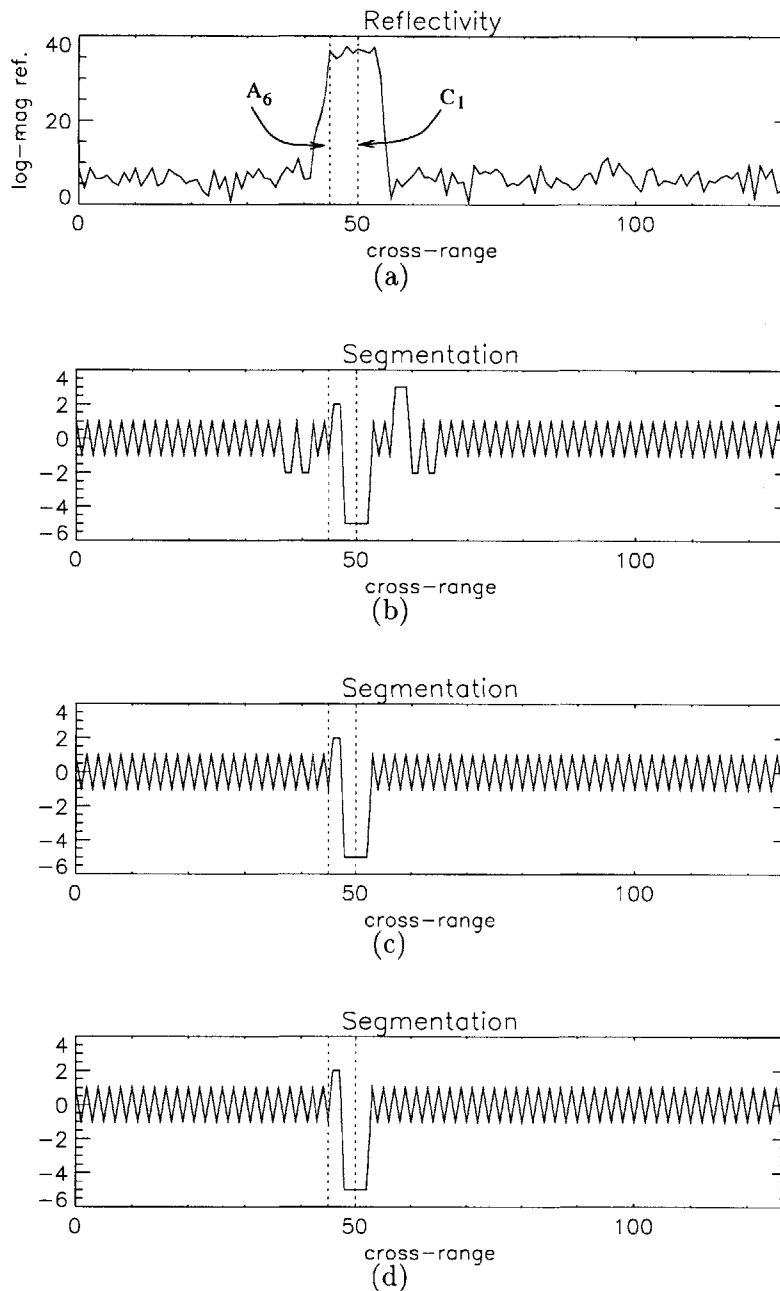
Figure 7.15. Segmentation of pixels at down-range DR3 in Figure 7.12. (a) Full-aperture reflectivity. (b) Segmentation based on the initial anisotropy attribution. (c) Segmentation based on the first anisotropy re-attribution. (d) Segmentation based on the second anisotropy re-attribution.
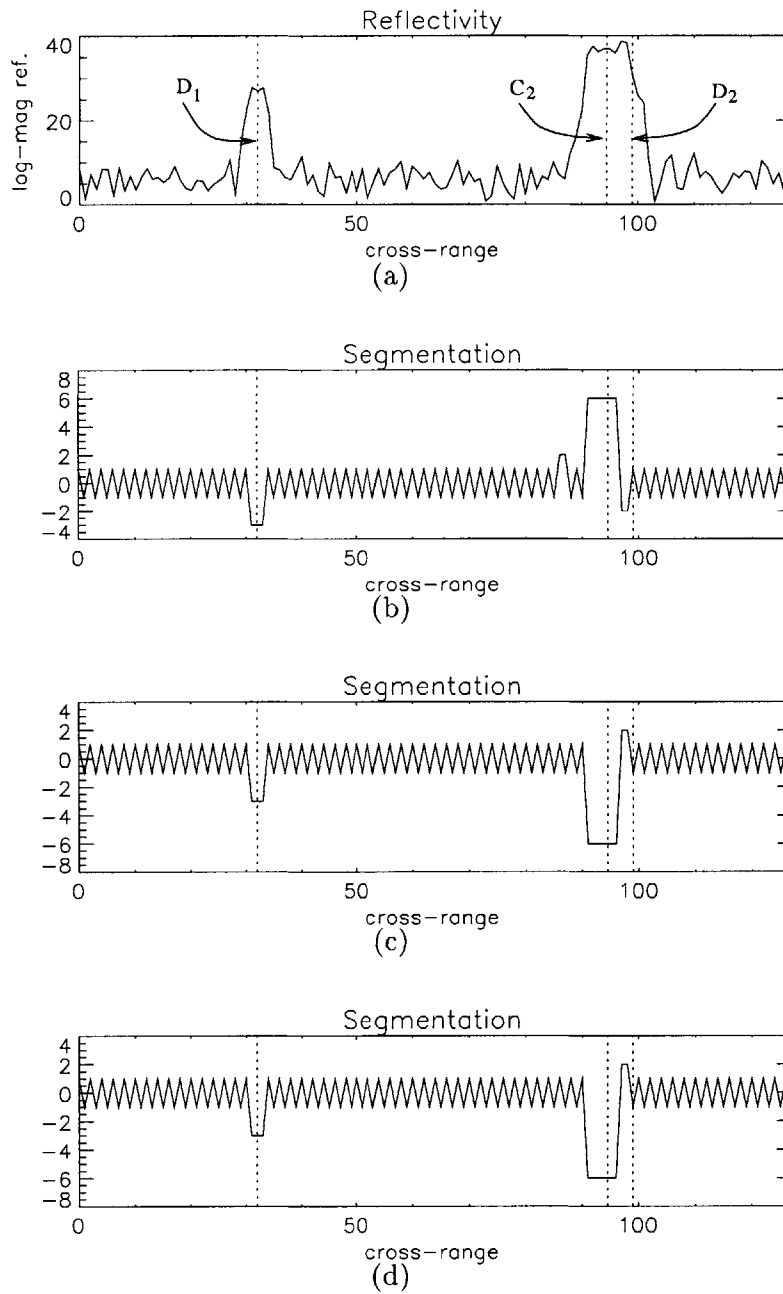
Figure 7.16. Segmentation of pixels at down-range DR4 in Figure 7.12. (a) Full-aperture reflectivity. (b) Segmentation based on the initial anisotropy attribution. (c) Segmentation based on the first anisotropy re-attribution. (d) Segmentation based on the second anisotropy re-attribution.

re-attribution significantly reduces the number of anisotropic classifications in clutter due to the strong response at the front of the vehicle[8]. Furthermore, the anisotropy for the obvious geometric scatterer at the front of the vehicle is still captured.

(a)   (b)
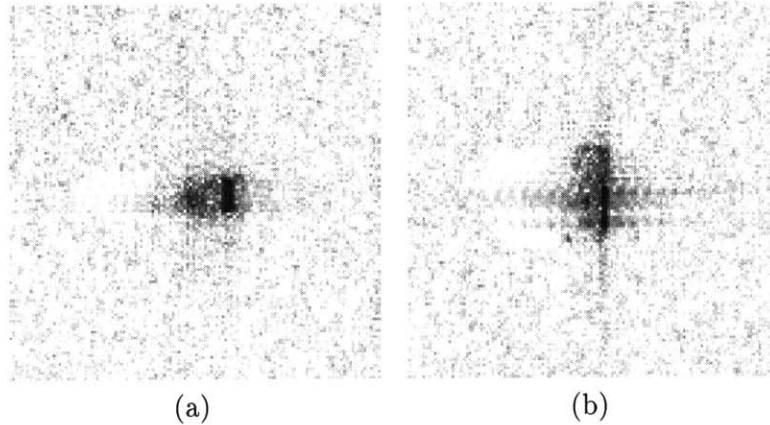
Figure 7.17. Standard full-aperture log-magnitude reflectivity image of a BMP-2 tank at an (a) 0° azimuth and (b) 90° azimuth.

Figure 7.19 similarly shows the anisotropy attributions from the initial and fourth iterations of the anisotropy attribution for the BMP-2 at a 90° azimuth. Again, the re-attribution has removed many of the anisotropy attributions which are not clearly associated with a geometric scatterer. The algorithm has, however, preserved the anisotropy attributions associated with the side of the vehicle which produces a strong specular response.

Both Figures 7.18 and 7.19 show a considerable decrease in the number of anisotropy attributions in the final iteration. Because the aggregation/re-attribution algorithm better compensates for interference than the algorithm of Chapter 5, it is likely that the anisotropy attributions that are removed by the iterative procedure are caused by unstable volumetric scattering. The removal of these attributions supports the reasoning in Chapter 5 for the weak dependence exhibited in the near-cardinal confusion matrix for anisotropy. In particular, at the end of Section 5.4, we speculated that the small values of the diagonal entries in the near-cardinal confusion matrix are due to the presence of unstable volumetric anisotropy (in addition to the obvious geometric anisotropy observed in the data). That claim is supported by Figures 7.18 and 7.19 where the majority of the anisotropy attributions are removed except for those attributions which can be readily identified (in the aggregation/re-attribution process) with a geometric scatterer, i.e. the front plate and broadside responses.

---

[8]Note that for the MSTAR images in Figure 7.17, the aperture is to the right of the image. Consequently, the right-most part of the vehicle in the image corresponds to the front.

(a)   (b)

(c)   (d)
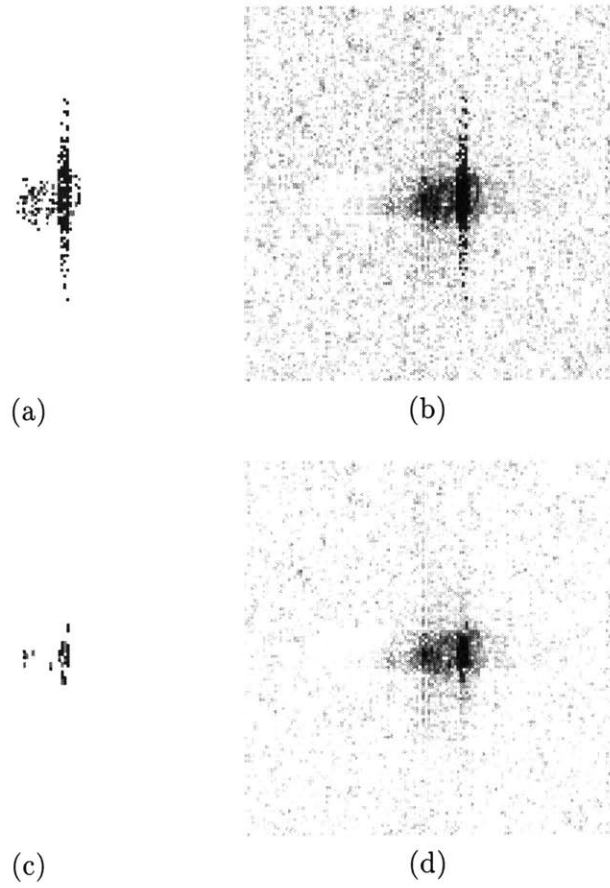
Figure 7.18. Anisotropy attributions and reflectivity images of the BMP-2 at a 0° azimuth. First column: images of the anisotropy attributions. Second column: images of the associated sub-aperture reflectivities. First row: initial anisotropy attribution. Second row: fourth iteration of the anisotropy attribution.

(a)   (b)

(c)   (d)
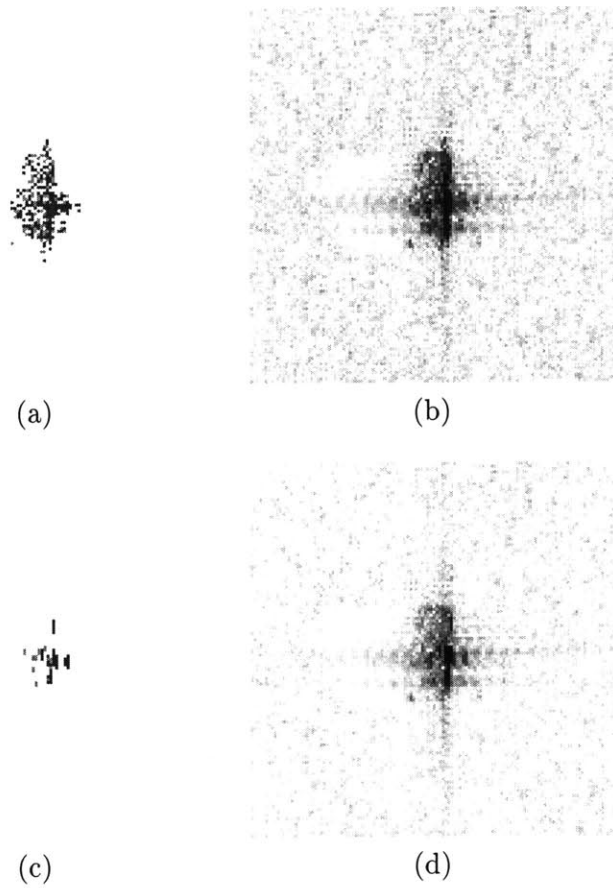
Figure 7.19. Anisotropy attributions and reflectivity images of the BMP-2 at a 90° azimuth. First column: images of the anisotropy attributions. Second column: images of the associated sub-aperture reflectivities. First row: initial anisotropy attribution. Second row: fourth iteration of the anisotropy attribution.

Chapter 8

Contributions and
Suggestions for Future Research

This thesis contributes to the field of SAR ATR through the exploitation of nonideal scatterer anisotropy. Two fundamental methodologies are used towards this end. The first is a nonparametric approach that classifies targets based on observed anisotropy behavior that is learned from the data. The second (and more developed) approach is a parametric characterization of the unimodal azimuthal response that is common from canonical scatterers. This chapter summarizes these contributions in Section 8.1 and provides suggestions for further research in Section 8.2.

■ 8.1 Summary of Contributions

This section summarizes the contributions of this thesis by chapter.

Sub-aperture Representation

The cornerstone of the analysis in this thesis is the sub-aperture pyramid presented in Chapter 3. This structure divides the aperture into various sub-apertures and arranges them in a manner that allows for the exploitation of scatterer anisotropy throughout the remainder of the thesis. One function of the pyramid is to represent the data in an over-complete basis that reveals anisotropic scattering. In particular, each level of the pyramid corresponds to a particular cross-range–azimuthal resolution trade-off. Not only does the pyramid provide a library of representations, but these representations are structured such that anisotropy in the data exhibit distinctive patterns in the pyramid. In particular, unimodal azimuthal scattering appears as a "streak" down the pyramid of reflectivity estimates where properties of the streak convey information about the anisotropy.

The data in the pyramid are used in two ways. The first is a nonparametric approach in which the patterns of the data are learned with a Parzen density estimator. The second approach uses the sub-apertures to define anisotropy hypotheses and uses the sub-aperture reflectivity measurements as observations to test the hypotheses. The sub-aperture scattering hypotheses approximate the unimodal anisotropic scattering

predicted by physics-based models for many canonical scattering types. In particular, the hypotheses characterize the degree of anisotropy which is related to scatterer size.

Nonparametric Estimation of Anisotropy

Chapter 4 of this thesis extends the density-based classification approach presented by DeBonet et. al.[20]. In both approaches, an image is characterized by generating a set of features whose statistical relationships are learned with a nonparametric density estimator. This learning process is applied to an image of an unknown target and images of reference vehicles. A classification decision is then made by comparing the density learned from the unknown target image to the densities learned from the reference images.

The work in Chapter 4 extends DeBonet's method in two ways. The first is our use of the sub-aperture feature set which is motivated by scattering physics, in contrast to DeBonet's wavelet-based feature set that is motivated by image texture. The sub-aperture feature vector is taken as the pyramid of log-magnitude sub-aperture reflectivities for a given location. These reflectivities convey the variations of the azimuthal response produced by a scatterer.

The other extension of DeBonet's work is in the manner in which the divergence comparing two densities is approximated. DeBonet uses a summation motivated by the law of large numbers. This approximation, however, can lead to the use of an excessive number of samples in the summation. In particular, DeBonet's approximation tends to evaluate the summand at a high number of clutter samples due to the predominance of clutter in the image. The approximation proposed in Section 4.2.4, however, is based on a Riemann sum where the samples are uniformly spread throughout the feature space thus preventing too many samples from being allocated toward clutter. The result is a significant reduction in the computation required.

Sub-aperture Anisotropy Attribution

Chapter 5 develops a method for attributing anisotropy. The basis of the attribution is the sub-aperture scattering model introduce by Chaney et. al.[12] and Allen et. al.[2] that characterizes the azimuthal response from a scatterer. The attribution is obtained by developing an LRT for anisotropy hypotheses defined from the sub-aperture pyramid. The first test presented is a GLRT in which the ML reflectivity estimate (under the associated anisotropy hypothesis) is used for the value of the reflectivity parameter. The resulting GLLR is called the *basic GLLR* statistic and corresponds to a difference in average sub-aperture energy between the two sub-aperture hypotheses comprising the likelihood ratio. Thus, average sub-aperture energy is a sufficient statistic for this model.

While the basic GLLR statistic is consistent with our intuition for characterizing azimuthal concentration, it is not robust to interference from neighboring scatterers which the sub-aperture scattering model does not include. This issue is addressed in two ways. The first address is to recognize that the ML estimate (under a given anisotropy

hypothesis) of reflectivity has a corresponding resolution that varies with the hypothesis, and thus, the hypotheses are not consistent with each other. To deal with this issue, a scaling of the measured full-aperture reflectivity is used to estimate the sub-aperture reflectivity, where the scaling factor provides the appropriate normalization for using the full-aperture to measure a partial-aperture response. Under this approach, all the reflectivity estimates are imaged at the finest resolution, and thus, the hypotheses are consistent with each other. This new reflectivity estimate results in a slight modification of the basic GLLR statistic. In particular, a term penalizing the amount of energy outside the hypothesized sub-aperture is added to the statistic. This modified GLLR statistic significantly improves the accuracy of the anisotropy attribution, but the effects of interfering scatterers are still evident.

The second method used to address interfering scatterers explicitly includes their effects in the scattering model. In particular, the sub-aperture scattering model is extended to allow for scatterers at neighboring pixel locations. In this extended model, the nuisance reflectivity parameter corresponds to a *vector of reflectivities*. These reflectivities are estimated in an ML fashion using a Gaussian prior on the reflectivities to regularize the estimation problem. The estimated reflectivities are then used to determine a likelihood statistic for the sub-aperture hypotheses. The LRT using this MSM statistic performs better than the modified GLLR but is still vulnerable to interference from neighboring scatterers.

To address the effects of interfering scatterers beyond the use of the modified GLLR and MSM statistics, Section 5.3 proposes performing the hypothesis tests in a telescopic, decision-directed fashion to take advantage of the structure provided by the sub-aperture pyramid. The telescopic testing procedure starts at the top of the pyramid and traverses a path down the pyramid following the most likely sub-aperture hypothesis at each step. Not only does this testing reduce the number of sub-apertures that need to be tested, but the telescopic testing enforces a consistency between the sub-aperture hypotheses which reduces the number of anisotropic classifications due to interfering scatterers.

The chapter concludes with application of the anisotropy attribution to the MSTAR data. To evaluate the utility provided by the anisotropy attribution, confusion matrices for the attribution are empirically estimated. In particular, the confusion matrices are separately presented for the case of near-cardinal and off-cardinal orientations. The near-cardinal orientations exhibit a mild dependence of the "test" attribution on the "truth" attribution, although this dependence is not as strong as we had expected. The anisotropy attribution for off-cardinal orientations is even less stable in the sense that the "test" attribution is independent of the "truth" attribution. We conjectured that the source of the instability is volumetric scattering which arises from the interference of irresolvable scatterers and produces an erratic, scintillating response in azimuth. Thus, the anisotropy characterization of these volumetric scatterers is also erratic as the azimuthal energy concentration fluctuates and is thus sensitive to slight changes in imaging orientation. The aperture used in the MSTAR data is too narrow to observe

anisotropy due to canonical scatterers except at near-cardinal orientations. The remainder of the anisotropy attributions are thus believe to be due to volumetric scatterers. This limited observation of canonical scattering explains the weak dependence observed at near-cardinal azimuths and the independence observed at off-cardinal azimuths in the MSTAR data.

Relation Between Phenomenology and Anisotropy Attribution

The results on the MSTAR data in Chapter 5 motivated the study in Chapter 6 of anisotropy phenomenology and its effects on anisotropy attribution. Two sources of anisotropy are considered in this chapter. The first is the class of geometric scatterers which have a simple geometric structure and produce a consistent anisotropic response. The second source considered is the class of volumetric scatterers that are composed of irresolvable point scatterers that interfere in a manner resulting in a chaotic azimuthal response which is occasionally characterized as anisotropic. For each case, the information conveyed by the anisotropy attribution is evaluated. In particular, anisotropy plots are shown to convey the dependence of the anisotropy attribution on the scatterer "size". For geometric scatterers, we find that there is a strong dependence of the anisotropy attribution on the scatterer size. Furthermore, the attributions are consistent with the actual concentration of the azimuthal response. However, for volumetric scatterers, we demonstrated that the strength of the dependence varies with the underlying scatterer density. In particular, for low-density volumetric scatterers, the anisotropy attribution is independent of scatterer size and the full-aperture attribution is usually produced. For high-density volumetric scatterers, the anisotropy attribution is highly dependent on scatterer size because the scatterer resembles a geometric scatterer.

It is argued and demonstrated in this chapter that by using a wider aperture, one can reduce the incidence of observed volumetric anisotropy while preserving the observed geometric anisotropy which is more meaningful in characterizing the scatterer. The task, however, is not as simple as just increasing aperture length. In particular, the effects of location uncertainty must be addressed. Increasing the aperture length increases the location uncertainty for physically large geometric scatterers, but reduces the location uncertainty for volumetric scatterers. To address this apparent conflict, we presented a decision-directed peak extraction method that uses the anisotropy attribution to adapt the resolution of the peak extraction to the underlying scatterer.

Chapter 6 also demonstrates the benefit of additional anisotropy hypotheses that wide-aperture data allows to be defined. Larger sub-apertures can be used to form sub-aperture hypotheses which better characterize the anisotropy produced by smaller scatterers. In narrow-aperture settings, these small scatterers may produce an azimuthal response whose mainlobe is wider than the aperture and thus can only be characterized as full-narrow-aperture. However, with a larger aperture, the localized concentration of this response in azimuth can be better characterized as the full-narrow-aperture attribution is no longer the minimal anisotropy characterization available.

Iterative Anisotropy Attribution

Chapter 7 returns to the task of improving the anisotropy attribution. An iterative approach is proposed to model and remove the effects of interfering scatterers. The proposed iteration is a two step procedure that first clusters pixels according to their corresponding scatterers and then re-attributes pixels using the clustering information to suppress the effects of neighboring scatterers. The clustering is performed with a greedy algorithm that maximizes the likelihood of an observed set of features corresponding to pixel reflectivity, pixel anisotropy, and scatterer size. Due to the manner in which the observed features are modeled, an additive cost function can be used to evaluate the likelihood of a segmentation. Evaluation of all possible segmentations is computationally prohibitive. Thus, the ML segmentation is approximated by a greedy algorithm that at each iteration merges the two regions which result in the largest increase of the likelihood.

The re-attribution procedure extends the regularization approach of Section 5.2. In particular, the vector of reflectivities (of the focused pixel and its neighbors) is modeled as a noisy observation of the finest resolution reflectivity measurements instead of a zero-mean Gaussian prior (as done in Chapter 5). Incorporating this observation model into the estimation of the reflectivity nuisance parameter produces a term that regularizes the estimate towards the observed reflectivities. In addition, as described in Chapter 7, the aggregation of pixels into distinct, physically extended scatterers allows us to devise an improved method for removing interference from such neighboring scatterers.

The results from using the iterative clustering–re-attribution procedure are quite accurate. Significant improvements are made over the attribution of Chapter 5. In particular, virtually all of the anisotropy misclassifications (due to interference from a neighboring scatterer) in clutter are removed. Furthermore, the attribution of weak scatterers that are corrupted by neighboring strong scatterers are also improved.

■ 8.2 Suggestions for Future Research

The research presented in this thesis opens several new directions for further research which are described in this section.

Sub-band Analysis

The focus of this thesis has been on sub-aperture analysis to exploit aspect dependent scattering in azimuth. This work is motivated by the physics-based PO and GTD scattering models which predict aspect dependencies based upon the canonical scatterer type, e.g. flat plate, dihedral, trihedral, top-hat, sphere, etc.. The PO and GTD scattering models also predict a similar dependence in frequency. Frequency is analogous to azimuth but for the down-range component instead of cross-range. The PO and GTD models predict that frequency concentration characterizes down-range extent in the same way that azimuthal concentration characterizes cross-range extent. The PO response for several canonical scatterers is presented in Eqs. (2.16)-(2.19) of Chapter 2.

Because of the similarity between frequency and azimuth, much of the sub-aperture analysis should carry over sub-band analysis. However, there are two differences that one should keep in mind when adapting the sub-aperture analysis to frequency. One difference is that the sinc-like dependence is actually a function of the reciprocal of frequency instead of frequency. Thus, the mode of the sinc will occur at the highest frequency in the band. However, besides this sinc component of the response there is an additional scaling factor of the form $f^{n/2}$ where n is an integer (positive, negative, or zero) that is determined by the canonical scatterer type. Negative values of n may shift the location of the mode to something other than the highest frequency. Furthermore, this frequency dependent scaling factor may significantly modify the envelope on the response, especially in ultra-wideband data. It is suspected that because the methods in Chapters 5 and 7 have been developed to be robust to slight functional deviations in the modeled response, this additional factor in the frequency dependence should not significantly detract from the applicability of the analysis to frequency instead of azimuth. However, such supposition should be studied and verified, and not taken as fact.

The application of the sub-aperture analysis to frequency is not expected to pose significant new challenges; however, this expectation assumes that one does sub-band analysis *instead* of sub-aperture analysis. If one wishes to do both sub-band and sub-aperture analysis simultaneously, the problem becomes considerably more difficult. Problems manifest themselves at the very core of the analysis with the generalization of the sub-aperture pyramid. An intuitive generalization is to extend the pyramid to a two dimensional structure resembling the quad-trees used in the multiscale analysis of images. In such an extension, the intervals corresponding to sub-apertures now generalize into "squares" that are the product of a sub-aperture and sub-band. Such a generalization seriously restricts the scattering phenomena that can accurately be captured in a concise manner. In particular, there is no reason to believe that just because scattering energy is concentrated in azimuth (or frequency) that the energy is concentrated in frequency (or azimuth). For example, a scatterer whose down-range extent is significantly smaller than its cross-range extent will have a broad energy distribution in frequency, but a high energy concentration in azimuth. Such a scatterer would not be concisely captured in the proposed pyramidal representation. With such scatterers in mind, we believe that one would have to generalize the pyramid to

$$S \times S = \{S_{m,i} \times S_{n,j} | S_{m,i}, S_{n,j} \in S\}$$

in order to capture scatterers whose cross-range and down-range extents do not match.

Besides extending the pyramidal representation, there is also the issue of how to model neighboring scatterers which are now located in both down-range and cross-range. To model all neighboring pixels, as done in this thesis, is computationally burdensome due to significant increase in the model order from N (neighboring pixels in cross-range at a single range) to N^2 (neighboring pixels in both cross-range and down-range). Considering that the iterative approach in Chapter 7 involves a matrix inversion, the

high cost of using all N^2 neighboring pixels is clear. Practical application of such models will almost certainly need to exploit the sparsity of scatterers in the imagery.

Replacing the Sub-aperture Scattering Model

The sub-aperture scattering model is used to model anisotropy both because of its simplicity and its general characterization of azimuthal concentration. An interesting direction for further research is to evaluate the efficacy of other models for the azimuthal response. A drawback of the sub-aperture scattering model is that the corresponding response in the image domain is a sinc function. Because the sinc function has an infinite extent, it is clear that no scatterer will truly fit this response. Furthermore, the slow spatial decay of this response may leave the attribution unnecessarily vulnerable to strong interfering scatterers. With this thought in mind, it may seem obvious that some other function, which decays faster in the spatial domain, should be used. Examples of alternative azimuthal response model are the sinc function (corresponding to a boxcar in the spatial domain) and the Gaussian pulse (corresponding to a Gaussian pulse in the spatial domain). One must be cautious, however, in using other azimuthal scattering models. In particular, there is the issue of how modeling errors in the azimuthal domain will effect the resulting anisotropy test. As an example, consider using a boxcar in the spatial domain (or equivalently a sinc in the azimuthal domain). The sensitivity of the anisotropy test to the misalignment of the zero crossings of the modeled sinc to the actual response[1] needs to be taken into consideration. This misalignment may be due to either a difference in the fundamental frequency of the two responses or a difference in the azimuthal location of the mainlobe. Regardless of the source, if the misalignment is severe enough, then there may be sub-apertures in which the sidelobes of the modeled and actual azimuthal response are 180° out of phase resulting in a significant error in that sub-aperture.

We believe that gains can be achieved by using a better model than the sub-aperture scattering model. Our intuition is that a sinc azimuthal response (or any other azimuthal response with significant oscillations) will not be significantly more effective than the boxcar response due to problems with the alignment of the zero-crossings. However, we suspect that smooth unimodal functions, such as the Gaussian, would perform better than the sub-aperture scattering model. Because such a function does not change sign, there is no concern about the alignment of zero-crossings. Furthermore, because the model function is smooth, it will exhibit faster decay in the image domain than the sub-aperture scattering model.

Measure of Sub-aperture Stability

In Chapter 6, we discuss how the use of a wider aperture can reduce the incidence of volumetric anisotropy. This reduction can be achieved because volumetric scatterers produce a chaotic azimuthal response whose energy is distributed across the aperture,

---

[1]Many scatterers produce an azimuthal response which oscillates.

while geometric scatterers produce a highly concentrated azimuthal response. However, this approach does not provide perfect rejection of volumetric scatterers. An interesting direction for further research is to characterize each anisotropy attribution with a confidence measure conveying how likely it corresponds to geometric anisotropy rather than volumetric anisotropy. Such a confidence measure could be used in tandem with the anisotropy attribution to weight its importance in a classification algorithm.

The intuitive motivation for obtaining this measure is the same as the intuition behind the reduction of volumetric anisotropy in wide-aperture data. In particular, we believe that a measure that characterizes the goodness-of-fit of the observed azimuthal response to the hypothesized azimuthal response should convey how likely an anisotropy attribution is due to a geometric scatterer rather than a volumetric scatterer. Because volumetric scatterers produce a chaotic azimuthal response, even if the response is declared anisotropic, the hypothesized azimuthal response will be a poor fit to the volumetric scatterer's actual response. A geometric scatterer on the other hand should demonstrate a very high similarity to the modeled azimuthal response due to its structured, highly concentrated azimuthal response.

This idea is similar in spirit to the work of Smith *et. al.*[51] where they propose an ad-hoc "peak-stability" measure to characterize the deviation of the azimuthal response from the isotropic response. We believe that a principled statistic can be obtained which also applies to sub-aperture scattering models.

Extending the Hybrid-Resolution Peak Extraction

The hybrid-resolution peak extraction/attribution used in Chapter 6 uses a decision-directed method to determine which resolution should be used for the anisotropy attribution. The method first examines the anisotropy attribution for the fine-resolution peak extraction, and if the attribution is anisotropic, then the coarse-resolution peak extraction is used for the attribution. This approach was restricted to the use of two resolutions for simplicity and to demonstrate proof-of-concept for the suppression of volumetric anisotropy in wide-aperture data. The approach should be generalized to use more than just two pre-determined resolutions for the peak extraction. We believe it would be more effective (both for the peak extraction and the anisotropy attribution) to perform the peak extraction and anisotropy attribution in alternating steps. In particular, the peak extraction is repeatedly performed at successively coarser resolutions until the anisotropy attribution produced for the peak extraction conveys a scatterer size that is consistent with the resolution at which the peak of the scatterer is extracted.

Appendix A

DeBonet's Wavelet-based Classifier

In [20], DeBonet et. al. introduce a SAR classifier that exploits the differences in cross-scale dependencies in image data to distinguish various targets. The classification setting and underlying principles of the algorithm proposed by DeBonet are the same as those described in Section 2.4. In this appendix, DeBonet's approach is described in more detail using the notation and ideas from Section 2.4. The cross-scale dependencies are learned in a nonparametric fashion by applying the Parzen density estimator to feature vectors constructed from a multiresolution representation of the image data. Classification decisions are then made by comparing the density estimate from an unknown target to density estimates from labeled targets as in Section 2.4.3. The remainder of this appendix describes these steps in more detail. Section A.1 presents the wavelet-based feature set used to represent the data. Section A.2 describes how these features are learned. Section A.3 then discusses how the estimated densities are used to classify vehicles.

■ A.1 Wavelet-based Representation

DeBonet bases his data representation on the wavelet transform[1] which represents the image data at multiple resolutions. The SAR data used in DeBonet's approach is the log-magnitude SAR image, i.e. the phase information is discarded. Starting from this image, DeBonet generates a sequence of multiresolution images by recursively applying the wavelet transform and down-sampling by a factor of two to obtain the low-frequency component of the image as illustrated in the left-hand side of Figure A.1. In this figure, W represents a circular convolution with the scaling function for the wavelet, $\downarrow 2$ represents a down-sampling by two in each image dimension, and G_m represents the image at scale m with $m = 0$ denoting the finest resolution and $m = M$ denoting the coarsest resolution. These images correspond to a sequence of nested sub-space representations at successive resolutions. DeBonet applies a series of filters $f_{m,1}, f_{m,2}, \ldots, f_{m,N}$ to G_m to reveal particular structures at that scale as illustrated in the right side of Figure A.1. In particular, DeBonet uses oriented Laplacian filters for the filters $f_{m,i}$. The image resulting from the application of filter $f_{m,i}$ to image G_m is denoted as $F_{m,i}$.

---

[1]It is assumed that the reader is familiar with wavelets. For a thorough treatment of wavelets, the reader is referred to Cohen[14], Vetterli[57], Wickerhauser[59], and Daubechies[18].
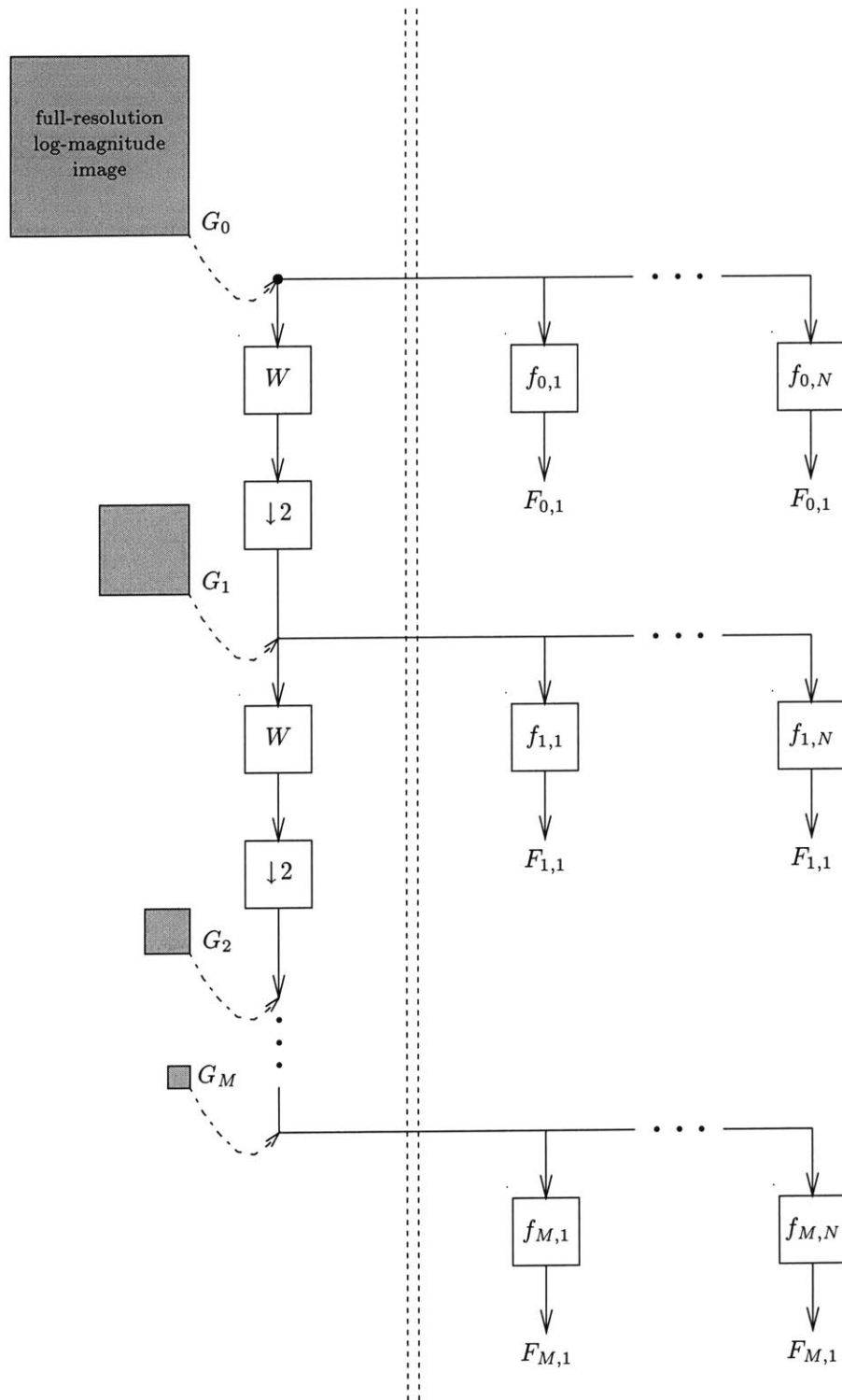
Figure A.1. Generation of the wavelet-based feature vectors. Left-hand side: Construction of the multiresolution SAR images G_m. Right-hand side: Filtering of multiresolution images to reveal image structure.

A feature vector is assembled for each pixel at the finest resolution by collecting from each image $F_{m,i}$ the components corresponding to that location. In particular, if the finest resolution filtered image pixels $F_{0,i}(x,y)$ are indexed by $x \in \{1,\ldots,N_1\}$ and $y \in \{1,\ldots,N_2\}$, then the feature vector[2] $V(x,y)$ for location (x,y) can be written as

$$
V(x,y) = \left[F_{0,1}(x,y), F_{0,2}(x,y), \ldots, F_{0,N}(x,y), \right.
$$
$$
F_{1,1}\left(\left\lceil\frac{x}{2}\right\rceil, \left\lceil\frac{y}{2}\right\rceil\right), F_{1,2}\left(\left\lceil\frac{x}{2}\right\rceil, \left\lceil\frac{y}{2}\right\rceil\right), \ldots, F_{1,N}\left(\left\lceil\frac{x}{2}\right\rceil, \left\lceil\frac{y}{2}\right\rceil\right), \ldots
$$
$$
\left. F_{M,1}\left(\left\lceil\frac{x}{2^M}\right\rceil, \left\lceil\frac{y}{2^M}\right\rceil\right), F_{M,2}\left(\left\lceil\frac{x}{2^M}\right\rceil, \left\lceil\frac{y}{2^M}\right\rceil\right), \ldots, F_{M,N}\left(\left\lceil\frac{x}{2^M}\right\rceil, \left\lceil\frac{y}{2^M}\right\rceil\right) \right]
$$
$$
\text{(A.1)}
$$

where $\lceil \cdot \rceil$ represents the least integer function. Such a feature vector is depicted in Figure A.2 where all the filtered images for each scale are collected into a vector-valued image, and these vector images can be arranged into a quad-tree due to the dyadic sampling variation in scale. The feature vector for pixel (x,y) then corresponds to the branch of the tree extending from the root node $\left(\left\lceil\frac{x}{2^M}\right\rceil, \left\lceil\frac{y}{2^M}\right\rceil\right)$ at scale M to (x,y) at scale 0.
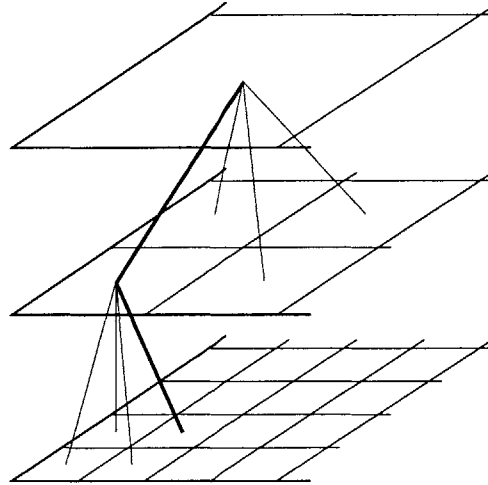
Figure A.2. Illustration a multiresolution feature vector used by DeBonet. Each vector-valued image is composed of the filtered images for the corresponding resolution. Each branch (such as the heavily shaded path) in the resulting quad-tree constitutes a feature vector.

---

[2]In order to keep the notation clear, V is used instead of X to denote the sample feature vector because x is used to denote location.

■ A.2 Learning Dependencies

DeBonet uses the wavelet-based feature set to reveal the cross-scale dependencies in a SAR image. To characterize these dependencies, he uses the Parzen density estimator in Eq. (2.20) with a hyper-cube kernel, i.e.

$$\hat{f}(\boldsymbol{v}) = \frac{1}{N_1 N_2 h^d} \sum_{i=1}^{N_1 N_2} \left\| \frac{\boldsymbol{v} - \boldsymbol{V}_i}{h} \right\|_\infty \tag{A.2}$$

where the $\boldsymbol{V}_i$ are the feature vectors generated from an image, $\|\cdot\|_\infty$ denotes the infinity norm, and each side of the hyper-cube[3] measures $2h$.

■ A.3 Decision Rule

To classify targets in images, DeBonet compares density estimates from *reference data*, i.e. images whose contents are known, to density estimates from *test data*, i.e. images whose contents are to be classified. The measure for this comparison is the Chi-Square divergence which is a second-order approximation to twice the KL divergence[15], i.e. $2D(\hat{f}_0\|\hat{f}_l)$, and is given as

$$\chi^2(\hat{f}_l, \hat{f}_0) = \int \frac{\left(\hat{f}_l(\boldsymbol{V}) - \hat{f}_0(\boldsymbol{V}) \right)^2}{\hat{f}_l(\boldsymbol{V})} \, d\boldsymbol{V} \tag{A.3}$$

for a reference density estimate $\hat{f}_l$ and test density estimate $\hat{f}_0$. This χ^2 divergence is computed between the density estimated from the test data and each of the densities estimated from the reference data. A classification decision is then made according to the rule

$$\hat{l} = \begin{cases} l^* \triangleq \arg\min_l \left\{ \chi^2(\hat{f}_l, \hat{f}_0) \right\} & : \text{if } \chi^2(\hat{f}_{l^*}, \hat{f}_0) \le T \\ \text{no classification} & : \text{if } \chi^2(\hat{f}_{l^*}, \hat{f}_0) > T \end{cases} \tag{A.4}$$

for a threshold T. Thus, the test data is declared to have arisen from the same source distribution as the reference data set that minimizes the χ^2 divergence as long as the resulting χ^2 divergence does not exceed the threshold T. If the minimum χ^2 divergence exceeds this threshold, then a "no classification" decision is made. The threshold thus imposes a minimal confidence requirement in order to make a classification.

---

[3]The use of the hyper-cube is no less general than the use of a hyper-rectangle as the kernel because the filters $f_{m,i}$ can be used to scale the data.

Exact computation of the integral in Eq. (A.3) is computationally intensive. DeBonet thus approximates the integral as

$$\chi^2(\hat{f}_l, \hat{f}_0) \approx \sum_{V \in \mathcal{S}_l} \frac{1}{\hat{f}_l(V)} \frac{\left(\hat{f}_l(V) - \hat{f}_0(V)\right)^2}{\hat{f}_l(V)}$$

$$= \sum_{V \in \mathcal{S}_l} \left(\frac{\hat{f}_l(V) - \hat{f}_0(V)}{\hat{f}_l(V)}\right)^2 . \tag{A.5}$$

In this approximation, DeBonet replaces the integral with a summation evaluated at the sample feature vectors in the reference data set $\mathcal{S}_l$. In order to account for their nonuniform distribution, the elements in the summation are weighted according to their local sparsity, i.e. by the reciprocal of the density estimate. Because the points used in the summation vary with the reference data set, the test density must be separately estimated for each reference data set to which it is compared.

Appendix B

MSTAR Data Set

Moving and Stationary Target Acquisition and Recognition (MSTAR) is a Defense Advanced Research Projects Agency (DARPA) program designed to explore and improve SAR ATR. To help achieve this goal, MSTAR released a large data set for public distribution. The MSTAR data set was collected using the Sandia National Laboratory SAR sensor platform operating at an X-band frequency. In particular, the radar had a center frequency of 9.6GHz and a bandwidth of 0.59GHz which produces a one foot down-range resolution. To similarly obtain a one foot cross-range resolution, a 2.8° synthetic aperture was used.

The MSTAR data set contains images of a variety of vehicles taken from various orientations. In particular, compact discs are distributed with the following data:

- *MSTAR Targets*: This data subset contains images of the BMP-2 tank (9563, 9566, c21)[1], BTR-70 transport (c71), T-72 tank(132, 812, s7), and SLICY[2] taken at 15° and 17° elevation angles with full aspect coverage, i.e. images formed every 1° in azimuth[3].

- *MSTAR/IU Mixed Targets*: This data subset contains images of the 2S1 gun (b01), BDRM-2 truck (E-71), BTR-60 transport (k10yt7532), D-7 bulldozer (92v13015), T-62 tank (A51), ZIL-131 truck (E12), ZSU-23/4 gun (d08), and SLICY taken at 15°, 17°, 30°, and 45° elevation angles.

- *MSTAR/IU T-72 Variants*: This data subset contains images of the T-72 tank (A04, A05, A07, A10, A32, A62, A63, A64) under various articulations taken at 15°, 17°, 30°, and 45° elevation angles.

- *MSTAR Clutter*: This data subset contains images of clutter from rural and urban areas of Huntsville, Alabama taken at a 15° elevation angle.

The work in this thesis only utilizes the 15° and 17° data in the "Targets" and "Mixed Targets" data subsets. The 17° data are generally used for the learning aspects of the

---

[1]Th serial numbers of the imaged vehicles are specified in parentheses.

[2]Sandia Labs Implementation of CYlinders (SLICY) is an artificial target composed of a few simple canonical scatterers constructed to aid the study of scatterer phenomenology and ATR.

[3]There are some azimuths omitted in the data set, thus there are slightly fewer than 360 images.

classification process because the dense uniform sampling in azimuth (every 1°) allows for all possible azimuthal dependencies to be observed. The 15° data are then used as the test data for performance evaluation.

Bibliography

[1] "MSTAR (Public) Targets: T-72, BMP-2, BTR-70, SLICY", see
http://www.mbvlab.wpafb.af.mil/public/MBVDATA.

[2] M. Allen and L. Hoff. Wide-angle Wideband SAR Matched Filter Image Formation
for Enhanced Detection Performance. In *Proc. of the SPIE, Algorithms for SAR
Imagery*, volume 2230, pages 302–314, Apr. 1994.

[3] S. Barbarosa, L. Marsili, and G. Mungari. SAR Super-Resolution Imaging by Signal
Subspace Projection Techniques. *AEU Int. J. Electron. Commun.*, 50(2):133–138,
1996.

[4] G. Benitz. Adaptive High-Definition Imaging. In *Proc. of the SPIE, Algorithms
for SAR Imagery*, volume 2230, pages 106–119, Apr. 1994.

[5] G. Benitz. High-Definition Vector Imaging. *Lincoln Lab. Journal*, 10(2):147–169,
1997.

[6] G. Benitz. High-Definition Vector Imaging for Synthetic Aperture Radar. In *31st
Asilomar Conference on Signals, Systems, & Computers*, volume 2, pages 1204–
1209, Nov. 1997.

[7] G. Benitz and D. DeLong. Extensions of High Definition Imaging. In *Proc. of the
SPIE, Algorithms for SAR Imagery*, volume 2487, pages 165–180, Apr. 1995.

[8] B. Bhanu. Automatic Target Recognition: State of the Art Survey. *IEEE Trans.
AES*, 22(4):364–379, July 1986.

[9] B. Bhanu, D. E. Dudgeon, E. G. Zelnio, A. Rosenfeld, D. Casasent, and I. S. Reed.
Guest Editorial Introduction to the Special Issue on Automatic Target Detection
and Recognition. *IEEE Trans. IP*, 6(1):1–6, January 1997.

[10] M. L. Bryant and F. D. Garber. SVM Classifier Applied to the MSTAR Public
Data Set. In *Algorithms for Synthetic Aperture Radar Imagery VI*, volume 3721
of *Proceedings of SPIE*, pages 355–360, Orlando, FL, April 1999.

[11] J. Capon. High-Resolution Frequency-Wavenumber Spectrum Analysis. In *Proc. IEEE*, volume 57(8), pages 1408–1418, Aug. 1969.

[12] R. Chaney, A. Willsky, and L. Novak. Coherent Aspect-Dependent SAR Image Formation. In *Proc. of the SPIE, Algorithms for SAR Imagery*, volume 2230, pages 256–274, Apr. 1994.

[13] H. Chiang and R. Moses. ATR Performance Prediction Using Attributed Scattering Features. In *Proc. of the SPIE, Algorithms for SAR Imagery*, volume 3721, pages 785–796, Apr. 1999.

[14] A. Cohen and J. Kovacevic. Wavelets: The Mathematical Background. *Proceedings of the IEEE*, 84(4):514–522, Apr. 1996.

[15] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc., New York, 1991.

[16] H. Cox and R. Zeskin. Reduced Variance Distortionless Response (RVDR) Performance with Signal Mismatch. In *25st Asilomar Conference on Signals, Systems, & Computers*, volume 2, pages 825–829, 1992.

[17] J. Curlander and R. McDonough. *Synthetic Aperture Radar*. John Wiley and Sons, Inc., New York, 1991.

[18] I. Daubechies. Ten Lectures on Wavelets, 1992. SIAM Press.

[19] J. DeBonet. Multiresolution Sampling Procedure for Analysis and Synthesis of Texture Images. In *ACM SIGGRAPH*, 1997. in Computer Graphics.

[20] J. DeBonet, P. Viola, and J. Fisher. Flexible Histograms: A Multiresolution Target Discrimination Model. In *Proc. of the SPIE, Algorithms for SAR Imagery*, Apr. 1998.

[21] S. DeGraaf. SAR Imaging via Modern 2-D Spectral Estimation Methods. *IEEE Trans. on Image Proc.*, 7(5):729–761, May 1998.

[22] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons, 1973.

[23] D. E. Dudgeon and R. T. Lacoss. An Overview of Automatic Target Recognition. *Lincoln Laboratory Journal*, 6(1):3–10, Spring 1993.

[24] C. Fosgate, H. Krim, W. Irving, W. Karl, and A. Willsky. Multiscale Segmentation and Anomaly Enhancement of SAR Imagery. *IEEE Trans. on Image Proc.*, 6(1):7–20, Jan. 1997.

[25] N. S. Friedland. Feature Characterization in the Context of Model-Based Search. In *Algorithms for Synthetic Aperture Radar Imagery III*, volume 2757 of *Proceedings of SPIE*, pages 253–270, Orlando, FL, April 1996.

[26] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, Inc., Boston, 1990.

[27] A. Gersho and R. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

[28] W. Härdle. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge [England], 1989.

[29] W. Irving, P. Fieguth, and A. Willsky. An Overlapping Tree Approach to Multiscale Stochastic Modeling and Estimation. *IEEE Transactions on Image Processing*, 6(11), November 1997.

[30] W. Irving, L. Novak, and A. Willsky. A Multiresolution Approach to Discriminating Targets from Clutter in SAR Imagery. *IEEE Transactions on Aerospace and Electronic Systems*, 33(4):1157–1169, October 1997.

[31] W. W. Irving. Multiresolution Approach to Discriminating Targets From Clutter SAR data acquiredin SAR Imagery. In *SPIE Symposium*, Apr. 1995.

[32] H. Joe. Estimation of Entropy and Other Functionals of a Multivariate Density. *Ann. Inst. Statist. Math*, 41(4):683–697, 1989.

[33] J. B. Keller. Geometrical Theory of Diffraction. *Journal of the Optical Society of America*, 52:116–130, Feb. 1962.

[34] E. R. Keydel and S. W. Lee. Signature Prediction for Model-Based Automatic Target Recognition. In *Algorithms for Synthetic Aperture Radar Imagery III*, volume 2757 of *Proceedings of SPIE*, pages 306–317, Orlando, FL, April 1996.

[35] A. Kim, S. Dogan, J. Fisher, R. Moses, and A. Willsky. Attributing Scatterer Anisotropy for Model Based ATR. In *Proc. of the SPIE, Algorithms for SAR Imagery*, volume 4053, Apr. 2000.

[36] A. Kim, J. Fisher, A. Willsky, and P. Viola. Nonparametric Estimation of Aspect Dependence for ATR. In *Proc. of the SPIE, Algorithms for SAR Imagery*, volume 3721, pages 332–342, Apr. 1999.

[37] H. Krim and M. Viberg. Two Decades of Array Signal Processing Research. *IEEE Signal Processing Magazine*, pages 67–92, July 1996.

[38] G. Leung and J. H. Shapiro. Toward a Fundamental Understanding of Multiresolution SAR Signatures. In *Algorithms for Synthetic Aperture Radar Imagery IV*, volume 3070 of *Proceedings of SPIE*, Orlando, FL, April 1997.

[39] J. Li and P. Stoica. An Adaptive Filtering Approach to Spectral Estimation and SAR Imaging. *IEEE Trans. on Signal Proc.*, 44(6):1469–1484, June 1996.

[40] M. McClure and L. Carin. Matching Pursuits with a Wave-Based Dictionary. *IEEE Trans. on Sig. Proc.*, 45(12):2912–2927, Dec. 1997.

[41] Monzingo and Miller. *Introduction to Adaptive Arrays*. John Wiley and Sons, 1980.

[42] D. Munson. An Introduction To Strip-Mapping Synthetic Aperture Radar. In *Proc. Int. Conf. Acoust., Speech, and Signal Proc.*, pages 2245–2248, Apr. 1987.

[43] D. Munson and R. Visentin. A Signal Processing View of Strip-Mapping Synthetic Aperture Radar. *IEEE Trans. on Acoustics, Speech, and Sig. Proc.*, 37(12):2131–2146, Dec. 1989.

[44] G. J. Owirka, S. M. Verbout, and L. M. Novak. Template-based SAR ATR Performance Using Different Image Enhancement Techniques. In *Algorithms for Synthetic Aperture Radar Imagery VI*, volume 3721 of *Proceedings of SPIE*, pages 302–319, Orlando, FL, April 1999.

[45] L. Potter and R. Moses. Attributed Scattering Centers for SAR ATR. *IEEE Trans. on Image Proc.*, 5(1):79–91, Jan. 1997.

[46] J. A. Richards. *Target Model Generation from Multiple Synthetic Aperture Radar Images*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, May 2001.

[47] S. Ross. *A First Course in Probability*. MacMillan, New York, 1988.

[48] G. T. Ruck, D. E. Barrick, W. D. Stuart, and C. K. Krichbaum. *Radar Cross Section Handbook*. Plenum Press, New York, 1970.

[49] R. O. Schmidt. Multiple Emitter Location and Signal Parameter Estimation. *IEEE Trans. AP*, 34:276–280, 1994.

[50] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

[51] W. Smith, T. Irons, J. Riordan, and S. Sayre. Peak stability derived from phase history in synthetic aperture radar. In *Proc. of the SPIE, Algorithms for SAR Imagery*, volume 3721, pages 450–461, Apr. 1999.

[52] M. Soumekh. *Fourier Array Imaging*. Prentice Hall, Englewood Cliffs, NJ, 1994.

[53] N. Subotic, B. Thelen, J. Gorman, and M. Reiley. Multiresolution Detection of Coherent Radar Targets. *IEEE Trans. on Image Proc.*, 6(1):21–34, Jan. 1997.

[54] V. Vapnik. *Statistical Learning Theory*. Wiley, 1998. New York.

[55] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer, 2000. New York.

[56] V. Velten, T. Ross, J. Mossing, S. Worrell, and M. Bryant. Standard SAR ATR Evaluation Experiments using the MSTAR Public Release Data Set. Technical report, Model Based Vision Lab, 1998.

[57] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding.* Prentice Hall, Englewood Cliffs, NJ, 1995.

[58] D. Wehner. *High Resolution Radar.* Artech House, Norwood, MA, 1987.

[59] M. Wickerhauser. Lectures on Wavelet Packet Algorithms, Nov. 1991.

[60] J. Wissinger, R. B. Washburn, N. S. Friedland, A. Nowicki, D. R. Morgan, C. Chong, and R. Fung. Search Algorithms for Model-Based SAR ATR. In *Algorithms for Synthetic Aperture Radar Imagery III*, volume 2757 of *Proceedings of SPIE*, pages 279–293, Orlando, FL, April 1996.

[61] C. P. Yeang and J. H. Shapiro. Target Detection Theory for Stripmap SAR Using Physics-Based Multiresolution Signatures. In *Algorithms for Synthetic Aperture Radar Imagery V*, volume 3370 of *Proceedings of SPIE*, pages 646–660, Orlando, FL, April 1998.