AFRL-SN-WP-TR-2001-1100

A UNIFIED MULTIRESOLUTION FRAMEWORK FOR AUTOMATIC TARGET RECOGNITION (ATR)

JOHN FISHER
ERIC GRIMSON
ALAN WILLSKY

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
200 TECHNOLOGY SQUARE
CAMBRIDGE, MA 02139

SEPTEMBER 2001

FINAL REPORT FOR PERIOD OF 01 AUGUST 1997 – 31 DECEMBER 2000

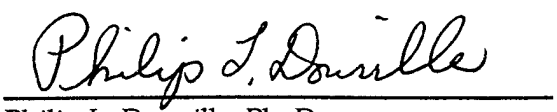
SENSORS DIRECTORATE
AIR FORCE RESEARCH LABORATORY
AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7318

NOTICE

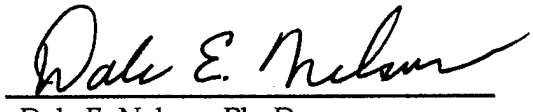
USING GOVERNMENT DRAWINGS, SPECIFICATIONS, OR OTHER DATA INCLUDED IN THIS DOCUMENT FOR ANY PURPOSE OTHER THAN GOVERNMENT PROCUREMENT DOES NOT IN ANY WAY OBLIGATE THE US GOVERNMENT. THE FACT THAT THE GOVERNMENT FORMULATED OR SUPPLIED THE DRAWINGS, SPECIFICATIONS, OR OTHER DATA DOES NOT LICENSE THE HOLDER OR ANY OTHER PERSON OR CORPORATION; OR CONVEY ANY RIGHTS OR PERMISSION TO MANUFACTURE, USE, OR SELL ANY PATENTED INVENTION THAT MAY RELATE TO THEM.

THIS REPORT IS RELEASABLE TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS). AT NTIS, IT WILL BE AVAILABLE TO THE GENERAL PUBLIC, INCLUDING FOREIGN NATIONS.

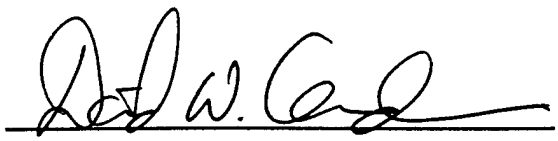
THIS TECHNICAL REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION.



Philip L. Douville, Ph. D.
Project Engineer
Target Recognition Branch



Dale E. Nelson, Ph. D.
Chief, Target Recognition Branch
Sensor ATR Technology Division
Sensors Directorate



Do not return copies of this report unless contractual obligations or notice on a specific document requires its return.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
September 2001	Final	08/01/1997 – 12/31/2000

4. TITLE AND SUBTITLE
A UNIFIED MULTIRESOLUTION FRAMEWORK FOR AUTOMATIC TARGET RECOGNITION (ATR)

5a. CONTRACT NUMBER
F33615-97-1-1014

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62301

6. AUTHOR(S)
JOHN FISHER
ERIC GRIMSON
ALAN WILLSKY

5d. PROJECT NUMBER
ARPA

5e. TASK NUMBER
AA

5f. WORK UNIT NUMBER
1D

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
200 TECHNOLOGY SQUARE
CAMBRIDGE, MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
SENSORS DIRECTORATE
AIR FORCE RESEARCH LABORATORY
AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7318

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
AFRL/SNAT

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-SN-WP-TR-2001-1100

12 DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14 ABSTRACT *(Maximum 200 Words)*
This report describes the development of multiscale, multiresolution methods for automatic target recognition (ATR). The methods are applied to and developed specifically for synthetic aperture radar data. Applications include high level reasoning over learned target models, information theoretic approaches for pose estimation, and model development from multiple views.

15. SUBJECT TERMS
Automatic Target Recognition, Synthetic Aperture Radar, Multiresolution, Multiscale, Information Theory

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	b. ABSTRACT	c. THIS PAGE			Philip Douville
Unclassified	Unclassified	Unclassified	SAR	128	19b. TELEPHONE NUMBER *(Include Area Code)* (937) 255-1115 x4378

Contents

List of Figures

List of Tables

1 Summary

In this final report we summarize our accomplishments in the research program supported by Grant F33615-97-1-1014 over the period from August 1, 1997 to December 31, 2000. The basic objective of this research program was to pursue an integrated set of research problems associated with automatic target detection and recognition (ATD/R) using multiscale, multiresolution, and/or multiaspect approaches.

One of our major successes to date in prior work was in developing multiresolution stochastic models for SAR imagery that accurately capture the scale-to-scale statistical variability of speckle in SAR imagery. In one application, we used models for natural clutter and for man-made objects together with our fast statistical likelihood calculation methods to develop an enhanced discrimination feature that, when integrated into Lincoln Labs' ATR algorithm and tested on a very large data set, resulted in a factor of 6 reduction in false alarms over the previous best results. In the second application, we used our models to segment natural clutter (trees and grass) and to enhance anomalous pixels (due to man-made scatterers) that did not produce the scale-to-scale variability consistent with natural clutter. The results are very accurate segmentations and enhanced visibility of anomalies as compared to widely used constant false alarm rate (CFAR) methods.

There are many additional applications for these multiscale models. For example, anomalies that result from man-made objects exhibit themselves as distinctive patterns across scale that differ significantly from the scale-to-scale textural variations. Consequently, chains of pixels across scale could in principle be viewed as robust and statistically meaningful features that can be further exploited for model-based recognition. In this work we turn our attention to the application and further development of multiscale features for higher level recognition and reasoning. Classically, target models include geometrical constraints on the appearance of features in space. In this new framework, models will also include information about the appearance of features across scale. The development of such models is a central objective of this project. Once we have such models, we can use our statistically optimal methods for evaluating likelihoods to evaluate match scores for hypothesized models and poses.

Towards that end we integrate the following methodologies (related sections in paranthesis) in developing a unified multiscale statistical approach to SAR signal processing:

multiscale modeling	Sections 2, 5
multiaspect, multiresolution modeling	Sections 3, 4, 5, 6, 7, 8
nonparametric statistics	Sections 2, 3, 5, 6, 7, 8
applied information theory	Sections 5, 6

We differentiate multiscale and multiresolution to indicate those methods which focus on modeling the log-magnitude image data versus the phase history, respectively, while multiaspect refers to those techniques which incorporate images acquired at disparate viewing angles. However, all of these methods can be viewed as statistical methods applied to the SAR view sphere. The application of these methodologies encompassed both the modeling of single scattering centers as well as objects (described by collections of these scatterers). Nonparametric statistics and information theory provide a richer means of modeling the complex uncertainties that arise when modeling SAR scattering at the distributed object level.

The principal investigators for this effort were Professors W. Eric Grimson, P. Viola, and Alan S. Willsky. Drs. J. Fisher and W. M. Wells served as senior investigators for this project, and a considerable number of graduate research assistants and additional thesis students not requiring stipend or tuition support participated in this effort. Furthermore, in the course of our research, we established working relationships with several organizations with long histories of involvement in ATD/R. In particular, we have interacted closely with several groups of engineers and scientists at Lincoln Laboratory, including researchers directly

involved in Defense Advanced Research Projects Agency (DARPA)-sponsored ATD/R programs. We have also interacted with engineers at Environmental Research Institute of Michigan (ERIM) and have several collaborations with Alphatech, Inc., which is involved in several ATD/R projects under DARPA support, including a Phase II Small Business Innovative Research (SBIR) program, a project under the joint Industry/University initiative in ATD/R, and two contracts in the MSTAR program. We believe that we have had considerable success in our research. Moreover, our work has also led to several significant transitions which have helped to establish and fuel the industrial interactions mentioned previously. In addition, a major element of the success that we have had can be seen in the continuing vitality of our efforts in this area. Indeed both our university research and our interactions with Lincoln, Alphatech, and ERIM have allowed us to identify research directions and areas for collaboration and transition that hold considerable promise for the some time to come. In the body of this report we provide a brief summary of the major elements of our research program all of which highlighted in detail in subsequent sections of this report which also include additional references. We also note that there are several manuscripts arising from this research currently under peer review for publication in various scholarly journals.

1.1 Research Description

In this section we briefly describe the major elements of the research in which we have been engaged during the tenure of this grant. We limit ourselves here to a succinct summary of these results and refer to the sections contained in the report for detailed developments.

1.1.1 Nonparamtric Multiscale Target/Clutter Models

This work, led by Professors Grimson and Viola and graduate student Jeremy De Bonet is a method for modeling SAR imagery as a multiscale random process. The method successfully captures aspects of both the structure and variability of SAR imagery exhibited in the log-magnitude domain. Section 2 details a SAR vehicle classifier based on this approach, which is applied to the public release MSTAR database and compared to the Wright-Patterson Air Force Base (WPAFB) baseline template classifier. Section 5 details an application to SAR scene registration for multiple aspect views. Note that the method described in section 5 includes a technique for automatic tie-point selection which could easily be adapted to the general problem of anomaly detection.

1.1.2 Attribution of Anisotropic Scattering Centers

This work, led by Professor Willsky, Dr. Fisher, and graduate student Andrew Kim, addresses the shortcoming of standard SAR imaging in that energy from scattering centers is assumed to occupy the full aperture. The well known phenomenon of speckle in SAR imagery belies to fault in this assumption. This work, detailed in sections 3 and 3, establishes a generic model of unimodal scattering derived from a multiresolution (multiaperture) representation of pixel reflectivities. Section 3 describes the initial approach as an extension of the multiscale models described in section 2, while section 4 further develops the model as well incorporating it into the Ohio State University matching engine (also supported under the same research program).

1.1.3 SAR Vehicle Model Building from Multiple Aspect Imagery

This work, led by Professor Willsky, Dr. Fisher, and graduate student John Richards (now working for Sandia National Laboratory) develops an approach for generating a vehicle scattering center models from

multiple aspect views. This technique is of particular interest for denied-target modeling or refining extant cad models. Section 7 describes the initial approach in the expectation-maximization framework while section 8 extends the approach to building target models.

1.1.4 Information Theoretic Approaches to SAR Scene Analysis

This work, led by Dr. John Fisher, looks at an information theoretic approach for extracting statistically relevant features for estimation. Section 6 describes the approach and demonstrates the efficacy for pose estimation of vehicles within SAR imagery. This work is currently being transitioned to Alphatech, Inc. to be used in a DOD-sponsored feature-aided tracking program.

2 Flexible Histograms: A Multiresolution Target Discrimination Model

In previous work we have developed a methodology for texture recognition and synthesis that estimates and exploits the dependencies across scale that occur within images[1, 2]. In this paper we discuss the application of this technique to synthetic aperture radar (SAR) vehicle classification. Our approach measures characteristic cross-scale dependencies in training imagery; targets are recognized when these characteristic dependencies are detected. We present classification results over a large public database containing SAR images of vehicles. Classification performance is compared to the Wright Patterson baseline classifier [3]. These preliminary experiments indicate that this approach has sufficient discrimination power to perform target detection/classification in SAR.

2.1 Introduction

The detection and classification of targets in imagery is a difficult problem. Any successful algorithm must be able to correctly classify the very wide range of images generated by a single target class. The sources of image variations in SAR are many: target rotation, translation, articulation, sensor noise, depression angle, overlay, camouflage, and many others. Typically classifiers attempt to deal with these variations in one of two ways: invariance and duplication. For example, translation invariance could be achieved by solely measuring the distance between detected scatters, since distance is invariant to translation. However, such a method does not work as well as the target rotates, since scatters may appear and then disappear. Alternately, a duplicative technique could be used at the cost of additional computation – by using several separate models for a single target class, each tuned to a different target orientation.

In this paper we compare a nonparametric multiscale approachwith a baseline classifier proposed and implemented by a group at WPAFB [3]. Their intent was to provide a reasonable standard for comparison of new SAR target recognition approaches. Their approach is duplicative: each target class is modeled by a number distinct templates each tuned to a different orientation. Targets are recognized if one of the target templates is nearby in the vector space defined by the cells of a SAR image.[1] The history of templates in recognition is extensive (see Duda and Hart[4] for an early review). Measuring the distance to a template is equivalent to measuring the negative log likelihood of a Gaussian process whose mean value is the template. Since each target is modeled as a collection of templates, the model is essentially a mixture of Gaussians. This is the optimal classification approach when we believe that rotation has been sampled finely enough, and that the other variations in SAR imagery can be modeled as a Gaussian noise. Over many years, template matching has proven to be a very effective technique. It is not surprising that the template-based classifier yields very good performance. However, it may be infeasible in practice to acquire the large number of example images needed.

In this paper we present an alternative statistical model for the variations in SAR imagery. Models are developed from far fewer target vehicle examples than are needed by template techniques. We show that it effectively classifies targets in the presence of a difficult set of confusion vehicles. These results have been quantified and compared with the performance of our reimplementation of the WPAFB baseline classifier.

[1] we refer to the amplitude of the complex signal in a resolution cell as a *cell*.

2.2 Multiresolution Models of Targets

Recently it has been shown that multiresolution models of statistical processes can be effective for target detection in SAR data [5, 6, 7]. These model have decomposed the SAR signal into a number of different images at different scales from coarse-to-fine. Irving *et al.* define a coarse to fine statistical process in which the distribution of higher frequency data is conditioned on lower frequency data. This conditional distribution is modeled as a Gaussian. Using these conditional distributions a multiscale Markov chain is defined which can model complex dependencies both across scale and space. A target detection algorithm is constructed from a pair of such distributions, one for targets and one for clutter. Each statistical model is trained from example data: a set of chips containing targets and another set of chips containing only clutter. From these chips, multiscale representations are computed, cross-scale conditional distributions are observed, and the parameters of the cross-scale Gaussian distribution are estimated.

In many ways our classification approach is a generalization of these multiscale detection algorithms, with three key differences: i) our multiresolution representation separates and distinguishes information represented at different orientations; ii) we consider full (non-Markovian) cross-scale conditional distributions; and iii) our cross-scale model is nonparametric and as a result, can model non-Gaussian processes.

2.2.1 A Coarse-to-Fine Generative Process

As in previous work[6, 7], the generative model that underlies our statistical classification approach is a conditional probability distribution across scale. This probability chain defines a statistical distribution for the entire SAR signal (including information at every frequency). This is accomplished by modeling the dependency of the highest frequencies on lower frequencies.

Unlike previous approaches, we define the conditional chain on a multiscale wavelet transformation. The wavelet transform is most efficiently computed as an iterative convolution using a bank of filters. First a pyramid of low frequency downsampled images is created: $G_0 = I$, $G_1 = 2\downarrow(g \otimes G_0)$, and $G_{i+1} = 2\downarrow(g \otimes G_i)$, where $2\downarrow$ downsamples an image by a factor of 2 in each dimension, $\otimes$ is the convolution operation, and g is a low pass filter. At each level, a series of filter functions is applied: $F_j^i = f_i \otimes G_j$, where the f_i's are various types of filters (see Figure 1). Computation of the F_j^i's is a linear transformation that can thought of as a single matrix W. With careful selection of g and f_i (as done in [8]) this matrix can be constructed so that $W^{-1} = W^\top$. Computation of the inverse wavelet transform is algorithmically similar to the computation of the forward wavelet transform.

For every resolution cell in an SAR data we define the *parent vector*:

$$\begin{aligned}\vec{V}(x,y) = \; & \left[F_0^0(x,y), F_0^1(x,y), \cdots, F_0^M(x,y),\right. \\ & F_1^0\left(\left\lfloor\tfrac{x}{2}\right\rfloor, \left\lfloor\tfrac{y}{2}\right\rfloor\right), F_1^1\left(\left\lfloor\tfrac{x}{2}\right\rfloor, \left\lfloor\tfrac{y}{2}\right\rfloor\right), \ldots, F_1^M\left(\left\lfloor\tfrac{x}{2}\right\rfloor, \left\lfloor\tfrac{y}{2}\right\rfloor\right), \ldots, \\ & \left. F_N^0\left(\left\lfloor\tfrac{x}{2^N}\right\rfloor, \left\lfloor\tfrac{y}{2^N}\right\rfloor\right), F_N^1\left(\left\lfloor\tfrac{x}{2^N}\right\rfloor, \left\lfloor\tfrac{y}{2^N}\right\rfloor\right), \ldots, F_N^M\left(\left\lfloor\tfrac{x}{2^N}\right\rfloor, \left\lfloor\tfrac{y}{2^N}\right\rfloor\right)\right]\end{aligned} \tag{1}$$

where N is the top level of the pyramid and M is the number of features. The parent vector can be visualized schematically as in Figure 2, in which the image is decomposed from coarsest (on the left) to finest resolution (on the right). Each element in these arrays represents the collected feature responses at that location and resolution.

We define a probability distribution over SAR images as a multiresolution wavelet tree across scale. In

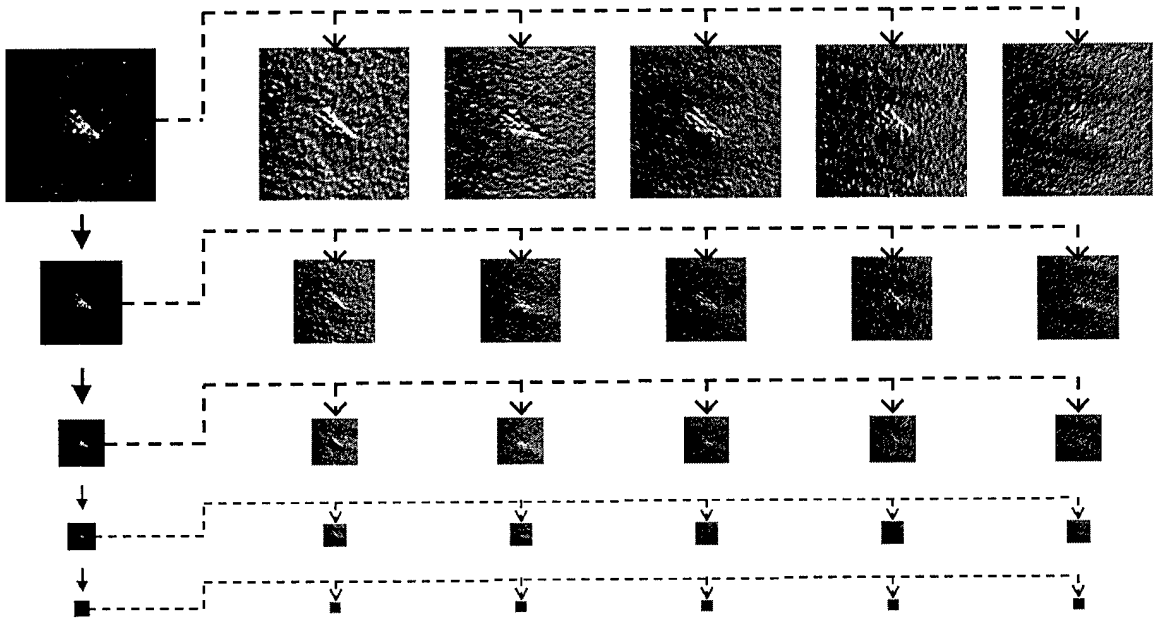


Figure 1: Wavelet Decomposition of a SAR Target Chip. In the upper left are the original SAR amplitudes.

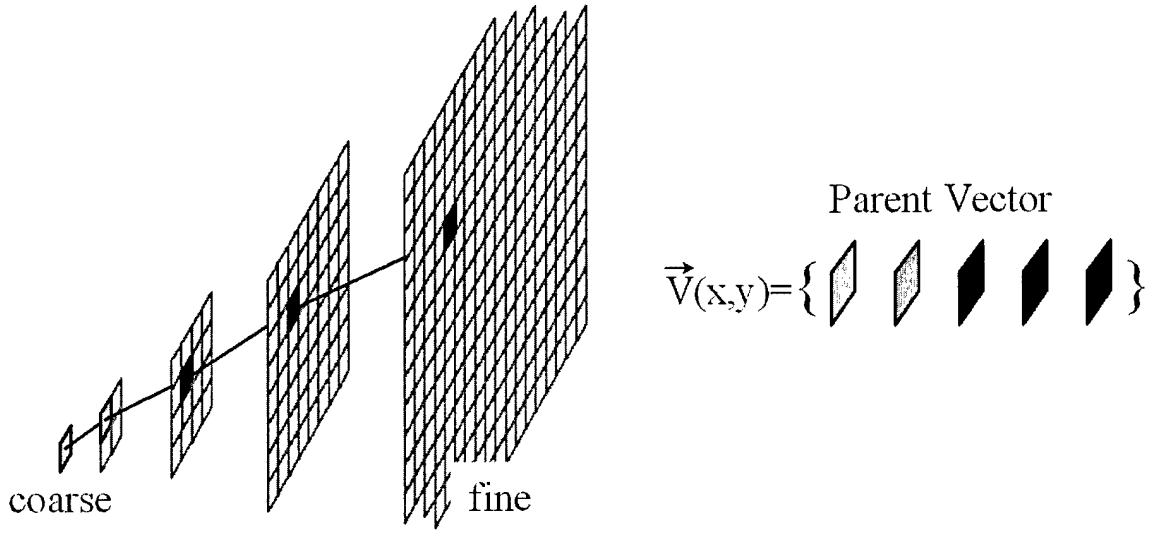


Figure 2: Parent Vector. Each cell in these arrays represents the collection feature responses at that location and resolution.

this tree the generation of the lower levels depend on the higher levels:

$$\begin{aligned} p(\vec{V}(x,y)) = \; & p(\vec{V}_M(x,y)) \times p(\vec{V}_{M-1}(x,y)|\vec{V}_M(x,y)) \\ & \times p(\vec{V}_{M-2}(x,y)|\vec{V}_{M-1}(x,y),\vec{V}_M(x,y)) \\ & \times \ldots \\ & \times p(\vec{V}_0(x,y)|\vec{V}_1(x,y),\; \ldots, \vec{V}_{M-1}(x,y),\vec{V}_M(x,y)) \end{aligned} \tag{2}$$

where $\vec{V}_l(x,y)$ is the a subset of the elements of $\vec{V}(x,y)$ computed from G_l. Usually we will assume ergodicity, i.e. that $p(\vec{V}(x,y))$ is independent of x and y. The generative process starts from the top of the pyramid, choosing values for the $\vec{V}_M(x,y)$ at all points. Once these are generated the values at the next level, $\vec{V}_{M-1}(x,y)$, are generated. The process continues until all of the wavelet coefficients are generated. Finally the image is computed using the inverse wavelet transform.

The probabilistic model is not made up of a collection of independent chains, one for each $\vec{V}(x,y)$, rather parent vectors for neighboring cells have substantial overlap as coefficients in the higher pyramid levels (which are lower resolution) are shared by neighboring cells at lower pyramid levels. Thus, the generation of nearby cells will be strongly dependent.

2.2.2 Estimating the Conditional Distributions

The conditional distributions in equation (2) must be estimated from observations. We choose to do this directly from the data in a nonparametric fashion. Given a sample of parent vectors $\left\{\vec{S}(x,y)\right\}$ from an example image, we estimate the conditional distribution as a ratio of Parzen window density estimators:

$$p(\vec{V}_l(x,y)|\vec{V}_{l+1}^M(x,y)) = \frac{p(\vec{V}_l^M(x,y))}{p(\vec{V}_{l+1}^M(x,y))} \approx \frac{\sum_{x',y'} R(\vec{V}_l^M(x,y),\vec{S}_l^M(x',y'))}{\sum_{x',y'} R(\vec{V}_{l+1}^M(x,y),\vec{S}_{l+1}^M(x',y'))} \tag{3}$$

where $\vec{V}_l^k(x,y)$ is a subset of the parent vector $\vec{V}(x,y)$ that contains information from level l to level k, and $R(\cdot)$ is a function of two vectors that returns maximal values when the vectors are similar and smaller values when the vectors are dissimilar. We have explored various $R(\cdot)$ functions. In the results presented the $R(\cdot)$ function returns a fixed constant if all of the coefficients of the vectors are within some threshold θ and zero otherwise.

The cross-scale conditional distributions defined in equation (3) can be used to define the distribution of parent vectors in a target image. This directed conditional structure is very useful when one wishes to sample from the distribution – no computationally expensive operations such as Gibbs sampling are necessary (this is in distinct contrast to Markov Random Fields[9] or in the FRAME method[10]). Given this simple definition for $R(\cdot)$ sampling from $p(\vec{V}_l(x,y)|\vec{V}_{l+1}^M(x,y))$ is very straightforward: find all x',y' such that $R(\vec{S}_{l+1}^M(x',y'),\vec{S}_{l+1}^M(x,y))$ is non-zero and pick from among them to set $\vec{V}_l(x,y) = \vec{S}_l(x',y')$.

In previous work we have demonstrated that images generated by sampling from the above distribution are of very high quality[1, 11]. By synthesizing new images from models built from SAR imagery, we can visualize the ability of this model to capture the complex structural patterns inherent in SAR. An image synthesized by sampling from this distribution is shown in Figure 3.

2.3 Target Classification

Given a particular distribution over parent vectors, it might at first seem as though classification of a target chip could be best accomplished by computing the likelihood that that chip was generated by each of the

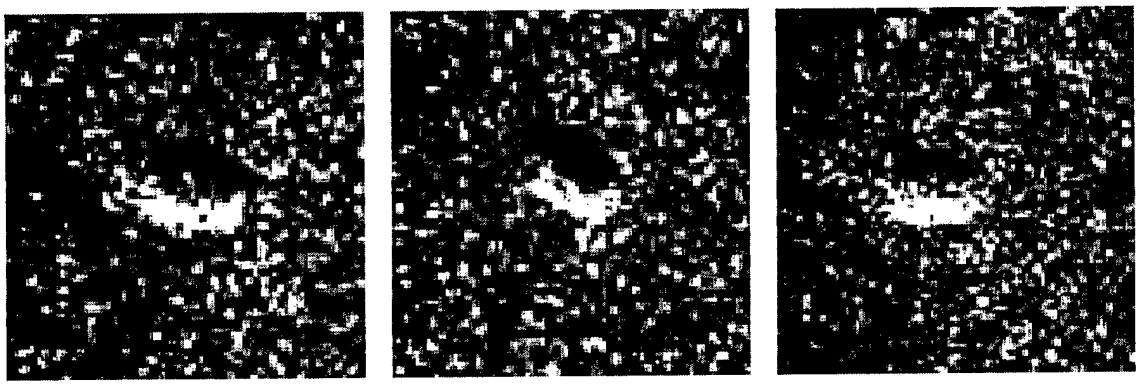


Figure 3: Target chips synthesized by sampling directly from the probabilistic distribution generated by the flexible histogram model generated from several example vehicles.

multiscale target models. If target chip is viewed as a single sample from a vector valued random process, this is the optimal Bayesian classification scheme. However, each target chip is actually a collection of *many samples* from the ergodic process characterized by the distribution. The distinction between these two views is subtle, and is crucial for obtaining a powerful classification measure.

Here is a simple example to illustrate this: consider a process which creates random binary images by repeatedly flipping a single coin. Each cell is colored white with an independent probability of 0.75. Suppose you were asked to decide which of two images is more likely to have been generated by this process: the first has 75 white cells and 25 black ones; the second image has 100 white cells (and 0 black). Intuitively, it seems more likely that the first image was generated by this process. But if we treat each image as single sample from a joint distribution, the probabilities indicate otherwise. The probability of generating the first image is much lower than that of the all white image (roughly 3×10^{-25} compared to roughly 3×10^{-13}). Why does this approach fail to pick out the correct image? It does not take into account that the overwhelming majority of samples which are generated by this process will have about 75 white cells. While it is true that the first image is more likely than the second, it is much less *typical*. Formally, typical images are those whose entropy is approximately the same as the entropy of the class to which they belong. The fact that most images are typical is known as the Asymptotic Equipartition Property [12].

A better way to decide which of these two images was generated by the above process is to measure which is more typical. This is done using the Kullback-Liebler (KL) divergence or *cross-entropy*[2]. The cross-entropy is a measure of the difference between two distributions:

$$D(p||q) = \int p(\vec{V}(x,y)) \log \frac{p(\vec{V}(x,y))}{q(\vec{V}(x,y))}\, dx\, dy \tag{4}$$

$$= \int p(\vec{V}(x,y)) \log p(\vec{V}(x,y)) - \int p(\vec{V}(x,y)) \log q(\vec{V}(x,y))\, dx\, dy \tag{5}$$

$$= -H(p) - \int p(\vec{V}(x,y)) \log q(\vec{V}(x,y))\, dx\, dy \tag{6}$$

[2]Cross entropy is not symmetric and is therefore not a metric.

It can be viewed as the difference between two expected log likelihoods: the log likelihood of samples of $p(x)$ under the distribution $p(x)$, and the log likelihood of samples of $p(x)$ under $q(x)$. For the first image we estimate the probability of white cells to be $p_1 = 0.75$. For the second image set we estimate $p_2 = 1.0$. The true probability of a white cell is $q = 0.75$. For the first image, the cross-entropy is $D(p_1||q) = 0.28$, while the second is $D(p_2||q) = 0.0$, a perfect fit.

We have implemented a cross-entropy discrimination measure, and have tested it both on optical images and SAR data[2]. Here we present a discrete version which is far more efficient computationally.

2.4 Flexible Histograms: Efficient Divergence Approximation

We approximate the distribution in equation (2) by the Parzen density estimator:

$$p(\vec{V}(x,y)) = \sum_{x',y'} R\left(\vec{V}(x,y), \vec{V}(x',y')\right) \tag{7}$$

where $R(\cdot)$ is the boxcar Parzen density kernel defined by:

$$R(\cdot) = \Theta\left(T - \left\| \left[\vec{V}(x,y) - \vec{V}(x',y')\right](D)^{-1} \right\|_{\infty}\right) \tag{8}$$

where each component $d_{i,i}$ of diagonal matrix D scales the corresponding feature response; and $\Theta(\cdot) = 1$ if its argument is greater than 0, and is 0 otherwise; Z is a normalization factor. The L_∞ norm ($||\cdot||_\infty$) is equal to the largest difference along any dimension. The value of $p(\vec{V})$ is proportional to the count of parent vectors which fall within a $N \times M$ dimensional hypercube centered at $\vec{V}(I,x,y)$. We can define a histogram whose bins are given by $B(x,y) = p\left(\vec{V}(x,y)\right)$. We call this representation a flexible histogram because the centers of the bins are determined by the the parent vectors that appear in the training data. In this way, when building models for different targets the bins used in the histogramming process are specialized. The process of construction is outlined schematically in Figure 4.

By using the boxcar Parzen density kernel in equation (8), a significant computational advantage is obtained. Because the parent vectors for neighboring cells share coefficients at lower resolutions, the computation of $p\left(\vec{V}(x,y)\right)$ is not independent for regions neighboring (x,y). Thus, if a parent vector fails to fall into the histogram bin because some feature at a particular resolution is beyond the threshold distance, then all parent vectors which share that coefficient can be eliminated from consideration. In practice this reduces the complexity of the algorithm from $O(N^2)$ in the number of cells to an average case complexity of $\Omega(N)$.

The cross-entropy between two distributions can be effectively approximated by a chi-squared (χ^2) test on the bin counts in the flexible histogram (the χ^2 measure is in fact a two term Taylor series expansion of the cross-entropy[12]). Using the χ^2 measure in this way can be viewed as a comparison of Parzen density estimates:

$$\chi^2 = \frac{(B(x,y) - B'(x,y))^2}{B(x,y)} \propto \sum_{x,y} \left\{ \frac{\sum_{x',y'} R\left[\vec{V}(I,x,y), \vec{V}(I,x',y')\right]}{\sum_{x',y'} R\left[\vec{V}(I,x,y), \vec{V}(I',x',y')\right]} \right\}^2 \tag{9}$$

where $R(\cdot)$ is the Parzen kernel in equation (8). This process is outlined in Figure 5.

A classification decision is made by comparing the likelihood measure of a target image to a threshold η_{model}. Using standard χ^2 tables a value for η_{model} for any desired level of certainty can be determined.

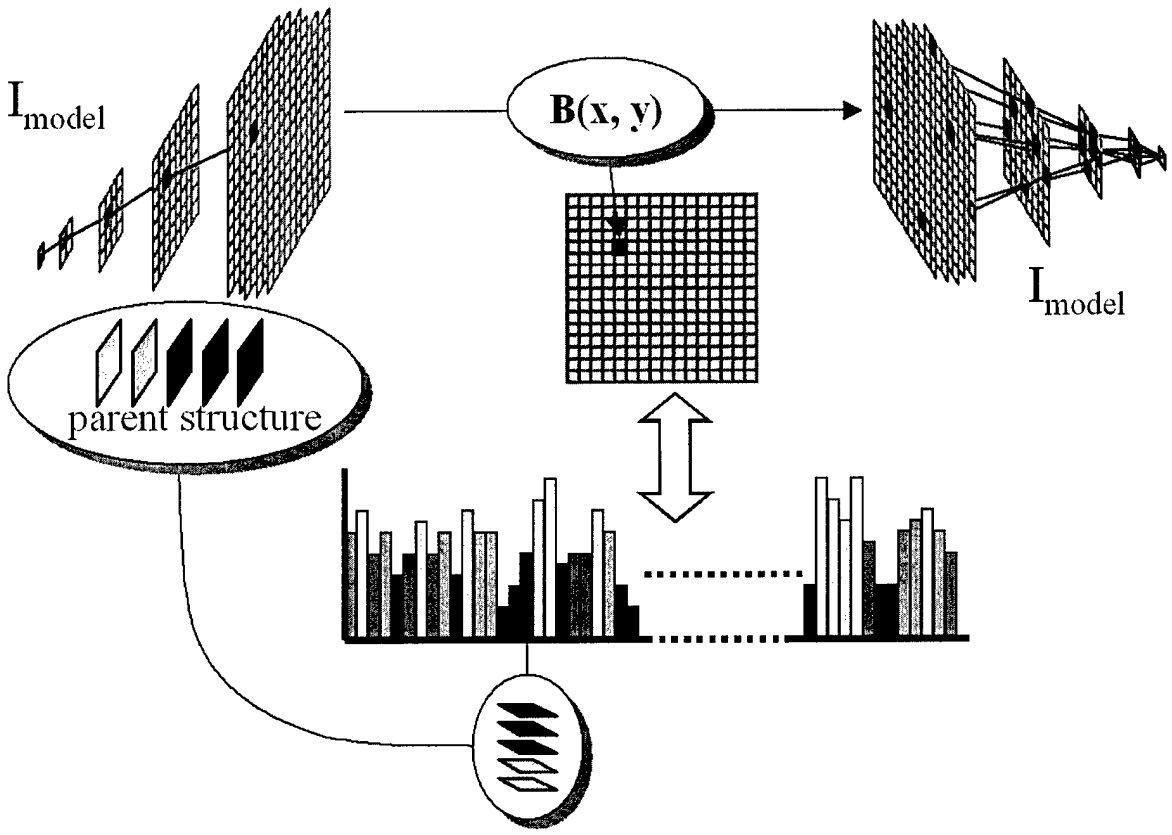


Figure 4: The flexible histogram accumulates the frequency of parent vectors which are within an $N \times M$ dimensional hypercube centered at each parent vector in the model image, to provide a measure of multiresolution redundancy.

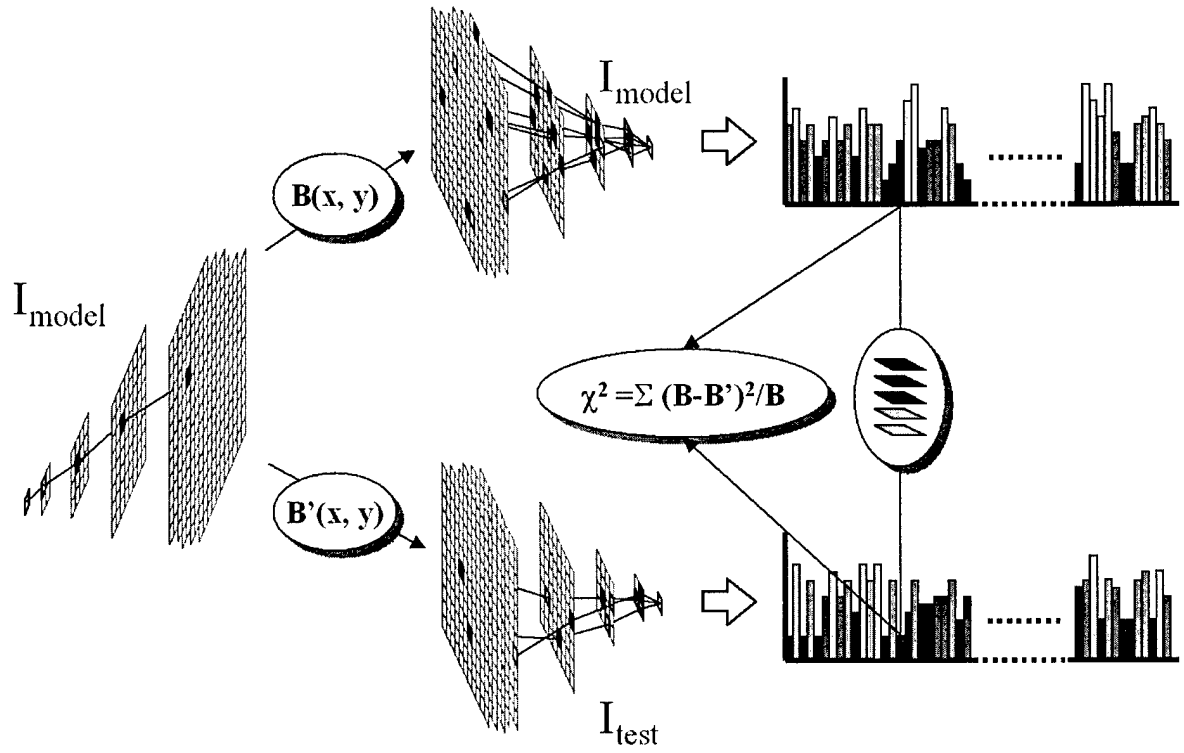


Figure 5: By comparing flexible histograms with the χ^2 measure we obtain a measure of the multiresolution similarity between the model image I_{model} and the test image I_{test}.

However, we do not assume that the flexible histogram completely describes the target class distribution; choosing a fixed χ^2 likelihood could eliminate true-positives which fall below the chosen threshold, yet which lie far above the false-positives and should therefore be detected. By choosing η_{model} empirically, we maximize the percentages of true-positives while guaranteeing an expected level of false-positives.

2.5 Incorporating multiple model images

Given access to multiple example images of the target vehicle, it is reasonable to assume that that a discrimination system should be able to perform better by taking advantage of the additional information. With slight modification, we can extend the single model flexible histogram technique to incorporate information from multiple examples of the target.

Consider now the situation where we have a set model images $\{I_{M0}, I_{M1}, \ldots, I_{Mn}\}$. We begin with the initial flexible histogram generated by accumulating the frequency of similar parent vectors in I_{M0} measured with respect to itself; information from the additional images can be incorporated into this histogram by adding the frequencies for each of the parent vectors in I_{M1} through I_{Mn} which are within the bins defined by I_{M0}. The additional parent vectors thus have the effect of refining the the frequency-counts in the histogram bins, but does not affect the number of bins present. This improves the target model by incorporating the relative frequencies of parent vectors in the additional model images, which will increase the accuracy of the histogram approximation to the distribution. And the flexible histogram becomes a more accurate estimate of the redundancy of the entire *target class.*

Using each of the n images as a base image for determining the bins in a flexible histogram, we can build n such estimators of class membership. Thus we can build a set of histograms $\{B_1, B_2, \ldots, B_n\}$ where histogram B_i is a measurement of the redundancy of each of $\{I_{M0}, I_{M1}, \ldots, I_{Mn}\}$ measured with respect to the parent vectors in I_{Mi}.

2.6 Experimental Setup

We now discuss experiments testing the performance of the flexible histogram method for detection and classification of SAR images of vehicles. The images used in this experiment are drawn from a library of SAR images of vehicles measured at X-band. The images are of approximately 1-foot-by-1-foot resolution. Our experiments are similar to those described in Velten *et al.* [3] which describes a template-based approach to classification and reports results on a database which has high intersection with the one we use here. The primary difference being that the confusion class in these experiments contains a greater number of vehicle types.

In one sense, getting meaningful detection or classification results is problematic. In our experiments we are only considering images of SAR vehicles. Presumably, in a full ATR system, large regions of clutter would be considered and some simpler test (e.g. a CFAR detector) would pass along regions of interest (ROIs) to a classifier. As observed by Velten *et al.* [3], the set of images we are using represent a more difficult set of ROIs than clutter or other man-made objects and so these experiments are a more stringent test than would be encountered in a practical ATR system. Furthermore, our primary goal is to provide a comparison between two approaches and for that purpose the data set we are using (with its restricted imagery) is perfectly valid. It is from the comparison standpoint that we would like these results to be viewed primarily.

2.6.1 Template-Based Approach Used for Comparison

For comparison purposes, we have implemented the algorithm described in Velten *et al.* [3]. The algorithm described there compares a thresholded and normalized template, M, to a thresholded normalized image chip (or ROI), S. As described, the template, M, is computed as an average over images within a designed aspect range. The detection statistic is the minimum L_1 difference between the template and the ROI computed over all template aspect ranges for a give vehicle type. In our implementation, we compute 36 templates for each vehicle model. Templates are computed using anywhere from three to nine images (from a training vehicle) depending on the number of aspect views that fall within the each 10 degrees of aspect. The total number of images used to compute each vehicle template model was 233 images.

2.6.2 Flexible Histogram Approach Used for Comparison

The results described here are based upon models for the target vehicle developed from 10 training images of each vehicle. The images were chosen uniformly in aspect, every 36^o. From these images 10 flexible histograms were generated for each vehicle using the extended procedure described in section 2.5. The detection statistic is the minimum χ^2 difference between the 10 pairs of flexible histograms generated for the training and testing images (measured with respect to each of the ten training images).

The choice of relatively small number, 10, images to generate each vehicle model is far less than the 233 images used in the template based approach, and was made to reflect a more reasonable estimation of the information available in real applications.

2.6.3 Description of Training and Test Set

As in Velten *et al.* [3] we consider three vehicle types, BMP-2, BTR-70, and T-72, in our classifier. Figure 6 shows example SAR images of the vehicles used to train the class models for both the template-based approach and the flexible histogram approach.

Images of different vehicles (same type of vehicle, different serial number) are used as the test set for the recognition class. Examples are shown in figure Figure 7. The images of training vehicles were collected at a depression angle of 17^o, while performance was tested on images of vehicles collected at 15^o. As stated, testing vehicles have different serial numbers than the vehicles used for training, except in the case of the BTR-70 (only one serial number in the data set). Another set of vehicles are used as confusers and are shown in figure Figure 8.

In this experiment we have a total of 777 independent testing vehicles for the recognition class (exclusive of 196 BTR-70 images which were collected from the same vehicle at different depression angles) and 1643 confusion images, for a total of 2616 images tested in this experiment.

2.6.4 Detection Results

Figure 9 compares the detection performance of both methods over the same image set. The detection statistic from a single vehicle model is used as a discriminant versus the rest of the test set. Figure 9 (a) (flexible histogram) and Figure 9 (d) (template method), for example, use the BMP-2 statistic as a discriminant against the confusion vehicles of figure Figure 8 as well as the remaining vehicles in the recognition class (i.e. T-72 and BTR-70). The results of figure Figure 9 do not include detection statistic from the training vehicle (except for the BTR70, for reasons stated above). As is readily observed, the flexible histogram approach compares favorably to the template-based approach for all three vehicle types.

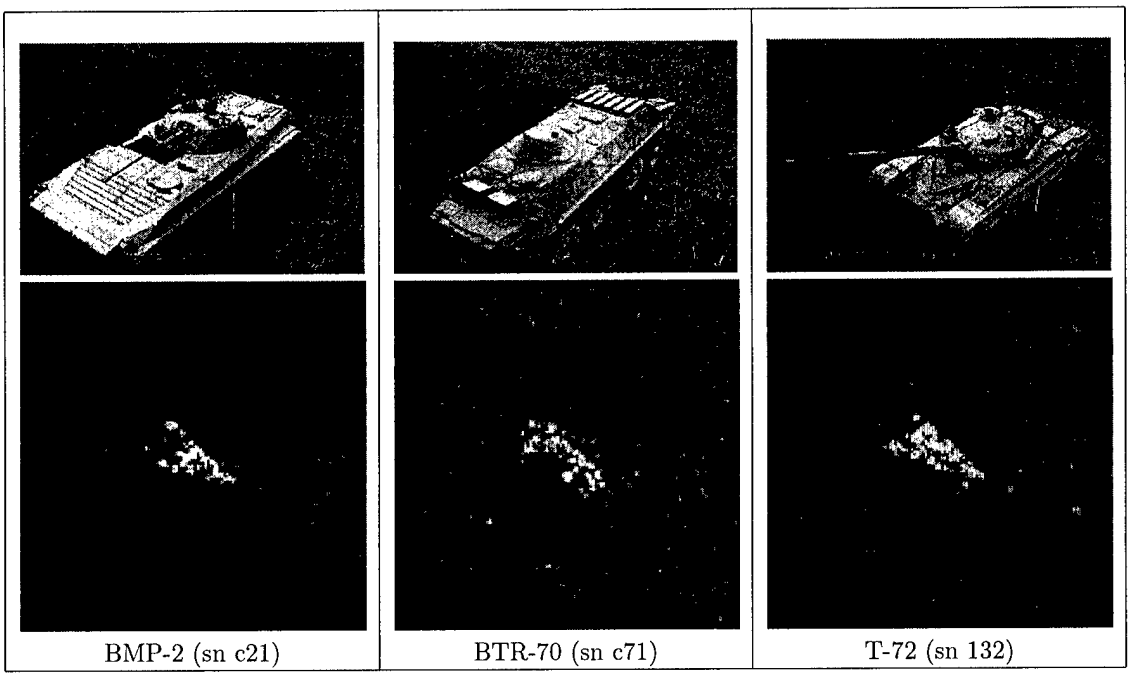


Figure 6: Training Vehicles for Recognition Class.

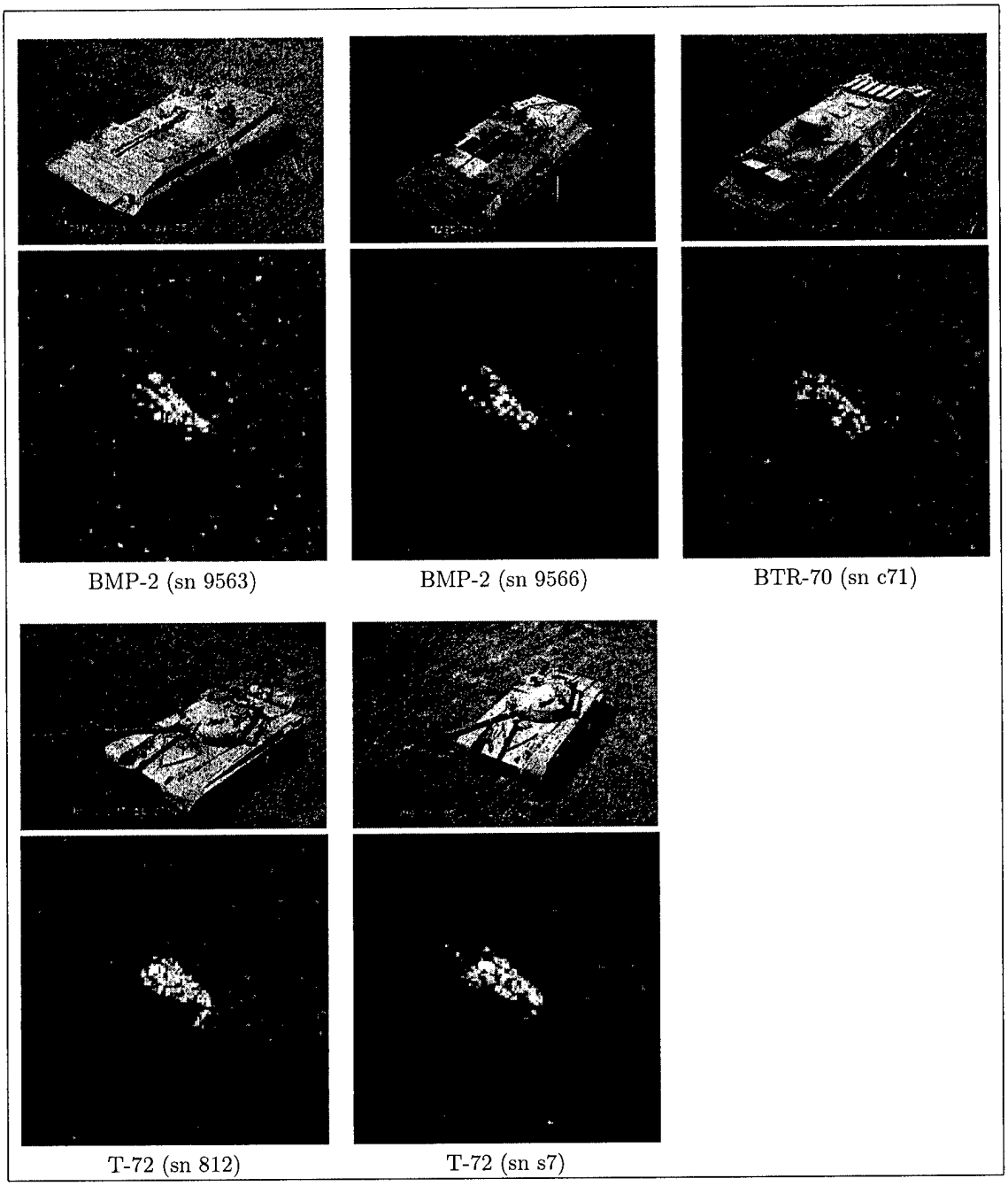


Figure 7: Testing Vehicles for Recognition Class.

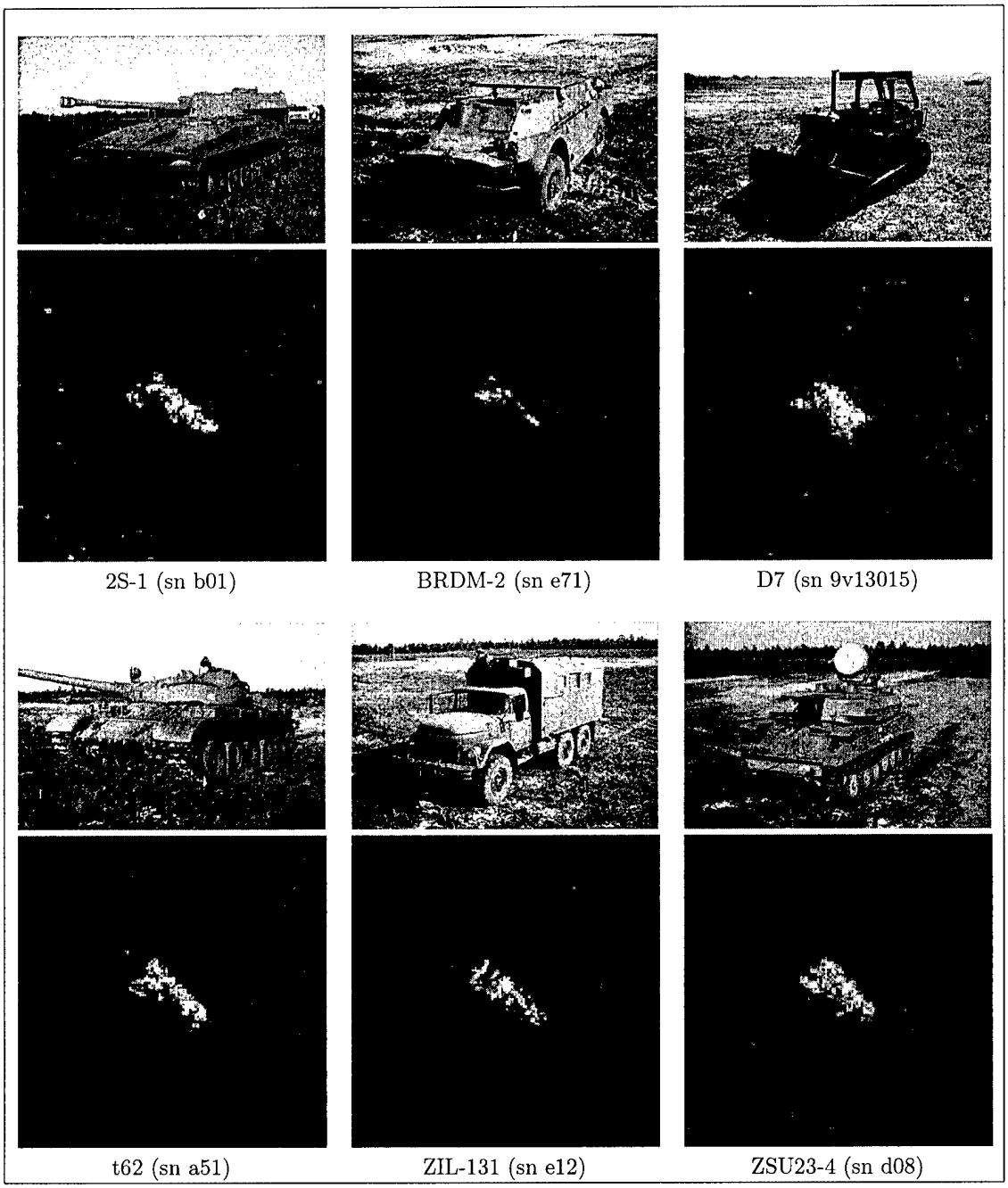


Figure 8: Testing Vehicles for Confusion Class.

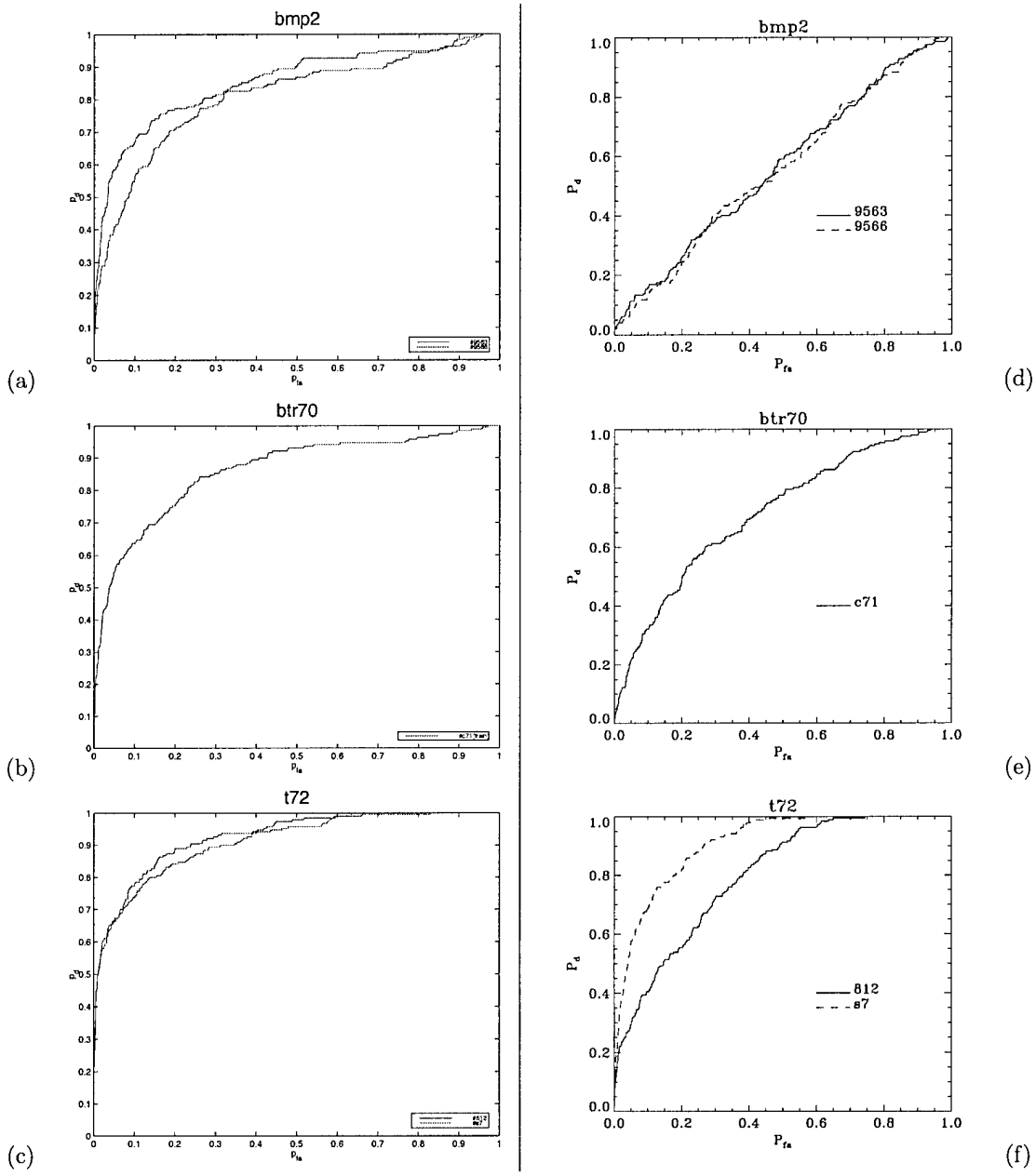


Figure 9: ROC curves for models generated by models for BMP2 (a,d), BTR70 (b,e) and T72 (c,f) vehicle types using the flexible histogram model (a-c) and the template matching method (d-f).

2.6.5 Classification Results

As stated, in this set of experiments, we lack a true *detection* statistic and so for the purposes of our experiment we use the following for both methods:

1. Given an ROI, compute each model statistic.
2. Choose the most likely (max or min statistic) over all of the vehicle models as the class label.
3. Compare the designated statistic against a threshold.
4. If it passes the threshold, label the ROI with the winning class, otherwise designate the ROI as "not classified".

In the steps described above, the threshold is model dependent and is set such that 90% of test vehicles of that model type will pass. The classification results (as a percentage) of this experiment are shown in table Table 1. Numbers in parentheses are the corresponding classification rate for the template-based approach. As the detection rate is the same for both methods it is not surprising that the classification performance on recognition vehicles is nominally the same. We do observe that the "not classified" rate is less than 10% for some cases, in particular the T-72. This is due to the confusion among the models (e.g. some of the rejected T-72s do pass the BMP-2 test).

It is in the performance on the confusion class that we note the largest difference. This is not surprising in light of the ROC curves from the detection results.

2.7 Discussion

The flexible histogram multiresolution target discrimination approach described here makes explicit the requirement that to be considered similar, images must contain similar distributions of *joint* feature responses over multiple resolutions. SAR classification rates shown here, though preliminary, compare favorably to the Wright Patterson baseline classifier. Further analysis, including experiments on larger data sets, and on data which includes images with targets and clutter in close proximity, are required to further refine the system and obtain a better estimate of its overall potential.

Future research suggested by this work includes specializing the flexible histogram technique to the particular characteristics of SAR imagery. In the work presented here, we only considered images generated from the magnitude of the SAR signal. In future work we plan to examine using the full complex signal to increase model specificity. Additionally, the particular wavelet decomposition used here is generic, and in future work we will be considering different decompositions which may be able to better match the images generated from SAR.

Table 1: Confusion matrix for the flexible histogram approach versus template-based results in parentheses at 90% detection rate.

Flexible Histogram				
	BMP2-C21	BTR-C71	T72-132	not classified
BMP2-9563 @ 15°	0.837	0.053	0.011	0.099
BMP2-9566 @ 15°	0.858	0.063	0.005	0.074
BTR70-C71 @ 15°	0.042	0.853	0.005	0.100
T72-812 @ 15°	0.021	0.053	0.858	0.068
T72-S7 @ 15°	0.021	0.126	0.779	0.074
BRDM2 @ 15°	0.174	0.174	0.132	0.520
D7 @ 15°	0.542	0.137	0	0.321
T62 @ 15°	0.626	0.121	0.068	0.185
ZIL131 @ 15°	0.379	0.168	0	0.453
ZSU-23-4 @ 15°	0.474	0.089	0.021	0.416
2S1 @ 15°	0.616	0.179	0.079	0.126
Template-Based Approach				
	BMP2-C21	BTR-C71	T72-132	not classified
BMP2-9563 @ 15°	0.903	0.005	0.005	0.087
BMP2-9566 @ 15°	0.847	0.015	0.015	0.122
BTR70-C71 @ 15°	0	0.898	0	0.102
T72-812 @ 15°	0.041	0	0.795	0.164
T72-S7 @ 15°	0.047	0.010	0.921	0.021
BRDM2 @ 15°	0.285	0.332	0.029	0.354
D7 @ 15°	0.529	0.153	0.223	0.095
T62 @ 15°	0.524	0.117	0.264	0.095
ZIL131 @ 15°	0.274	0.310	0.314	0.102
ZSU-23-4 @ 15°	0.394	0.095	0.453	0.058
2S1 @ 15°	0.493	0.234	0.069	0.204

3 Nonparametric Estimation of Aspect Dependence for ATR

In conventional SAR image formation, idealizations are made about the underlying scattering phenomena in the target field. In particular, the reflected signal is modeled as a pure delay and scaling of the transmitted signal where the delay is determined by the distance to the scatterer. Inherent in this assumption is that the scatterers are isotropic, i.e. their reflectivity appears the same from all orientations, and frequency independent, i.e. the magnitude and phase of the reflectivity are constant with respect to the frequency of the transmitted signal. Frequently, these assumptions are relatively poor resulting in an image which is highly variable with respect to imaging aspect. This variability often poses a difficulty for subsequent processing such as ATR. However, this need not be the case if the nonideal scattering is taken into account. In fact, we believe that if utilized properly, these nonideal characteristics may actually be used to aid in the processing as they convey distinguishing information about the content of the scene under investigation. In this paper, we describe a feature set which is specifically motivated by scattering aspect dependencies present in SAR. These dependencies are learned with a nonparametric density estimator allowing the full richness of the data to reveal itself. These densities are then used to determine the classification of the image content.

3.1 Introduction

The problem of target recognition is generally quite difficult, particularly in the context of SAR. Speckle poses an obvious impediment, but the task is further complicated by the high variability of the image with respect to imaging aspect. Actual vehicles are composed of a multitude of different fundamental scattering types. Individually, most of these atomic scatterers will exhibit an aspect dependence, but to further complicate matters, the effects of their mutual interaction will depend on their positioning relative to the impinging radar signal resulting in a highly convoluted aspect dependence.

Although the speckle and aspect dependence in SAR will limit performance when ignored or treated as a nuisance, we conjecture that they can actually be used to enhance ATR performance. The starting point of this work is a tree representation composed of multiresolution subaperture SAR images formed from recursive partitions of the full aperture. With access to images formed from multiple subaperture lengths and offsets, one can differentiate between different signatures across the aperture. Speckle also provides information which can be utilized in this structure. Speckle arises from the constructive and destructive interference associated with coherent imaging. Although the representation used is composed of only of log-magnitude imagery, the relative phase between subapertures is implicit as a result of the recursive partitioning. Thus, not only does the representation expose the magnitude of the aspect dependence but also the phase.

A nonparametric density estimator is used to learn the aspect dependence of targets within this subaperture data representation. In addition to revealing the aspect dependencies of a scatterer, the proposed subaperture feature set affords an efficient density estimation procedure via pruning methods. Classification of an object is determined by using the estimated density functions with an efficient approximation of the maximum likelihood (ML) classifier.

The algorithm proposed here is similar to the flexible histogram technique proposed by DeBonet[13]. Both techniques use nonparametric density estimators to characterize the behavior of the data in the feature space which is followed by a Chi-Square test to compare densities. The prominent distinction between the two techniques is in the feature set used. Motivated by the obvious visual structure present in high resolution SAR imagery, DeBonet et al. proposed a steerable wavelet feature set to extract phenomena such as edges and ridges in the image. In contrast, we are motivated by the scattering physics in SAR and propose a subaperture feature set designed to reveal the aspect dependent scattering behavior in the image.

Another difference between the two techniques is in the comparison of densities. Motivated by the high computational complexity of the approach by DeBonet et al., all densities are evaluated at a fixed set of points that is invariant to both the test image and hypothesis being checked. Use of this invariant set of points affords a significant reduction in the computational order of the algorithm.

The remainder of the paper presents these ideas in more detail as follows. Section 3.2 describes the subaperture feature set used to exploit aspect dependencies. Section 3.3 describes the density estimator used and presents the test applied to the density estimates. Section 8.4 presents results of this technique applied to the MSTAR database. Section 3.5 contains concluding remarks and a summary.

3.2 Subaperture Analysis

An essential component of the framework described in this paper is the feature extraction prior to the density estimation. The features selected should contain class discriminating information from SAR data. In this section, we present a set of filters motivated by the scattering physics involved in SAR image formation to accomplish this task. The resulting data representation reveals the anisotropy of the scattering with a multiscale tree structure which allows for efficient estimation of the probability density.

There has recently been a considerable amount of work [14, 15, 16, 17] studying the fundamental behavior of anisotropic scatterers in SAR for the purpose of target recognition. With the exception of the approach by McClure and Carin[16], these limit consideration to isolated atomic scatterers, and in the case of McClure and Carin's approach, they treat the scattering behavior of multiple scatterers as a simple linear superposition of those scatterers. This superposition allows a computationally tractable algorithm, but neglects the electromagnetic coupling among neighboring scatterers. In contrast to the atomic scattering approach, we present a method based on a feature set which is intended to focus on the anisotropic scattering behavior of the vehicle as a whole instead of the individual scatterers of which it is composed. It does this by analyzing the scattering response across the aperture as the cross-range resolution is varied.

3.2.1 Subaperture Coverings

Our goal is to exploit the aspect dependencies associated with man-made objects. Towards this end, we introduce a set of subaperture images. In order for these subapertures to clearly expose the dependencies, the subapertures should have some structure to them. Before describing one such structure, we first define some notation. A subaperture will be defined by a half-open interval $I = [a, b)$ with $0 \leq a < b \leq 1$ where a and b specify the endpoints of the subaperture normalized so that the full aperture is $[0, 1)$. For example, $[0, .5)$ and $[.5, 1)$ denote two disjoint half-apertures. Since the phenomena we seek to detect are localized in the aperture, each subaperture will correspond to a single convex interval. Naturally, the collection of subapertures used should cover the unit interval so that none of the data is ignored, i.e. the union of all the intervals should correspond to $[0, 1)$. Such a collection of apertures will be referred to as a *subaperture covering.*

There are numerous coverings from which to choose. We will consider only those which can be mapped to a tree, one example of which is given in Figure 10. Each subaperture within the same level of the tree will be restricted to have the same length. A subaperture I is the descendant of another subaperture J if $I \subset J$. Furthermore, for any two subapertures I and J with the length of J larger than that of I, either $I \cap J = I$ or $I \cap J = \emptyset$, i.e. the smaller aperture is either entirely included in the larger one or disjoint from it. Within this tree structure, it will frequently be convenient to index the subapertures by two subscripts m and l specifying the resolution (or subaperture length) and center of the subaperture as depicted in Figure 10. For a tree with $M + 1$ levels, the root level of the tree will be denoted with $m = M$ and successively smaller

apertures will have corresponding smaller integer values down to $m = 0$ at the bottom of the tree. Thus, larger values of m correspond to larger subaperture lengths.

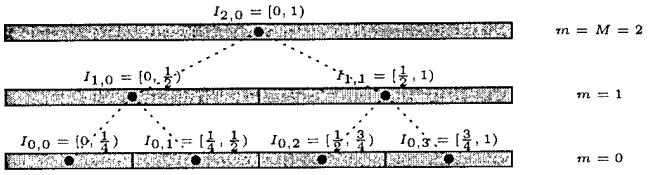


Figure 10: A subaperture covering in the form of a binary tree of subapertures that is obtained by iteratively partitioning each subaperture into two disjoint intervals of equal length.

The subaperture covering in Figure 10 is obtained by iteratively partitioning each subaperture into two disjoint half-apertures. For another example, consider the covering generated by dividing each subaperture into half-overlapping half-apertures as illustrated in Figure 11(a). Note that for this covering, at a given level $m < M - 1$, there are duplicate subapertures. Thus, the subapertures can be presented as in Figure 11(b) giving a more intuitive visualization of the covering.

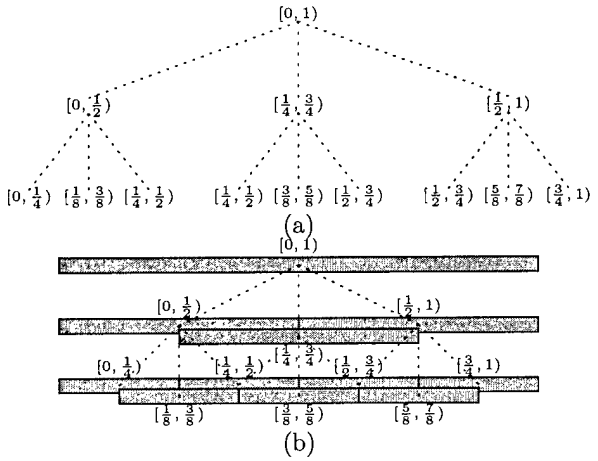


Figure 11: Example of a subaperture covering obtained by iteratively partitioning each subaperture into three half-overlapping half-apertures. (a) Tertiary tree showing parent-child subaperture structure. (b) Simplified graph representation with redundant subapertures removed.

Different types and sizes of a scatterer will yield different aspect dependencies. The motivation for using the subaperture tree is that it is expected to reveal distinguishing aspect dependencies in the scattering. For example, a small metal sphere will have a strong response in all directions and thus produce a strong reflectivity estimate from each of the subapertures. However, as depicted in Figure 12, a flat plate produces a significantly stronger response when oriented broadside with respect to the radar as compared to off-broadside orientation. Thus, the reflectivity estimates will vary across the subapertures with the largest estimate coming from the subaperture oriented broadside to the plate. Furthermore, because various sized

subapertures are used, the duration of the broadside flash is captured in this representation. In particular, while the subaperture is contained within the mainlobe of the response, the reflectivity estimate will be consistently large, but as the subaperture is extended beyond the mainlobe, the additional energy received will be relatively insignificant and result in a lower reflectivity estimate. This comparison between a sphere and flat plate illustrates the extraction of features associated with atomic scatterers, but we expect this to extend to extracting the more complex scattering pattern from a collection of scatterers associated with an entire target.

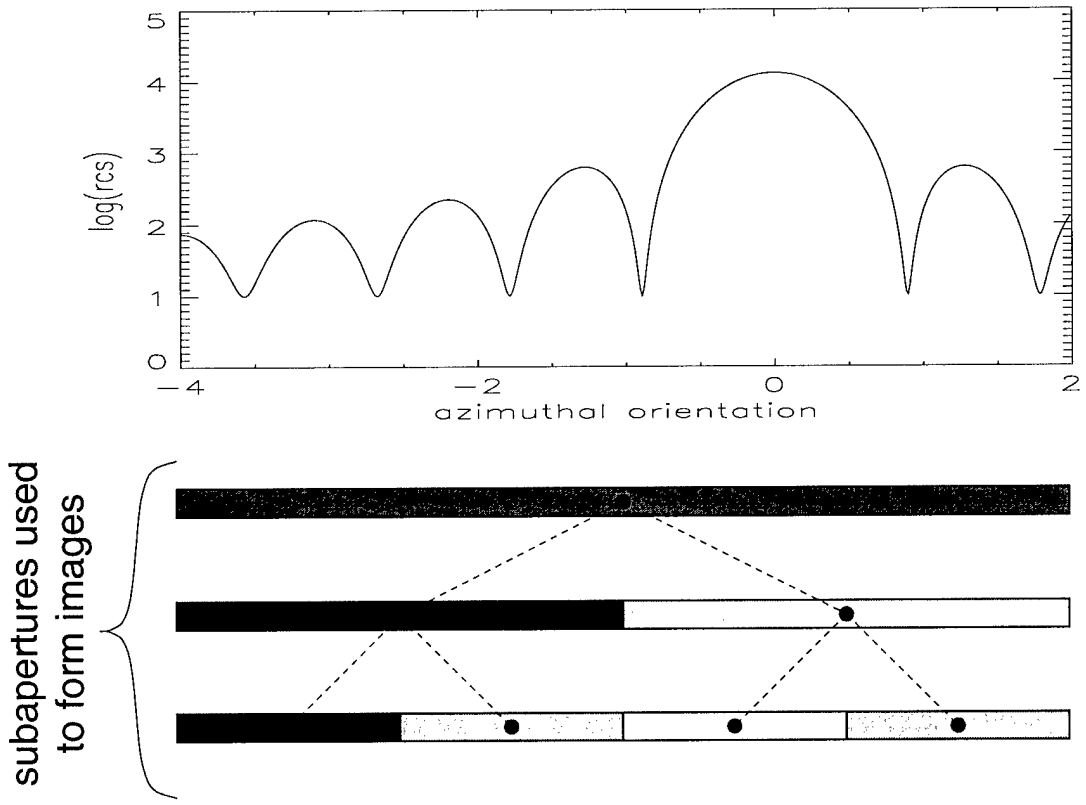


Figure 12: The predicted response of a $1m \times 1m$ flat plate and a depiction of the resulting reflectivity estimate for each of the subapertures. Lighter shaded subapertures convey larger reflectivity estimates.

A slightly different viewpoint of this subaperture feature set comes from considering the cross-range resolution versus azimuthal[3] resolution trade-off. Cross-range resolution is inversely proportional to the aperture length. Thus, at lower levels of the subaperture tree, spatial resolution has been sacrificed for azimuthal resolution, i.e. the ability to better observe anisotropic phenomena. This is the classic time–

[3]For clarity, we will use the term "cross-range" when referring to the cross-range in the spatial (image) domain, and we will exclusively use the term "azimuthal" when referring to the corresponding dimension in the sensor domain.

frequency resolution tradeoff in Fourier analysis, and each level of the tree represents the data under a particular cross-range–azimuth resolution. The presence of multiple resolutions is attractive because we expect the best representation for different objects to be nonunique as the importance between resolution in the two domains is balanced.

3.2.2 Subaperture Feature Vector

We will now describe the subaperture feature vector. For each spatial location, every subaperture in the covering is used to produce an estimate of the log-magnitude of the reflectivity. These estimates are taken to comprise a feature vector $\mathbf{x} \in \mathbb{R}^d$, where d is the number of subapertures. A feature vector is computed for each spatial location in the SAR target field producing an image of feature vectors that together describe the aspect dependence throughout the image. We consider subaperture coverings in which the lengths of the subapertures are such that the ratio of sizes of larger to smaller apertures is always an integer, thus allowing the feature vectors to be mapped to a tree based on the area covered by a resolution cell. Note that it is immaterial whether the disjoint or the overlapping covering is used here since the resolution is completely determined by the lengths of the subapertures. Thus, for the half-aperture splitting we will always end up with a binary tree with vector valued entries. However, the number of subapertures with a given length, i.e. the type of half-aperture covering, will determine the dimensionality of each of the nodes on the binary tree.

Prior to moving on to the next section, it is important to make clear the difference between the two different trees with which we are working. The first is the subaperture tree which arranges the subapertures used to produce reflectivity estimates. Each node on this tree is a pair of numbers defining an interval which specifies a subaperture section. This tree has nothing to do with spatial location. The second tree is the feature vector tree which is used to represent the reflectivity estimates over the entire target field under different subapertures. This tree is composed of vector valued reflectivity estimates. A path down this tree represents a feature vector and different levels of the tree correspond to estimates at different resolutions.

Although the feature vector tree and subaperture trees are different, there are relationships between the two. In fact, the topology of the subaperture tree determines the topology of the feature vector tree and vice-versa. Figure 13 depicts these relations for the half-overlapping half-aperture covering. First, the number of levels in the two trees are the same because the levels of both trees are defined by resolution. However, the root node of the subaperture tree corresponds to the leaves of the feature vector tree and vice-versa. To see why this reversal arises, recall that the root node of the subaperture tree is the full aperture, and use of the entire 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 apertures in the subaperture tree are at the bottom of their tree, but their corresponding nodes are at the root of the feature vector tree since they give the coarsest cross-range resolution. Due to this bijective relation for resolution, the same label m will be used to denote scale in both trees. However, to account for the reversed progression of resolution down the tree, we reverse the values of m at each scale on the feature vector tree, i.e. $m = 0$ for the root node and $m = M$ for the leaves. The number of children that each node on feature vector tree spawns is equal to the ratio of subaperture lengths between adjacent levels in the subaperture tree. For example, any covering with half-aperture splitting will produce a binary tree of feature vectors. The dimensionality of the nodes at any level in the feature vector tree is equal to the number of subapertures for the associated resolution since each subaperture corresponds to a reflectivity estimate.

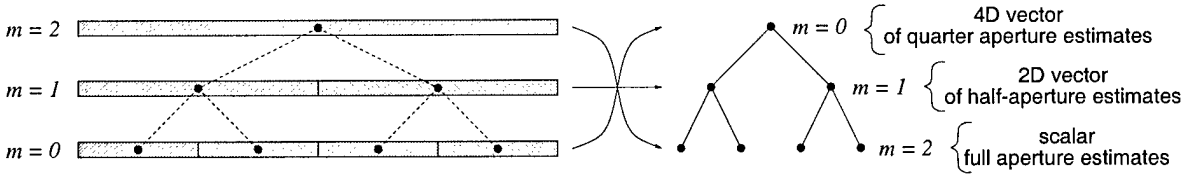


Figure 13: The relationship between the subaperture and feature vector trees for the half-overlapping half-aperture covering.

3.2.3 Phase Information

Speckle is commonly thought of as a nuisance in SAR imagery, but in fact, it can provide a valuable source of information that is implicit in this subaperture feature set. Speckle arises from the constructive and destructive interference associated with coherent imaging systems such as SAR. It is by way of this interference that one can determine the relative phase between pixels at different resolutions. Consider a portion of the disjoint half-aperture covering illustrated in Figure 14 where a, b, and c represent the complex reflectivity of multiresolution pixels corresponding to the same resolution cell center. Then, the *complex* reflectivities b and c can be combined to produce a finer resolution reflectivity estimate $a = be^{-j\phi} + ce^{j\phi}$ where ϕ determines the cross-range sampling rate and is thus known so the phase term can be absorbed into b and c. However, our subaperture representation only provides the magnitudes $|a|$, $|b|$, and $|c|$. But, this corresponds to being given three sides of a triangle. Thus, fixing one of the angles, say the phase of a is 0, then the other two angles can be determined up to a sign. Thus, relative to the phase of a, the value of b and c can be determined, up to a sign, in this data representation even though it is composed solely of real-valued log-magnitude imagery.

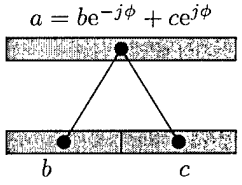


Figure 14: Relation between reflectivity a and its corresponding subaperture reflectivities b and c.

3.3 Classification

Having presented the subaperture feature set, we now proceed to describe a method to exploit it for classification. The first step is to learn the scattering behavior using a nonparametric density estimator. These density estimates are then used to determine classification by matching the pdf of the test data to the "closest" pdf in the hypothesis class.

3.3.1 Density Estimation

Knowledge of probability densities is essential in classification. Thus, the relevant densities must somehow be estimated to some degree. Some techniques 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: parametric and nonparametric. We choose a nonparametric density estimator because it can fully capture the complex relationships we expect to observe in SAR imagery. Before describing the estimator used, we will define some notation. Capital boldface letters are used to denote random vectors, e.g. $\mathbf{X}$, and samples of it are denoted with a corresponding lower case letter, e.g. $\mathbf{x}$. Calligraphic letters, e.g. $\mathcal{X}$, will be used to denote the corresponding space of possible samples. The distribution for a random variable will be denoted by $F(\cdot)$ and the corresponding pdf[4] will be denoted by $f(\cdot)$.

The nonparametric estimator we use is the Parzen[18] density estimator which is defined as follows: given N i.i.d. samples $\mathbf{x}_1, \ldots, \mathbf{x}_N \sim F$, the Parzen density estimate for f is

$$\hat{f}(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} \kappa(\mathbf{x} - \mathbf{x}_i; \mathbf{h}) \tag{10}$$

where $\kappa(\cdot)$ is the kernel and $\mathbf{h}$ is a vector of *kernel widths* for each dimension. 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 value than the probability at some distant value. An attribute of this technique is that when a pdf with finite second moment is used as the kernel, the density estimator is consistent, i.e. $\hat{f}$ converges to the true pdf in probability, if the kernel width is an appropriately decreasing function of n [18].

3.3.2 Score Function

We wish to classify a set of i.i.d. samples from an unlabeled distribution given a set of samples from each of the candidate distributions. To do this, we will employ the Parzen density estimator to obtain density estimates to describe both the unknown distribution and each of the known distributions based on the observed samples. An appropriate use of these distributions in this case is to compare them with the theoretic concept of Kullback-Leibler (KL) divergence between pdf's. The KL divergence between two pdf's f and g is defined as

$$D(f\|g) = \int f(\mathbf{x}) \log\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) d\mathbf{x} \tag{11}$$

which is the expected value, under f, of the log-likelihood ratio of f and g.[5] 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.

To see why the KL divergence is an appropriate criterion, consider the following problem. We are presented with L sets $\{\mathcal{S}_1, \ldots, \mathcal{S}_L\}$ of data each containing N samples. For each l, all the samples in $\mathcal{S}_l$ are generated as i.i.d. samples from the *unknown* distribution $F_l(\mathbf{x})$. We are now presented with another set $\mathcal{S}_0 = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ of data consisting of i.i.d. samples drawn from one of the distributions $F_l(\mathbf{x})$. The problem is to choose under which F_l the samples in the set $\mathcal{S}_0$ was produced. Because we do not know the actual

[4]Only absolutely continuous distributions will be considered, and thus the corresponding density function does exist.

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

pdf's f_l, we use their Parzen estimates $\hat{f}_l$ for each $l \in \{0, \ldots, L\}$. We now show the asymptotic equivalence between the minimization of KL divergence and the ML classifier. Starting from the ML classifier for the density generating the data set $\mathcal{S}_0$, we get

$$\begin{aligned}
\hat{l} &= \arg \max_{l \in \{1,\ldots,L\}} \{f_l(\mathbf{x}_1, \ldots, \mathbf{x}_N)\} \approx \arg\max_l \left\{ \hat{f}_l(\mathbf{x}_1, \ldots, \mathbf{x}_N) \right\} \\
&= \arg\max_l \left\{ \log\left(\frac{\hat{f}_l(\mathbf{x}_1, \ldots, \mathbf{x}_N)}{\hat{f}_0(\mathbf{x}_1, \ldots, \mathbf{x}_N)} \right) + \log\left(\hat{f}_0(\mathbf{x}_1, \ldots, \mathbf{x}_N) \right) \right\} \\
&= \arg\max_l \left\{ \log\left(\frac{\hat{f}_l(\mathbf{x}_1, \ldots, \mathbf{x}_N)}{\hat{f}_0(\mathbf{x}_1, \ldots, \mathbf{x}_N)} \right) \right\} \\
&= \arg\max_l \left\{ \log\left(\prod_{i=1}^{N} \frac{\hat{f}_l(\mathbf{x}_i)}{\hat{f}_0(\mathbf{x}_i)} \right) \right\} \text{ (since the } \mathbf{x}_i \text{ are assumed to be independent)} \\
&= \arg\min_l \left\{ \frac{1}{N} \sum_{i=1}^{N} \log\left(\frac{\hat{f}_0(\mathbf{x}_i)}{\hat{f}_l(\mathbf{x}_i)} \right) \right\} \\
&\approx \arg\min_l \left\{ \int \hat{f}_0(\mathbf{x}) \log\left(\frac{\hat{f}_0(\mathbf{x})}{\hat{f}_l(\mathbf{x})} \right) d\mathbf{x} \right\} = \arg\min_l \left\{ D(\hat{f}_0 \| \hat{f}_l) \right\}. \qquad (12)
\end{aligned}$$

All approximations are meant in the sense that the argument of the $\arg\max\{\cdot\}$ is approximate. The first approximation is that the estimated pdf's are close to the true underlying pdf's which follows from the consistency of the Parzen density estimator. The second approximation follows from the law of large numbers so that the relative frequency of the $\mathbf{x} \in \mathcal{S}_0$ converge to the estimated probability $\hat{f}_0(\mathbf{x})$. This result 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 ML classification.

For the proposed classifier, we will take the subaperture feature vector as the random vector $\mathbf{X}$ for the reasons described in Section 3.2. An advantage of this feature set is that it allows for efficient estimation of the pdf due to its tree structure. Computing the distances used in the Parzen kernel in a top-down manner allows for the density estimation for each point to be done in $\mathcal{O}(N_1 \log(N_2))$ instead of $\mathcal{O}(N_1 N_2)$ by exploiting the tree structure in cross-range. With the computational efficiency afforded by the tree structure comes a liability. A tree structure can be used because the feature vectors are multiresolution. But inherent to multiresolution feature vectors are coarse scale dependencies, due to overlapping resolution cells, thus violating the independence assumption used in Eq. (12). Thus, it is reasonable that other divergence measures may outperform the KL divergence. We have empirically found that one such statistic is the Chi-Square (χ^2) divergence, which is a first order approximation of the KL divergence. The χ^2 divergence for two pdf's f and g is given as

$$\chi^2(f, g) = \int \frac{(f(\mathbf{x}) - g(\mathbf{x}))^2}{g(\mathbf{x})} d\mathbf{x}. \qquad (13)$$

Because of its better performance, the χ^2 divergence will be used in place of the KL divergence for the results presented in the following section. The appropriate divergence measure for multiresolution feature vectors such as the subaperture feature set is an area currently under study.

In light that this is an approximation of the ML classifier, it is reasonable to ask why should the KL or χ^2 divergence approach above be used in place of a direct comparison of likelihoods. Our motivation for choosing the former is the gain in computational efficiency. Letting N_{test} be the number of different test

data sets, i.e. the number of times the classifier will be run, the computational order of the two approaches can be broken down as follows. If we were to directly perform ML classification, then a total of $(L+1)N_{test}$ likelihoods would have to be computed since $L+1$ likelihoods would have to be computed for each of the different test sets. Furthermore, the computation of each likelihood requires a density value to be estimated at each of the sample points. Since density estimation for a single point is in an $\mathcal{O}(N_1 \log(N_2))$ procedure, this produces an $\mathcal{O}(N_1^2 N_2 \log(N_2)\, LN_{test})$ algorithm to classify all the data. Now consider the KL or χ^2 approach used with a *fixed* set of k points giving a discrete approximation of $\mathcal{X}$. Here, only $L+N_{test}$ densities need to be estimated since the estimates are made at a fixed set of locations which are invariant with respect to the data samples, whereas in the direct ML application the estimates are made at the data samples themselves. Furthermore, the densities only need to be computed at k locations which could be much less than $N_1 N_2$ for reasonably smooth densities. This gives an algorithm which is $\mathcal{O}(kN_1 \log(N_2)\,(L+N_{test}))$. That k can be much less than the number of data samples is reasonable because the number of points at which the pdf is estimated should not depend on the number of data samples but the smoothness of the underlying pdf. Considering Nyquist's sampling criterion, it is intuitive that smooth pdf's will not need to be sampled as finely as pdf's with less regularity since evaluating the pdf at an excessive number of locations is redundant as each estimate will be largely dependent on the others. However, the accuracy and implications of the Riemann sum approximation to the integral with a sparse sampling of points is an open issue currently under investigation.

3.4 Experimental Results

Public release MSTAR data is used for the results presented here. These contain 128×128 images with a resolution of $0.3m$ in both range and cross-range. The transmitted signal had a bandwidth of $0.591GHz$ and a center frequency of $9.60GHz$.

This paper is based on the assumption that the subaperture feature set captures the aspect dependence in SAR imagery. An illustration of this is given in Figure 15 which shows the image generated from each subaperture in the disjoint half-aperture covering for a bmp2 tank at a 17° elevation and 0.19° azimuth. From these images, the aspect dependence of the broadside flash on the front side of the tank is clear. One can deduce that the duration of the flash is less than half of the aperture length and is centered at slightly greater than 0.5 on the aperture. That the flash duration is less than half the aperture is apparent from the quarter-aperture imagery, where the majority of the energy of this flash can be seen to be contained in the $[0.25, 0.5)$ and $[0.5, 0.75)$ subapertures. That more of the flash is contained in the $[0.5, 0.75)$ subaperture is apparent from the larger reflectivity as compared to the $[0.25, 0.5)$ subaperture. This along with the observation that the $[0.5, 1)$ subaperture produces a stronger response than the $[0, 0.5)$ subaperture indicates that the center of the flash is located slightly above 0.5 on the aperture. As is done for this example, we will use the disjoint half-aperture subaperture covering with three resolutions for the remainder of the experiments in this section.

The training data for this experiment consist of the following vehicles at a 17° elevation with their serial number in parentheses: bmp2 (c21), btr70 (c71), and t72 (132). The testing data consists of the following vehicles at a 15° elevation angle: bmp2 (9563 and 9566), btr70 (c71), t72 (812 and s7), 2s1 (b01), drbm2 (e71), d7 (9v13015), t62 (a51), zil131 (e12), and zsu23-4 (d08). The serial numbers of the testing set vehicles are different than those for the training set vehicles with the exception of the btr70 (c71) in which case there is only one serial number available.

The hypothesis space is discretized according to each of the three vehicle types and azimuthal pose. In these preliminary experiments, the azimuthal coordinate is sampled every 36°, thus producing a hypothesis space with a total of $L = 30$ hypotheses. Each hypothesis is associated with a single training image

corresponding to that vehicle type and azimuthal orientation.

Recall that the underlying pdf's for the data sets are estimated at a finite number of locations or kernel centers. For the results presented here, these locations were chosen as so to be approximately uniformly distributed over the union of supports of all the hypothesis pdf's. The number of kernel centers k is fixed apriori, and the uniform distribution is obtained via a slight modification of the k-means algorithm described in Appendix 3.6. The number of kernel centers used for the experiments presented here is $k = 512$.

The density estimator employed is the Parzen density estimator using a hyper-rectangle kernel, i.e. the kernel in Eq. (10) is

$$\kappa(\mathbf{x};\mathbf{h}) = \begin{cases} 1 & : \|\mathbf{x}^{(m)}\|_\infty < \mathbf{h}^{(m)}, \forall m \\ 0 & : \text{otherwise} \end{cases} \tag{14}$$

where $\mathbf{x}^{(m)}$ represents the subvector of $\mathbf{x}$ corresponding to scale m, and $\mathbf{h}^{(m)}$ represents the kernel width used for all feature vector components at scale m.

The role of the kernel width in density estimation is to balance the trade-off between the bias and variance of the estimated density. As our ultimate goal is discrimination, we choose the kernel width to satisfy a Neymann-Pearson criterion at a particular scale at a time. Holding the kernel widths constant for all but a single scale, the kernel width for that scale is chosen to minimize the probability of false alarm given a particular probability of detection[6]. The resolutions are repeatedly cycled through, finding the optimal kernel width for each scale until a maximum is reached.

In Figure 16, we show the receiver operating characteristic (ROC) curves for the bmp2 tank, btr70 transport, and t72 tank. The probability of detection is measured on all the testing data for the same vehicle; the probability of false alarm is measured on the entire testing set minus those for the target vehicle. For each vehicle, we compare the performance with the subaperture feature set to a similar technique by DeBonet et al.[13] in which a steerable wavelet pyramid is used for the feature set and also the Wright Patterson Air Force Base standard template matching algorithm[19]. These results are quite promising especially considering the small number of kernel centers used ($k = 512$ instead of $N_1 N_2 = 16,384$) and reduced computational complexity[7].

3.5 Conclusions

In this paper, we have presented a classification algorithm designed to utilize the aspect dependence of scatterers in SAR imagery to enhance classification performance. The aspect dependence of the imagery is exploited by the use of a subaperture feature set that captures the scattering behavior across the aperture. A nonparametric density estimator is then used to learn the particular patterns for each hypothesis class and test image. The classification of the test image is then determined via a χ^2 divergence between the estimated densities. By using a common set of points to evaluate the pdf's and the tree structure of the feature set, significant gains in the computational order of the algorithm can be achieved.

3.6 Discretization of the sample space

A slight modification of the k-means algorithm is used to generate an approximately uniform sampling of points at which to estimate the pdf's. Let k denote the number of points to choose. The steps in the

[6] All the data involved in this training of the kernel width is from the 17° depression data used for training, and does not involve any 15° depression data.

[7] The order of the technique proposed by DeBonet[13] is the same as direct ML classification since the pdf's are evaluated at the test data points and not on a common set of evaluation points.

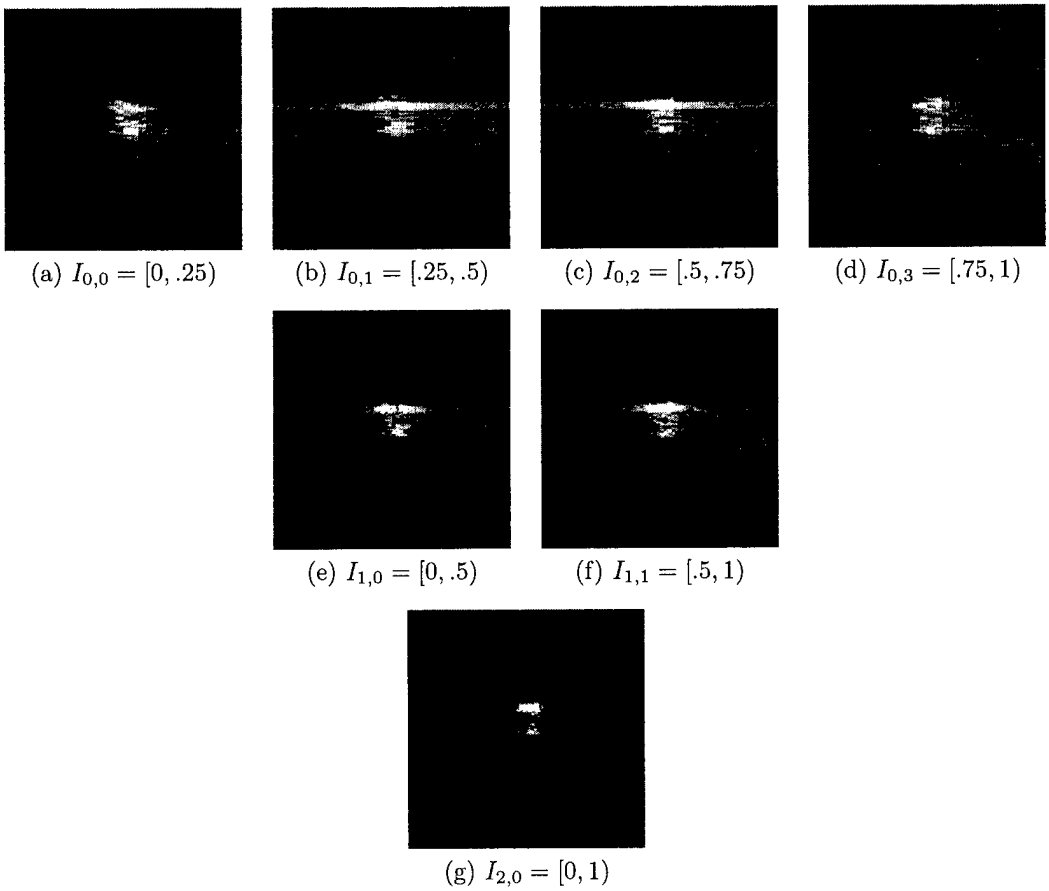
procedure are illustrated in the flow diagram in Figure 17. First, k points are initialized by setting them to randomly chosen training feature vectors. Regions are then defined by assigning each training vector to the closest point. "Closeness" is measured in terms of the supremum norm since a hypercube is used for the density estimation. Each of the k points is then redefined to be the center of the region, i.e. the center of the smallest bounding box containing all points in that region. A distortion measure is then computed as the root-mean-square (RMS) of the sizes of the regions, where the size of a region is taken as its side length. 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.

The intuition behind the algorithm is to adjust the sizes of the box regions until they are all the same size, i.e. the points are uniformly spaced with respect to the supremum norm distance. If there is a box larger than surrounding ones, then the after computing the center of the box and redefining the region, the larger box should shrink as its former members are moved to another region where they are closer to the center.

This procedure differs from the standard k-means algorithm in step 3 of Figure 17 and 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 its geometric center. The natural distortion measure used in this case is the RMS of the distances from each training vector to the nearest centroid. Using the k-means algorithm produces a sampling that is approximately 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. However, a uniform sampling is preferred in the context of this research because values where the density is small may be very important and thus should not be neglected. Given a particular point which has a small density value under one hypothesis and a large value under another, this location is important in deciding how well each hypothesis describes the observed data. Recall that we compute the χ^2 distance between pdf's f and g as in Eq. (13). Thus, locations where g is small would contribute significantly if f is not small since it would be an anomaly for those points to occur under g. This is consistent with the idea that it is the unusual, i.e. low probability, events that offer the most distinguishing features.

The dominance of clutter in the image chips has a significant impact on the allocation of pdf evaluation points. In particular, many of them will be associated with the nondiscriminating clutter. Such points will not help to distinguish between target types since they are measuring the scattering properties of clutter and not of the target. As a preliminary remedy to this waste of evaluation points, the modified k-means algorithm is only run over feature vectors that are either above or below a particular threshold so that the resulting evaluation points will be more representative of the bright scatterers and shadows associated with vehicles.

Although the modification of the k-means algorithm, like the original, is computationally burdensome, it can be performed offline and need only be done once to define where the pdf's for a specific vehicle are to be estimated.



(a) $I_{0,0} = [0, .25)$ (b) $I_{0,1} = [.25, .5)$ (c) $I_{0,2} = [.5, .75)$ (d) $I_{0,3} = [.75, 1)$

(e) $I_{1,0} = [0, .5)$ (f) $I_{1,1} = [.5, 1)$

(g) $I_{2,0} = [0, 1)$

Figure 15: Subaperture images of a bmp2 at 17° elevation and 0.19° azimuth. For each image, the front of the vehicle is the portion nearest the top of the image.

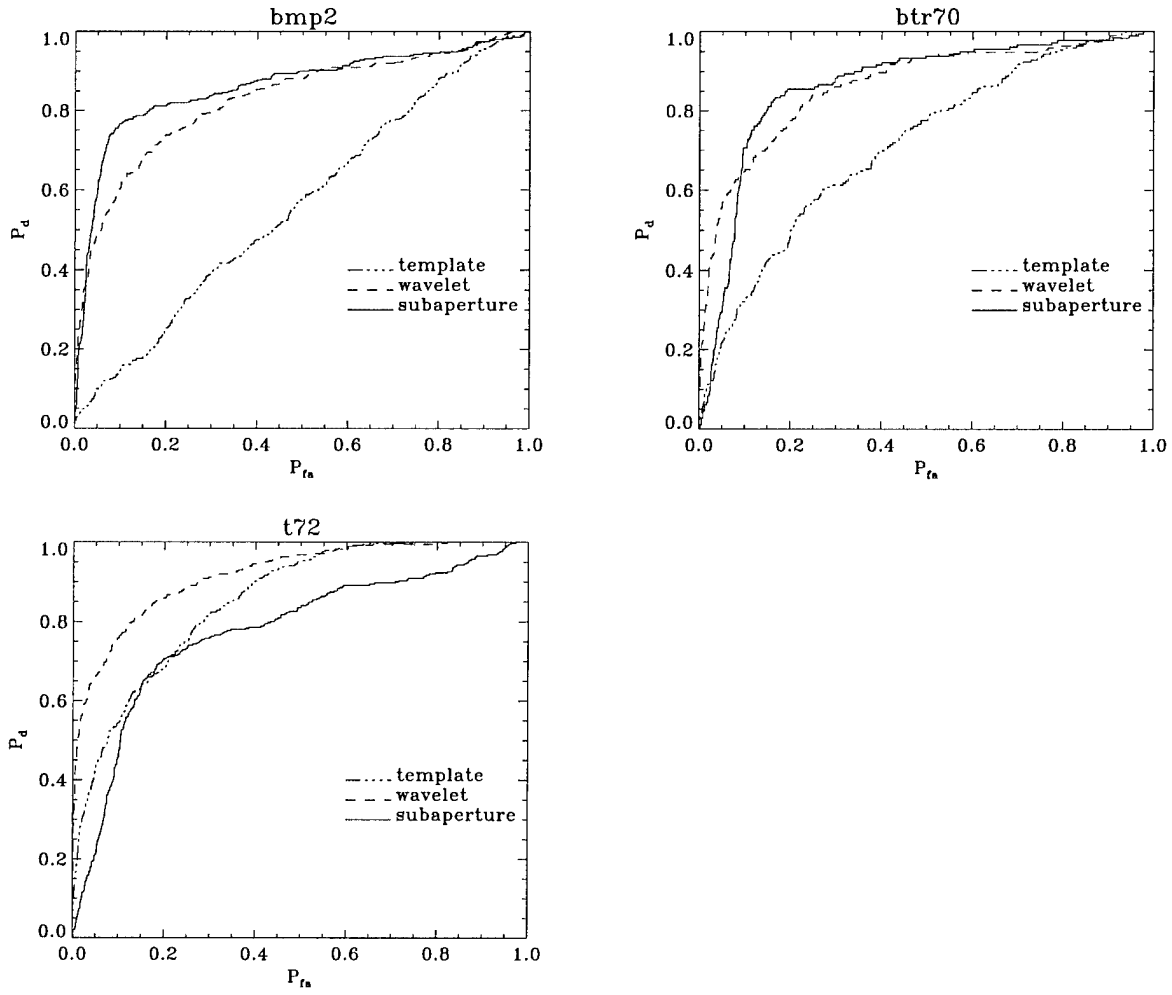


Figure 16: ROC detections curves using the subaperture feature set, DeBonet's steerable wavelet feature set, and template matching.

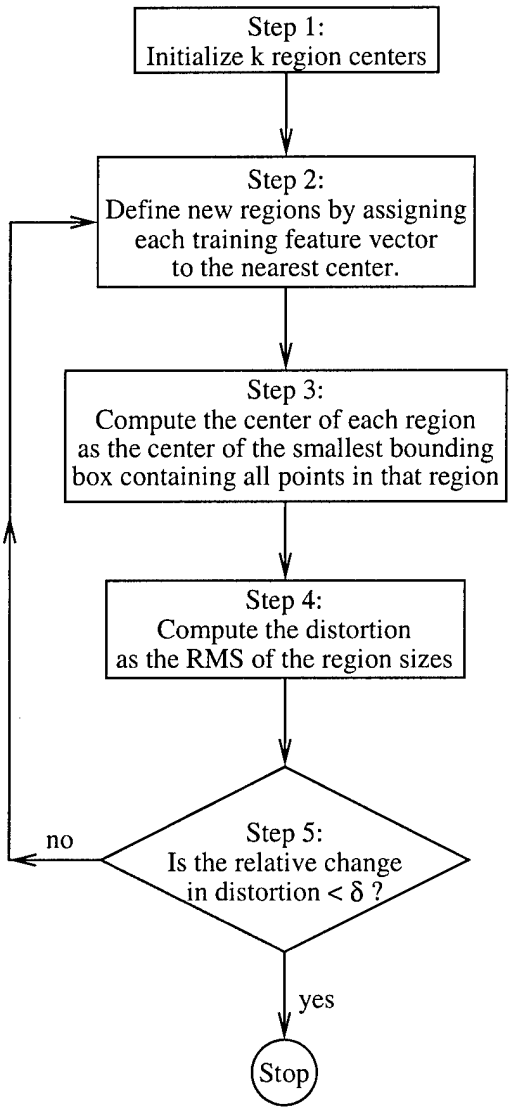


Figure 17: Flow diagram illustrating the clustering procedure used to generate a uniform sampling over the union of the supports of the training image pdf's.

4 Attributing Scatterer Anisotropy for Model Based ATR

Note *that this section includes work that was done jointly with Professor Randolph Moses and his graduate student Sinan Doogan at the Ohio State University. Professor Moses and his student were supported under both the DARPA IU and MEP-3 programs and the US Air Force Research Laboratory under grant F33615-979191020*

Scattering from man-made objects in SAR imagery often exhibit aspect and frequency dependences which are not well modeled by standard SAR imaging techniques. If ignored, these deviations may reduce recognition performance due to model mismatch, but when appropriately accounted for, these deviations can be exploited as attributes to better distinguish scatterers and their respective targets. Chiang and Moses[20] developed an ATR system that allows the study of performance under various scatterer attributions. Kim *et. al.*[21] examined a nonparametric approach for exploiting non-ideal scattering using a multi-resolution sub-aperture representation. Both of these works are extended here to examine the effect of anisotropic scattering attribution for model-based ATR. In particular, predicted and extracted peak scatterers are attributed with a discrete anisotropy feature. This feature can be obtained in a computationally efficient manner by performing a set of generalized log-likelihood ratio (GLLR) tests over a pyramidal sub-aperture representation. Furthermore, an approximate probabilistic characterization of the feature set allows for a natural inclusion into the approach of Chiang and Moses which will be used to evaluate the benefit of our attribution to the X-band MSTAR data and infer the phenomenology behind anisotropic scattering.

4.1 Introduction

Scatterers composing a target in SAR imagery often exhibit nonideal scattering in the form of aspect and frequency dependences. Standard SAR image formation ignores this variability resulting in unstable scintillating reflectivity estimates complicating the recognition problem. These deviations from the ideal point scattering model should not be viewed as a hindrance and approximated away, but instead, they should be seen as an attribute which can be used to distinguish scatterers and thus their respective targets.

This paper is an extension of two separate works presented at Aerosense 1999. Chiang and Moses[20] presented a full ATR system which allowed performance comparisons to be made between systems based on different feature attributes. It was used to demonstrate the improvement in ATR performance achieved by using models based on the Geometric Theory of Diffraction (GTD) with synthetic data. Kim *et. al.*[21] presented a classification technique that utilizes aspect dependence by learning these dependences in a nonparametric fashion on a multi-resolution pyramid of sub-apertures.

The work presented here uses the same sub-aperture structure introduced in by Kim *et. al.*, but simplifies the information conveyed into a single scalar parameter characterizing the azimuthal concentration of unimodal scattering. Motivated by canonical scattering models, we conjecture that knowledge of the azimuthal duration of a scatterer can be used to infer properties of its geometry. Scattering models such as the physical optics model or the Geometric Theory of Diffraction (GTD) predict that for many physically large scatterers there is an inverse relation between the size of the scatterer and the duration of its response in azimuth. Thus, knowing the anisotropy of a scatterer allows one to infer properties of the physical structure of the object under investigation thereby aiding the classification procedure. Incorporating the resulting anisotropy attribution into the feature based classifier and applying it to real SAR data allows us to study the utility of and phenomenology behind azimuthal anisotropy.

The remainder of this paper is organized as follows. Section 4.2 presents the multi-resolution sub-aperture pyramid used to represent the SAR data. Section 4.3 then describes the set of hypotheses that we will consider and develops the hypothesis tests on the sub-aperture pyramid. Section 4.4 describes the matching

algorithm that we use to evaluate our anisotropy attribution on collected SAR data. Section 4.5 presents experimental results demonstrating the utility of anisotropy attribution and discusses its underlying phenomenology. The paper concludes with a summary and discussion in Section 4.6.

4.2 Sub-aperture Analysis

The foundation of our analysis is the sub-aperture pyramid which we present in this section. This structure is motivated by the scattering physics involved in SAR and presents information in a way that allows for simple and intuitively reasonable hypothesis tests. Because of the linear structure of the aperture, we will associate it with an interval of the real line throughout this paper. In particular, the full-aperture will be denoted by the interval $[0, 1)$.

4.2.1 Definition

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 on a pyramidal structure. These sub-apertures will be used to represent both our set of candidate hypotheses and to form our reflectivity estimates. The prototypical sub-aperture covering that we will use throughout this paper is the half-overlapping half-aperture pyramid shown in the lower portion of Figure 18.

We take as a sub-aperture pyramid a set S of sub-apertures with the following structure. The set S is partitioned into smaller sets S_m which correspond to a particular degree of anisotropy. For reasons which will become apparent, we associate m with scale. S_0 refers to the set consisting of the largest sub-apertures, and S_M refers to the set of the smallest sub-apertures. A second subscript on S denotes a specific sub-aperture at the given scale. To obtain the necessary structure on the sub-aperture pyramid for what follows later, the following conditions are imposed on S:

(S1) $\forall S_{m,i}$, $S_{m,i} = [a, b)$ for some $0 \leq a < b \leq 1$,
(S2) $\forall S_{m,i}$ with $m \geq 1$, $\exists S_{m-1,j}$ such that $S_{m,i} \subset S_{m-1,j}$, and
(S3) $\forall S_{m,i}, \exists$ a partition $\mathcal{P}(m,i) \subset S_M$ of $S_{m,i}$.

The first condition simply restricts the sub-apertures to be a single connected interval. This is motivated by our search for concentrated unimodal scattering in azimuth. The second condition asserts that each sub-aperture, except those in S_0, has a parent which contains it. This allows us to construct a telescopic hypothesis test on a tree which will not only afford computationally efficiency but also robustness. The third condition requires the existence of a partition of each sub-aperture by coarsest scale sub-apertures. This will allow for the set of measurements given by S_M to form a sufficient statistic for all the measurements in S. Herein, the term *sub-aperture pyramid* will always refer to one satisfying conditions (S1)-(S3).

Any sub-aperture can be used to form a SAR image. The images formed with smaller values of m have a finer cross-range imaging resolution because of their larger apertures. This is our motivation for associating scale with m. Each sub-aperture $S_{m,i}$ generates an associated reflectivity estimate $q_{m,i}$. The collection of reflectivity estimates from the sub-apertures in S_M is denoted as $\boldsymbol{q}_M$. The measured reflectivity $q_{m,i}$ is not normalized with respect to the aperture length, i.e.

$$q_{m,i} = \int_{S_{m,i}} a(s)\, ds \tag{15}$$

where $a(s)$ is the azimuthal response of the scatterer[8]. Thus, when interested in the normalized reflectivity estimate, one should divided $q_{m,i}$ by the sub-aperture length $L_{m,i} = \lambda(S_{m,i})$, where λ denotes Lebesgue measure.

4.2.2 Interpretation and Motivation

Different types and sizes of scatterers will yield different aspect dependencies. The motivation for using the sub-aperture pyramid is that it is expected to reveal distinguishing aspect dependences in the scattering. For example, a small metal sphere will have a strong response in all directions and thus produce a strong reflectivity estimate from each of the sub-apertures. However, as depicted in Figure 18, a flat plate produces a significantly stronger response when oriented broadside with respect to the radar as compared to off-broadside orientation. Thus, the reflectivity estimates will vary across the sub-apertures with the largest estimate coming from the sub-aperture oriented broadside to the plate. Furthermore, because various sized sub-apertures are used, the duration of the broadside flash is captured in this representation. In particular, while the sub-aperture is contained within the main-lobe of the response, the reflectivity estimate will be consistently large, but as the sub-aperture is expanded, the additional energy received will be relatively insignificant and result in a lower reflectivity estimate when normalized with respect to the sub-aperture length. While this framework will capture general degrees of anisotropy, it is not overly-sensitive to azimuthal dependencies in that slight deviations on the scatterer geometry are modeled. This relieves of the burdens associated with such models that requires an excessive number of parameters.

A slightly different viewpoint of this sub-aperture feature set comes from considering the cross-range resolution versus azimuthal[9] resolution trade-off. Recall that the cross-range resolution is inversely proportional to the aperture length. Thus, at lower levels of the sub-aperture pyramid, spatial resolution has been exchanged for azimuthal resolution, i.e. the ability to better observe anisotropic phenomena. This is the classic time–frequency resolution tradeoff in Fourier analysis, and each level of the pyramid represents the data under a particular cross-range–azimuth resolution. The presence of multiple resolutions is attractive because we expect the best representation for different objects to be nonunique as the importance between resolution in the two domains is balanced.

4.3 Anisotropic Scattering Models

Having presented the sub-aperture pyramid, we now proceed to formulate our hypothesis testing problem for anisotropy. The hypotheses will be drawn directly from the sub-aperture pyramid. Two models will be presented here. The first is a simple single scatterer model with an intuitive sufficient statistic. This test however is susceptible to the influence of neighboring scatterers. This motivates the second model which explicitly accounts for the contributions from neighbors. The tests presented in this section are for a fixed scattering location which we assume to be specified. These locations could come from a peak extraction process or a pre-specified grid of points to produce an image of anisotropy.

[8]By azimuthal response, we mean the 1-D cross-range uncompressed signal for a given down-range location. The signal $a(s)$ is assumed to have already been appropriately demodulated to have zero phase modulation for the inspected cross-range location.

[9]For clarity, we will use the term "cross-range" when referring to the cross-range in the spatial (image) domain, and we will exclusively use the term "azimuthal" when referring to the corresponding dimension in the sensor domain.

4.3.1 Single Scatterer Model

Each sub-aperture $S_{m,i}$ defines an associated scattering hypothesis $H_{m,i}$ over the aperture s via

$$H_{m,i}: \; a(s) = 1_{S_{m,i}}(s) \tag{16}$$

where $1_A(\cdot)$ denotes the indicator function over the set A. Thus, our hypotheses correspond to scattering responses that are uniform over the sub-aperture in question and zero elsewhere. Naturally, this is an idealization for anisotropic scattering, but because we are only looking for a general characterization of anisotropy, it will serve our purposes here. Although we will call this a test of anisotropy, the ideal isotropic scattering hypothesis is included in our hypothesis set if 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 features to test these hypotheses would be all the measured sub-aperture reflectivities $\{q_{m,i}\}$. From the definition of the $q_{m,i}$ in Eq. (15) and partition property (S3), it is sufficient to consider the subset $\boldsymbol{q}_M \subset \{q_{m,i}\}$ since all sub-aperture reflectivities $q_{m,i}$ can be computed from $\boldsymbol{q}_M$ by summing all the $\boldsymbol{q}_{M,j}$ which form a partition of $S_{m,i}$. Thus, we will take $\boldsymbol{q}_M$ as our feature vector. The value of this feature vector under hypothesis $H_{m,i}$ is $\boldsymbol{b}(m,i)$ whose j^{th} element is given by

$$\begin{aligned} b(m,i)_j &= \int_{S_{M,j}} 1_{S_{m,i}}(s)\; ds \\ &= \lambda(S_{M,j} \cap S_{m,i}), \end{aligned} \tag{17}$$

i.e. it is the portion of the response one expects to see over the j^{th} sub-aperture at scale M. We now define our scattering model conditioned on anisotropy hypothesis $H_{m,i}$ as the signal plus noise model,

$$q_{M,j} = \int_{S_{M,j}} A 1_{S_{m,i}}(s) + \eta(s)\; ds, \tag{18}$$

where A is the scattering amplitude of the signal and $\eta(s)$ is circularly complex white Gaussian noise with spectral density σ^2. This leads to the model

$$\boldsymbol{q}_M = A\boldsymbol{b}(m,i) + \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, 2\sigma^2\Lambda), \tag{19}$$

where Λ is the noise covariance structure inherited from the sub-aperture pyramid. The noise in the measured reflectivities in Eq. (19) are characterized as zero-mean circularly complex Gaussians with covariances dictated by the amount of sub-aperture overlap. The elements of its covariance matrix Λ are given by $[\Lambda]_{i,j} = \lambda(S_{M,i} \cap S_{M,j})$, which for the half-overlapping half-aperture pyramid in Figure 18 is

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

To classify the anisotropy of a scatterer from our vector of sub-aperture measurements $\boldsymbol{q}_M$, we apply a log-likelihood ratio (LLR) test to the model in Eq. (19) where each log-likelihood is compared to the full-aperture hypothesis. Because there is the unknown reflectivity parameter A, we actually use a generalized

LLR (GLLR) test where for each hypothesis, we take A to be its maximum likelihood (ML) estimate under that hypothesis. Thus, for $H_{m,i}$, we take $\hat{A} = q_{m,i}/L_{m,i}$. This produces the GLLR

$$\ell_{m,i} = \frac{1}{4\sigma^2}\left[\frac{1}{L_{m,i}}|q_{m,i}|^2 - |q_{0,0}|^2\right]. \tag{20}$$

Thus, the most likely sub-aperture in this case is the one whose average energy is largest. We note the similarity here to the approach taken by Chaney *et. al.*[15] in which they replace, within the image, the standard reflectivity estimate with the maximum sub-aperture reflectivity estimate $q_{m,i}/L_{m,i}$, thus using normalized reflectivity (instead of normalized energy) as their criterion for choosing anisotropy. Their approach however is based on intuitive arguments and not a derived statistic.

Though simple and intuitive, the GLLR in Eq. (20) is susceptible to the effects of close proximity neighboring scatterers which are not included in our model in Eq. (18). Recall that the images formed by the smaller sub-apertures have a coarser resolution. Thus, if a scatterer were located outside the finest resolution cell but within a coarser resolution cell, then the finest resolution reflectivities $q_{0,j}$ would be 0, but $q_{m,i}$ would be large if the resolution cell associated with scale m included the scatterer. We illustrate this with the example shown in Figure 19. Here a scatterer with amplitude A is located outside the finest resolution cell of size δ but is contained within the coarser resolution cell of size 2δ associated with half-aperture estimates. If this scatterer is isotropic, then its response is the complex exponential illustrated. Integrating over the full aperture gives a 0 reflectivity estimate as expected, but integrating over a half aperture produces a normalized reflectivity with magnitude $A/\sqrt{2}$. Eq. (20) would then classify the center of the resolution cell as anisotropic, even though no scatterer is present.

The problem above is a consequence of not modeling the influence of neighboring scatterers. One of the ways in which the neighboring scatterer manifests itself is through its corruption of the estimated reflectivity as the size of the resolution cell varies. Choosing $\hat{A} = q_{m,i}$ is the maximum likelihood reflectivity estimate for the resolution cell associated with $S_{m,i}$. Alternatively, we may instead choose the best reflectivity estimate constrained to lie in the finest resolution cell which is the full aperture estimate $q_{0,0}$. Choosing $\hat{A} = q_{0,0}$ for all hypotheses can be shown to produce the GLLR statistic

$$\ell_{m,i} = \frac{1}{4\sigma^2}\left[\frac{1}{L_{m,i}}|q_{m,i}|^2 - \frac{1}{L_{m,i}}|q_{0,0} - q_{m,i}|^2 - |q_{0,0}|^2\right]. \tag{21}$$

This statistic is identical to that in Eq. (20) except for the extra term comparing the reflectivity estimates $q_{0,0}$ and $q_{m,i}$. Recall that Eq. (20) compared the average energy in a sub-aperture to the full-aperture. This new GLLR accounts for the average energy outside the current sub-aperture as well. Viewed differently, under our scattering model in Eq. (18), the values of $L_{m,i}q_{0,0}$ and $q_{m,i}$ should simply be noisy perturbations of each other. The new term penalizes when this is not the case. Under the example in Figure 19, since the the contribution of each of the half-apertures would be the same, the GLLR is equal to zero for all hypotheses, which is reasonable, since there is no underlying scattering at the focused location.

4.3.2 Multiple Scatterer Model

The modification in Eq. (21) addresses the problem when a neighboring scatterer is isotropic, however, when the neighboring scatterer is anisotropic, problems such as that illustrated by the example in Figure 19 can still arise as its contribution will not integrate out over the full-aperture estimate. To alleviate this problem we generalize the model in Eq. (18) to account for multiple scatterers.

First, we must generalize the sub-aperture scattering model to account for the modulations produced by neighboring scatterer. Again, we assume that the sub-apertures have been formed while being focused on cross-range location y_0. The possibility of other scatterers are considered at discrete locations $y_0 + k\Delta_p$, where Δ_p specifies a cross-range sampling resolution. To model the effect of a scatterer at location $y_0 + k\Delta_p$ on the measurements at location y_0, we simply modulate the focused response $1_H(\cdot)$ to account for the shift in the image domain. The observed effect over the sub-apertures is then given by

$$b^k(m,i)_j = \int_{S_{M,j}} e^{j2\pi \frac{k\Delta_p}{\Delta_r} s} 1_{H_{m,i}}(s)\ ds \tag{22}$$

where Δ_r is the null-to-null resolution associated with the full aperture. Incorporation of the neighboring scatterers is now modeled via superposition, i.e.

$$\boldsymbol{q}_M = \sum_k A_k \boldsymbol{b}^k(H_k) + \boldsymbol{\varepsilon} \tag{23}$$

where the noise model is the same as in Eq. (19). The summation over k in Eq. (23) should be over a range that at least includes all scatterers contained in the largest resolution cell. Thus, if L_* is the smallest aperture length, then we need to consider $k \in \{-K, \ldots, K\}$ where

$$\begin{aligned} K\Delta_p &\geq \frac{\Delta_r}{2L_*} = \text{size of coarsest resolution cell} \\ K &= \left\lceil \frac{\Delta_r}{2L_*\Delta_p} \right\rceil \end{aligned} \tag{24}$$

and $k = 0$ corresponds to the resolution cell under investigation. Rewriting Eq. (23) as a matrix equation we get

$$\begin{aligned} \boldsymbol{q}_M &= \left[\boldsymbol{b}^{-K}(H_{-K}) \ \mid \ \cdots \ \mid \ \boldsymbol{b}^K(H_K)\right] \begin{bmatrix} A_{-K} \\ \vdots \\ A_K \end{bmatrix} + \boldsymbol{\varepsilon} \\ &= B\mathbf{A} + \boldsymbol{\varepsilon} \end{aligned} \tag{25}$$

where B and $\mathbf{A}$ are appropriately defined. Thus, we can use weighted least squares (WLS) to implicitly estimate the values of the interfering A_k and account for their contribution to $\boldsymbol{q}_M$. In order to have our least squared error minimization correspond to ML, we need to have our inner product effectively whiten the noise. This is accomplished by using the inner product weighted by the inverse of the noise covariance, i.e. $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \frac{1}{2\sigma^2} \boldsymbol{u}^T \Lambda^{-1} \boldsymbol{v}$. We estimate the hypothesis as the one which minimizes the norm of $\boldsymbol{\varepsilon}$ when the WLS estimate of $\mathbf{A}$ is used. The ML estimate for $\mathbf{A}$ is obtained from this model as

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \left\{ \|\boldsymbol{\varepsilon}\|^2_{\Lambda^{-1}} \right\} = (B'\Lambda^{-1}B)^{-1} B'\Lambda^{-1}\boldsymbol{q}_M \tag{26}$$

which simplifies to Eq. (20) in the case of limiting the model order to $K = 0$.

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 a scatterer associated with it. For the initial work presented in this paper, we will restrict ourselves to hypotheses where only the $k = 0$ (the focused resolution cell) is allowed to vary.

Thus, our hypothesis space is effectively still $\mathcal{H}$. We note that this constraint will be relaxed in the future in order to appropriately take into account the scattering of neighboring scatterers.

To classify the anisotropy of the $k = 0$ scatterer, we use ML, i.e. we choose the sub-aperture hypothesis which minimizes the weighted norm of $\boldsymbol{\varepsilon}$ in Eq. (23)

$$\begin{aligned} \|\boldsymbol{\varepsilon}\|^2_{\Lambda^{-1}} &= \|\boldsymbol{q}_M - B\hat{\mathbf{A}}\|^2_{\Lambda^{-1}} \\ &= \frac{1}{2\sigma^2}\boldsymbol{q}'_M \left[\Lambda^{-1} - \Lambda^{-1}B(B'\Lambda^{-1}B)^{-1}B'\Lambda^{-1}\right]\boldsymbol{q}_M \end{aligned} \tag{27}$$

where the weighting matrix in the last line is independent of the data and thus can be precomputed for each of the candidate hypotheses.

The model in Eq. (23) is too unconstrained for the test given by the minimization in Eq. (27) to work. In particular, if one chooses K sufficiently large to account for all scatterers in the coarsest resolution cell, then the model order is greater than the number of sub-aperture measurements and $\boldsymbol{\varepsilon}$ can be made zero for all hypothesis of $k = 0$. Thus, we need to regularize the model. In order for the error $\boldsymbol{\varepsilon}$ to be made small under the incorrect model, many of the values in $\mathbf{A}$ generally have to be made unreasonably large and result in an unrealistic scenario. Thus, we impose a 2-norm regularization penalty in the estimation of $\mathbf{A}$. In particular, instead of minimizing the weighted squared error to estimate $\mathbf{A}$, we take

$$\begin{aligned} \hat{\mathbf{A}} &= \arg\min_{\mathbf{A}} \left\{\|\boldsymbol{\varepsilon}\|^2_{\Lambda^{-1}} + \gamma \mathbf{A}'R\mathbf{A}\right\} \\ &= (B'\Lambda^{-1}B + \gamma R)^{-1}B'\Lambda^{-1}\boldsymbol{q}_M \\ &= P\boldsymbol{q}_M \end{aligned} \tag{28}$$

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

$$R = \boldsymbol{I} - \boldsymbol{e}_K\boldsymbol{e}'_K = \text{diag}(1, \ldots, 1, 0, 1, \ldots, 1).$$

This produces the following value for the weighted error norm as

$$\begin{aligned} \|\boldsymbol{\varepsilon}\|^2_{\Lambda^{-1}} &= \|\boldsymbol{q}_M - B\hat{\mathbf{A}}\|^2_{\Lambda^{-1}} \\ &= \frac{1}{2\sigma^2}\boldsymbol{q}'_M \left[\Lambda^{-1} - 2\Lambda^{-1}BP + P'B'\Lambda^{-1}BP\right]\boldsymbol{q}_M \end{aligned} \tag{29}$$

which can be used for our hypothesis test.

4.3.3 Telescopic Testing

The sub-aperture pyramid which we use to form our measurements and base our hypotheses is convenient not only for providing anisotropy information, but also for providing an efficient means of performing the hypothesis tests. We can obtain an efficient approximation to the test by evaluating only a small subset of the candidate hypotheses. Due to the nested structure of the sub-apertures (condition (S2) in Section 4.2), we can perform the tests in a telescopic fashion by traversing down the tree of sub-apertures as depicted in Figure 20. We expect the likelihoods to increase as the hypothesized sub-apertures "shrink down to" the correct sub-aperture, and then to decrease as the hypothesized sub-apertures "shrink beyond" the correct sub-aperture. This motivates performing the hypothesis test in the following manner:

1. Start with the set of largest sub-aperture(s) at scale $m = 0$. Find the most likely hypothesis at that scale and denote it as H_{0,i^*}.

2. Consider those hypotheses at scale $n = m + 1$ for which $S_{n,j} \subset S_{m,i^*}$. Find the one which has the highest likelihood and denote it as H_{n,j^*}.

3. If the parent is more likely (i.e. $\gamma_{m,i^*} > \gamma_{n,j^*}$), then stop and return H_{m,i^*} as the estimated hypothesis.

4. If $m = M$, we are at the bottom of the tree so stop and return H_{n,j^*} as the estimated hypothesis. Otherwise, set $m = n$ and $i^* = j^*$ and goto step 2.

The intuition described above can be justified under the single scattering models. In particular, consider the expected value of $\ell_{n,j}$ when the true hypothesis is $H_{m,i}$ and the proportion of overlap between $S_{n,j}$ and $S_{m,i}$ is given by $\alpha = \frac{\lambda(S_{n,j} \cap S_{m,i})}{\lambda(S_{m,i})}$. For Eq. (20) in which $\hat{A} = q_{m,j}$ is used, the expected value is

$$\begin{aligned} \mathrm{E}\Big[\ell_{n,j} \,\Big|\, H_{m,i}\Big] &= \mathrm{E}\left[\frac{1}{L_{n,j}}|q_{n,j}|^2 - |q_{0,0}|^2 \,\middle|\, H_{m,i}\right] \\ &= \left(\frac{\alpha^2}{L_{n,j}} - 1\right)|A|^2 + (2L_{n,j} + 1)\sigma^2 \\ &\approx \left(\frac{\alpha^2}{L_{n,j}} - 1\right)|A|^2 \end{aligned} \tag{30}$$

where the approximation is for high SNR. From this, we see the intuitive behavior described above. When $S_{n,j}$ contains $S_{m,i}$, the overlap is $\alpha = 1$ and the GLLR increases with decreasing $L_{n,j}$. When the sub-aperture becomes too small, i.e. $S_{n,j} \subset S_{m,i}$, the overlap is $\alpha = L_{n,j}/L_{m,i}$, and thus the expected GLLR decreases when the sub-aperture becomes too small.

For Eq. (21) in which $\hat{A} = q_{0,0}$, the expected value of the GLLR is

$$\begin{aligned} \mathrm{E}\Big[\ell_{n,j} \,\Big|\, H_{m,i}\Big] &= \mathrm{E}\left[\frac{1}{L_{n,j}}|q_{n,j}|^2 - \frac{1}{L_{n,j}}|q_{0,0} - q_{n,j}|^2 - |q_{0,0}|^2 \,\middle|\, H_{m,i}\right] \\ &= 2\left(\frac{\alpha - 1}{L_{n,j}}\right)|A|^2 + \left(\frac{1}{L_{n,j}} - 1\right)(|A|^2 - 2\sigma^2). \end{aligned} \tag{31}$$

Again, we see the intuitive behavior described above. In particular, when $S_{n,j}$ contains $S_{m,i}$, the overlap $\alpha = 1$ and the GLLR increases with decreasing $L_{n,j}$. When the sub-aperture becomes too small, i.e. $S_{n,j} \subset S_{m,i}$, the overlap is $\alpha = L_{n,j}/L_{m,i}$, and the GLLR can be written as

$$\mathrm{E}\Big[\ell_{n,j} \,\Big|\, H_{m,i}\Big] = 2\frac{|A|^2}{L_{m,i}} - \frac{|A|^2 + 2\sigma^2}{L_{n,j}} - |A|^2 + 2\sigma^2$$

which is a decreasing function of $L_{n,j}$ for strong scatterers, i.e. $|A|^2 \gg 2\sigma^2$.

Because the difficult form of our regularized multiple scatterer log-likelihood given in Eq. (29), we have not yet shown the same pattern for this extended test, however intuition leads us to believe it holds here too.

4.3.4 Boxcar Model Deviations

The hypothesis set defined in Eq. (16) are simplified models to which real scatterers will not exactly correspond. One may question whether deviations from this model may drastically effect our hypothesis tests. For example, if the scattering has a $\mathrm{sinc}(\cdot)$ like dependence in azimuth, then the sidelobes will have a large response for a strong scatter and cannot be well modeled as background noise. To address this issue, we incorporate deviations from the boxcar response into our model. In particular, we start by assuming the underlying scattering pattern has been perturbed by white Gaussian noise. For the single scatterer model, this changes the model response in Eq. (17) to

$$\tilde{b}(m,i)_j = \int_{S_{M,j}} 1_{S_{m,i}}(s) + \nu(s)\, ds$$

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

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

This results in the new measurement model as

$$\boldsymbol{q}_M = A\boldsymbol{b}(m,i) + \mathbf{w}, \text{ with } \mathbf{w} \sim \mathcal{N}\big(\mathbf{0}, 2(|A|^2\rho^2 + \sigma^2)\Lambda\big).$$

Thus, we have essentially the same model as in Eq. (19) except that the variance of the noise now depends affinely on the square-magnitude of the underlying scatterer. The effect in Eqs. (20) and (21) is a simple scaling of the GLLR's of *all* the hypotheses. For the multiple scatterer model, the extension is similar and has the same result.

4.4 Bayes Classification

With a method for anisotropy attribution in hand, we describe in this section a classifier based on Bayesian probability theory. This matcher allows us to evaluate the utility and explore the phenomenology of anisotropy in SAR by incorporating the labeled anisotropy of scatterers into the feature set. A more thorough description of this classifier can be found in the paper by Chiang and Moses[20].

4.4.1 Classification Problem Statement

The Bayes matching problem is given as follows. At the input to the classifier stage, we are given a set of n feature vectors $Y = [Y_1, Y_2, \ldots, Y_n]^T$ extracted from a measurement, and for each candidate hypothesis[10] $H \in \mathcal{H}$ we are given a set of m predicted feature vectors $X = [X_1, X_2, \ldots, X_m]^T$ (where m may vary with H). We wish to find the hypothesis whose posterior likelihood of the observed features, Y, is maximum. From Bayes' rule, we have

$$P(H|Y) = \frac{f(Y|H,n)P(n|H)P(H)}{f(Y|n)P(n)}.$$

[10]The hypotheses in this section correspond to possible vehicle classifications. They are not the anisotropic hypotheses used in Section 4.3.

Since the denominator does not depend on hypothesis H, the MAP decision is found by maximizing the numerator $f(Y|H,n)P(n|H)P(H)$ over $H \in \mathcal{H}$. In this paper, we assume the priors $P(H)$ and $P(n|H)$ are uniform, so we need only compute $f(Y|H,n)$.

We incorporate uncertainty in both the predicted and extracted feature sets, and assume the predict and extract uncertainties are conditionally independent. This gives[22]

$$f(Y|H,n) = \int f(Y|\hat{X},H,n) f(\hat{X}|H,n)\, d\hat{X} \tag{32}$$

where $f(\hat{X}|H,n)$ models the feature prediction uncertainty, and $f(Y|\hat{X},H,n)$ models feature extraction uncertainty.

The computation of $f(Y|\hat{X},H,n)$ requires a correspondence between the elements of Y and $\hat{X}$, or equivalently between Y and X. We consider two correspondence mappings. The first is a probabilistic many-to-many map, in which we assume that

$$f(Y|X) = f(Y|H) = \prod_{j=1}^{n} f(Y_j|X,H) = \prod_{j=1}^{n} \left[B f_{FA}(Y_j) + \sum_{i=1}^{m} D_i(H) f(Y_j|X_i,H) \right] \tag{33}$$

where λ is the average number of false alarms features present, f_{FA} models false alarm probability of a particular feature vector, $P_k(H)$ is the detection probability of the i^{th} predicted feature under hypothesis H, $B = \lambda/[\lambda + \sum_{k=1}^{m} P_k(H)]$ is the probability that an extracted feature is a false alarm, and $D_i(H) = (1-B)P_i(H)/[\sum_{k=1}^{m} P_k(H)]$ is the probability that an extracted feature comes from the i^{th} predicted feature.

The second mapping considered is a deterministic one-to-one map, in which the correspondence is assumed to be a deterministic nuisance parameter and the match score is maximized over the correspondence. In this case the likelihood score is given by

$$\begin{aligned} f(Y|\Gamma,H,n) \;=\; & \left\{ P(n_{FA} \text{ false alarms}) \prod_{\{j:\Gamma_j=0\}} f_{FA}(Y_j) \right\} \\ & \cdot \left\{ \prod_{\{j:\Gamma_j=i>0\}} P_i(H) \right\} \cdot \left\{ f(Y_j|\Gamma_j=i,H,n) \prod_{\{i:\Gamma_j\neq i,\forall j\}} (1-P_i(H)) \right\} \end{aligned} \tag{34}$$

where Γ defines the feature correspondences, including the n_{FA} extracted features that correspond to no predicted features (denoted $\{j : \Gamma_j = 0\}$) and the predicted features that correspond to no extracted features (denoted $\{i : \Gamma_j \neq i, \forall j\}$).

For the case that $P(n_{FA}$ false alarms) obeys an exponential rule $P(n_{FA}$ false alarms) $= ce^{-\beta n_{FA}}$ for constants c and β, the search for the correspondence that maximized the above likelihood can be efficiently implemented.[22]

To implement either Eq. (33) or (34), we require a probability model for $f(Y|\hat{X},\Gamma,H,n)$. We assume that the uncertainties of the X_i are conditionally independent given H, and that the uncertainties of the Y_j are conditionally independent given H, X, and n. This yields

$$f(Y|\Gamma,H,n) = \prod_{j=1}^{n} f(Y_j|\Gamma,H,n).$$

Each extracted feature Y_j either corresponds to a predicted feature or is a false alarm. If Y_j is a false alarm, we assign $\Gamma_j = 0$, and we model the feature attribute as a random vector with probability density function

$$f(Y_j|\Gamma_j = 0, H, n) = f_{FA}(Y_j).$$

If Y_j corresponds to a predicted feature X_i, we write $\Gamma_j = i$ (for $i > 0$) and compute the feature match score from Eq. (32). In particular, from Eq. (32) it follows that for $i > 0$,

$$f(Y_j|\Gamma_j = i, H, n) = \int f(Y_j|\hat{X}_i, H, n) f(\hat{X}_i|X_i, H)\, d\hat{X}_i. \tag{35}$$

For the special case of Gaussian uncertainties, we have $f(Y_j|\hat{X}_i, H, n) \sim \mathcal{N}(\hat{X}_i, \Sigma_e)$, and $f(\hat{X}_i|X_i, H, n) \sim \mathcal{N}(X_i, \Sigma_p)$, so from Eq. (35) we obtain

$$f(Y_j|\Gamma_j = i, H, n) = f(Y_j|X_i, H, n) \sim \mathcal{N}(X_i, \Sigma_p + \Sigma_e). \tag{36}$$

Similarly, for features whose attributes are discrete-valued, the likelihood is the sum

$$P(Y_j|X_i, H, n) = \sum_{\hat{X}_i} P(Y_j|\hat{X}_i, H, n) P(\hat{X}_i|X_i, H, n). \tag{37}$$

4.5 Results

Public release MSTAR data is used for the results presented here. These images have a resolution of $0.3m$ in both range and cross-range. The transmitted signal had a bandwidth of $0.591GHz$ and a center frequency of $9.60GHz$.

All of the results in this section are based on the three-level half-overlapping half-aperture pyramid depicted in Figure 18. The multiple scatterer model will be used to characterize anisotropy. The number of neighboring scatterers considered is set by $K = 6$. The value of the regularization parameter on neighboring reflectivities is set to $\gamma = 0.5$. We incorporate a bias in our anisotropy test towards full-aperture scattering. In particular, to be declared anisotropic, an anisotropic likelihood must be at least twice the full-aperture likelihood. The purpose of this higher threshold is to aid in protecting against the effects of neighboring scatterers whose reflectivities may not have been estimated exactly, and will thus induce a modulation across the aperture which can be mistaken as anisotropy.

This paper is based on the idea of detecting anisotropic scattering in SAR imagery. To illustrate the anisotropy assignments made by our model, we show in Figure 21 the results for a BMP2 (serial number c21) at $0°, 20°, 40°, 60°$, and $80°$ azimuths with a $17°$ depression. Even though the aperture associated with this data set is relatively small (about $3°$), we note that we are still able to detect anisotropic scattering. In particular, it usually appears to be associated with the turret, barrel, or leading edge of the tank. We make particular note of the classifications at the $0°$ azimuth. Here we see many clutter pixels being classified as anisotropic. The cause of this error is the unmodeled behavior of neighboring anisotropic scatterers. Recall that in our current formulation, we only consider the possibility of neighboring scatterers which are isotropic. However, the front edge of the tank generates a strong anisotropic response which is not accurately captured in our current model. Extending our model to account for such scattering should alleviate problems such as this.

The images in Figure 21 show that scatterers are being classified as anisotropic, however it does not convey how useful that information is in characterizing targets or explaining phenomenology. To address

these issues, we consider empirical confusion matrices of anisotropy based upon the MSTAR data set. In particular, we consider the data set composed of the following vehicles: 2S1 (b01), BMP2 (c21), BRDM2 (E-71), D7 (92v13015), T72 (132), ZIL131 (E12), and ZSU23-4 (d08) where truth is taken to be the empirical results from 17° depression data and test data is taken from the 15° depression data at the same azimuth as the truth. For each pair of training and testing images at the same azimuth, a set of peaks 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" will refer to one computed in this fashion.

The confusion matrix for this data set is given in Table 2. One noticeable property from the confusion matrix is that regardless of the conditioned training anisotropy, the full-aperture hypothesis is the most likely testing anisotropy. We give the following reasons for this. First, recall that we bias our anisotropy decision towards the full-aperture hypothesis which partially accounts for this. The bias in the confusion matrix may also be attributed to incorrect correspondences. It is widely believed that anisotropic scatterers are less stable than isotropic scatterers and therefore are not always extracted as peaks[23]. If an anisotropic training scatterer is not extracted in the testing data and there is a nearby isotropic scatterer, then our correspondence will incorrectly match the two. We also note that there is the unmodeled dependence on depression.

Another prominent aspect of this confusion matrix is the apparent independence of testing anisotropy and training anisotropy which would support the argument that anisotropy is not a stable feature. Under the model that anisotropy is caused by irresolvable interfering scatterers, this is reasonable especially when one considers that the depression angle has been changed. However, this result counters intuition gained from the canonical scatterer model. To further evaluate the source and stability of anisotropic phenomena, we partition the data set into those images at a near-cardinal azimuth ($\pm 2.5°$ of a cardinal angle) and those at off-cardinal azimuths. The motivation being that for near-cardinal azimuths, we expect the influence of canonical scatterers to be most pronounced because of the natural rectangular shape of vehicles. Thus, near-cardinal azimuths should exhibit canonical anisotropic scattering associated with flat plates and other large simple scatterers oriented orthogonal to the impinging radar signal.

The confusion matrices for the near-cardinal and off-cardinal azimuths are given in Table 3. Here, we see striking differences. There is still a tendency to favor the full-aperture hypothesis, which we explain by the same reasoning as for the confusion matrix in Table 2. One significant difference in the near-cardinal confusion matrix from the other two is that there is a noticeable presence along the diagonal, signaling that anisotropy is more stable at these near-cardinal angles as expected for canonical scatterers. The off-cardinal confusion matrix however shows that training anisotropy is independent of testing anisotropy. This leads us to believe that there are at least two fundamental sources of anisotropy. The first is canonical scattering which dominates at cardinal azimuths and not much at other azimuths. The second is an unstable source of anisotropy which is more commonly exhibited at off-cardinal azimuths. A likely candidate for this unstable anisotropy is the scintillation produced by irresolvable interfering scatterers. Anisotropy arising from such interference is highly variable and changes unpredictably with depression, which may account for lack of correlation in the anisotropy classifications between the 15° and 17° data in the off-cardinal confusion matrix.

To explore how anisotropy attribution might help the recognition problem, we use anisotropy as a feature in the matcher described in Section 4.4. The confusion matrices in Table 3 are used to characterize uncertainty in anisotropy. Our experiments involve detection of the BMP2 and T72 from a test set composed of BMP2's, T72's, T62's, and BTR70's. In particular, for the predict data, we use peak extractions from the BMP2 (c21) and T72 (132) at a 17° depression. The extract data consists of the following vehicles at a 15° depression angle: BMP2 (9563 and 9566), T72 (812 and s7), BTR70 (c71), and T62 (a51). For the BMP2

Table 2: Anisotropy confusion matrix for 2S1, BMP2, BRDM2, D7, T72, ZIL131, and ZSU23-4.

Training \ Testing	full-aperture	1/2-aperture	1/4-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

Table 3: Anisotropy confusion matrix for vehicles at near-cardinal and off-cardinal angles.

	Near-cardinal			Off-cardinal		
Training \ Testing	full-ap.	1/2-ap.	1/4-ap.	full-ap.	1/2-ap.	1/4-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
quarter-ap.	0.61	0.09	0.30	0.83	0.12	0.05

detection statistic, we compare the likelihood ratio of the test vehicle under the BMP2 model to the T72 model, where we are treating the T72 as our model for "non-BMP2" vehicles. Similarly, we take the reciprocal for the detection statistic of the T72 letting BMP2's serve as the model for "non-T72" scatterers. We recognize that modeling the "other" class with a single vehicle is simplistic and crude, however it is not our current goal to build a full classifier, but to setup a framework where we can study anisotropic phenomena. The T62 and BTR70 are used in the testing set because they are well known to be difficult confusers. The resulting ROC's are displayed in Figure 22 for three different sets of scatterer features:

(F1) location
(F2) location and anisotropy
(F3) location and anisotropy (while restricting predict scatterers to be full-aperture).

Each feature set uses the top 10 amplitude scatterers, except the third set which uses the top 10 amplitude scatterers which are declared to be full-aperture in the predict stage. The motivation behind the conditioning in (F3) is that if anisotropic scatterers are unstable, then they are unlikely to match in the extract data, so we remove them from consideration.

At first glance of the ROC's in Figure 22, it appears that all the tests perform equally well and the anisotropy is not useful as an attribution. However, these vehicles contain quite complex scattering phenomena and many of the anisotropy declarations may be due to volumetric interference between irresolvable scatterers which would change unpredictably with depression. With this in mind, we examine the ROC's for the test vehicles at near-cardinal ($\pm 2.5°$) azimuths. These ROC's are shown in Figure 23. Although there is less statistical significance in these numbers due to the relatively small number of test vehicles at these orientations, there does appear to be a separation between each of the tests. The feature set using both location and anisotropy appears to perform the best which is what we would expect since the scattering at near-cardinal azimuths is heavily influenced by canonical scatterers. The worst performer of the three feature sets is (F3) which uses location and anisotropy, but only considers predicted scatterers which are full-aperture. Thus, this test is discarding the anisotropic scatterers it observed on the model, which is valuable information since the anisotropy exhibited from these canonical scatterers should be stable.

4.6 Summary and Discussion

We have proposed a general characterization of anisotropy based on a sub-aperture pyramid. The sub-aperture pyramid generates a tree of multi-resolution images at a variety of cross-range versus azimuthal resolution trade-offs allowing for the detection of anisotropic phenomena. With each sub-aperture in the pyramid, we associate a hypothesis that the azimuthal scattering is confined to and uniform over that sub-aperture. This then leads to a sequence of hypothesis tests to classify the anisotropy for a pixel which can be approximated with an efficient pruning algorithm due to the tree-structure over the sub-apertures.

This characterization of anisotropy allows us to explore the underlying phenomenology of anisotropic scattering. In particular, our results show that while apparent at all orientations, there seems to be markedly different sources for anisotropy. At near-cardinal azimuths, anisotropic scatterers are stable as we would expect under canonical scattering models. However, at off-cardinal angles, anisotropy is erratic and difficult to predict from a different depression. This suggests that there is a different source of anisotropy at these intermediate azimuths. A likely candidate for this anisotropy is the scintillating scattering produced by volumetric scattering. Such a group of irresolvable scatterers would exhibit anisotropic behavior due to their interference and would be unpredictable with changes in depression like we observed in our experiments.

As demonstrated by the ROC curves, these different sources of anisotropy will need to be addressed separately in order to fully utilize the information contained in each. The classification approach here used here is useful at near-cardinal angles where anisotropy is stable, but not at off-cardinal angles. The sub-aperture models used here are in their elementary stages and as they develop, they should further aid in the studying of anisotropic scattering in SAR. Even though they are motivated by canonical scattering, they detect anisotropic behavior regardless of the source and can be used to study the phenomenon in general.

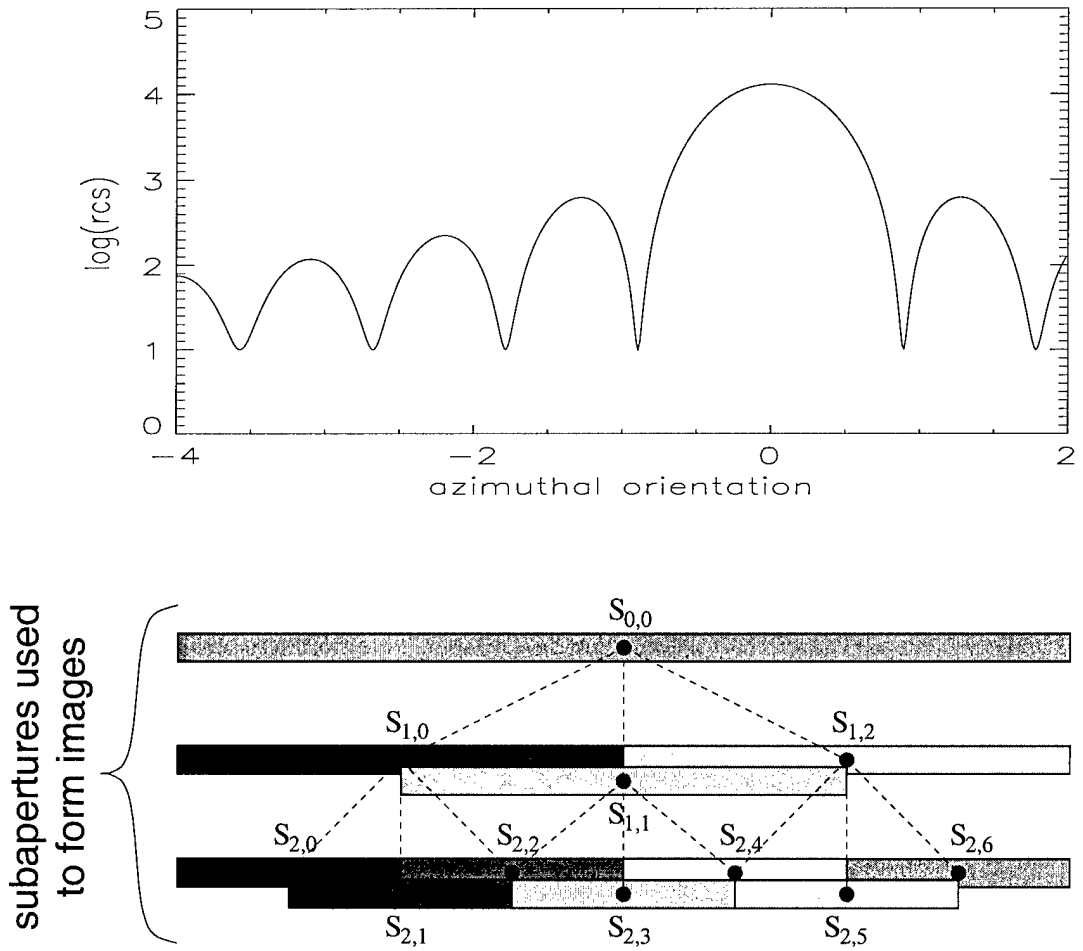


Figure 18: The response of a $.5m \times .5m$ flat plate and a depiction of the reflectivity estimate for each of the sub-apertures. Lighter shaded sub-apertures convey larger reflectivity estimates.

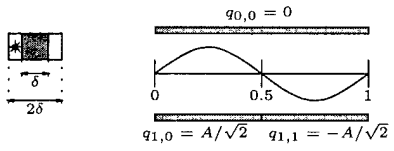


Figure 19: Illustration of a scatterer not in the finest resolution cell that can produce a false anisotropy classification in Eq. (20).

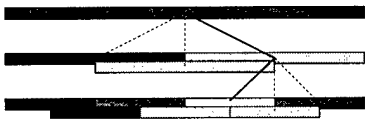


Figure 20: Illustration of how the anisotropy testing can be done in a decision directed fashion by starting with the largest aperture and at each scale, inspecting only the children of the most likely sub-aperture.

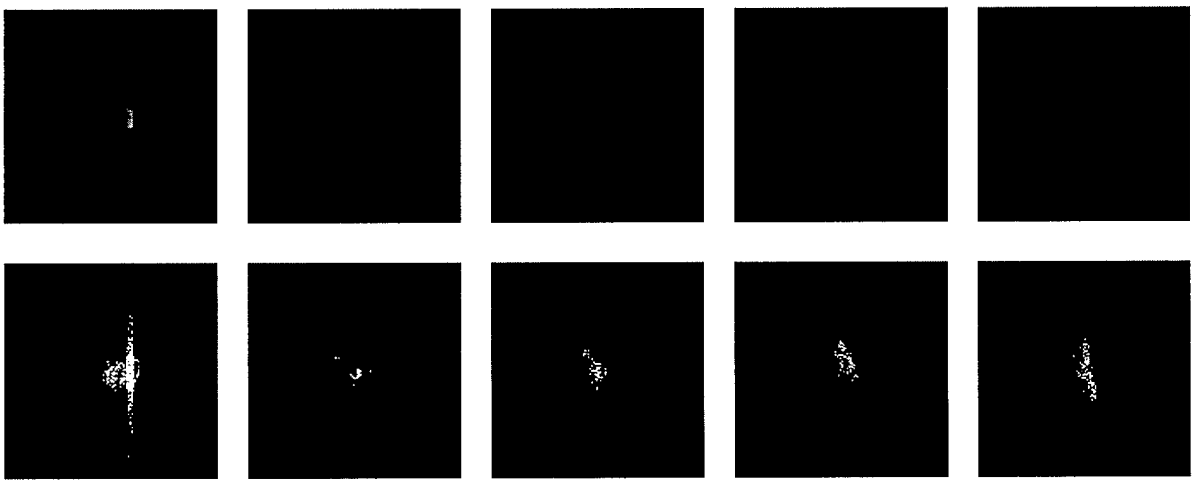


Figure 21: Anisotropy characterization of several instances of a BMP2. Top row: Log-magnitude reflectivity image. Bottom row: Log-magnitude reflectivity image with pixel locations declared to be anisotropic masked out in white.

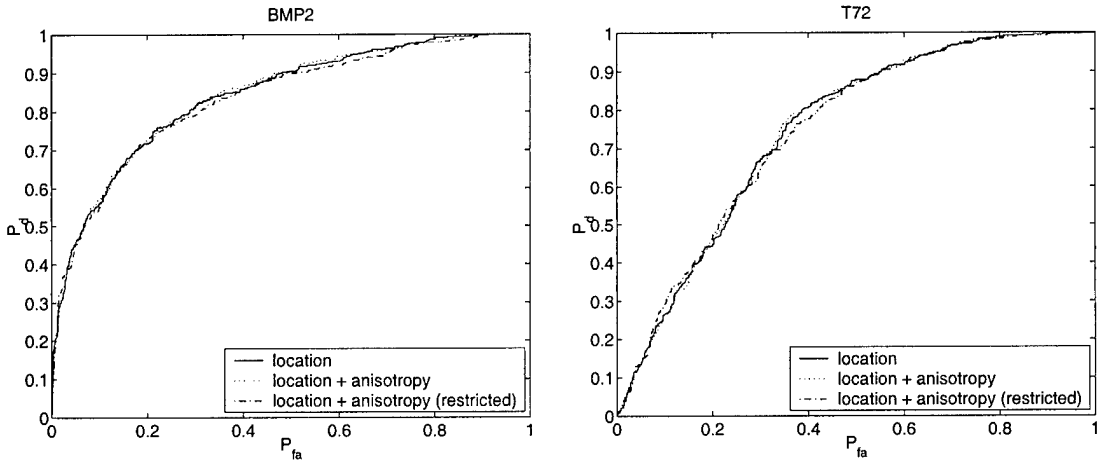


Figure 22: ROC curves for the BMP2 (left) and T72 (right) using features (F1)-(F3).

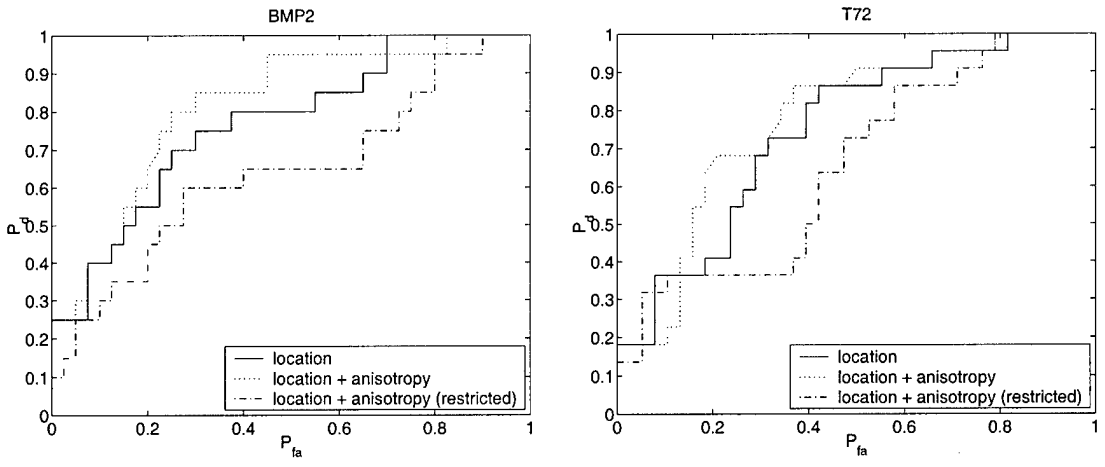


Figure 23: ROC curves for the BMP2 (left) and T72 (right) at near-cardinal angles using features (F1)-(F3).

5 Structure-driven SAR image registration

We present a fully automatic method for the alignment SAR images, which is capable of precise and robust alignment. A multiresolution SAR image matching metric is first used to automatically determine tie-points, which are then used to perform coarse-to-fine resolution image alignment. A formalism is developed for the automatic determination of tie-point regions that contain sufficiently distinctive structure to provide strong constraints on alignment. The coarse-to-fine procedure for the refinement of the alignment estimate both improves computational efficiency and yields robust and consistent image alignment.

5.1 Introduction

The ability to bring multiple synthetic aperture radar (SAR) images into alignment is crucial in many applications. For example, a need for accurate alignment commonly occurs when multiple partially overlapping image are "stitched" together to form a single larger image. Even with geocoding information, precise pixel-level registration is need to accurately align the images along seams. Precise alignment is also needed in applications performing change detection; for example, detecting vehicle or missile movement. Here alignment must be accurate enough to guarantee that each stationary target is properly registered, ensuring that correspondence is correctly determined.

Here we present a fully automatic method for alignment SAR images, which is capable of precise and robust alignment.

5.2 Overview of the Algorithm

The automatic registration algorithm presented here can be broken into several stages, as illustrated by Figure 24. In an initial preprocessing stage, the input images are amplitude equalized to enhance the relative importance of scene elements such as roads or trees, which can often be critical in accurately determining alignment. These images are then aligned by a coarse-to-fine registration procedure, in which processing is first done on low resolution versions of the input image, producing an initial alignment transformation. This alignment is then refined by reestimating the transformation with higher resolution images, beginning from the estimate obtained at the lower resolution.

At the coarsest resolution, a set of tie-point regions are automatically determined from the base image. Alignment transformations are evaluated by comparing the tie-point regions in the base image to the corresponding points in the transformation of the second image. Regions are compared using the flexible histogram texture comparison method [24, ?, ?].

A stochastic hill climbing algorithm is used to find successively better alignment transformations. When the hill climbing procedure converges, refinement of the alignment transformation begins at the next higher resolution. For each successive resolution, the tie-points used in the previous resolution are mapped into the higher resolution and then refined. The hill climbing procedure is then re-initiated from the convergence point achieved at the previous resolution.

After convergence at the finest resolution, in the experiments described here, an alignment transformation was obtained to within a pixel of perfect alignment as measured by manual registration.

5.3 Histogram Equalization

Unlike the problem of vehicle detection, where natural objects are distractors, when performing SAR image registration low-reflection scene elements – such as roads, water, or forest – contain useful information. To

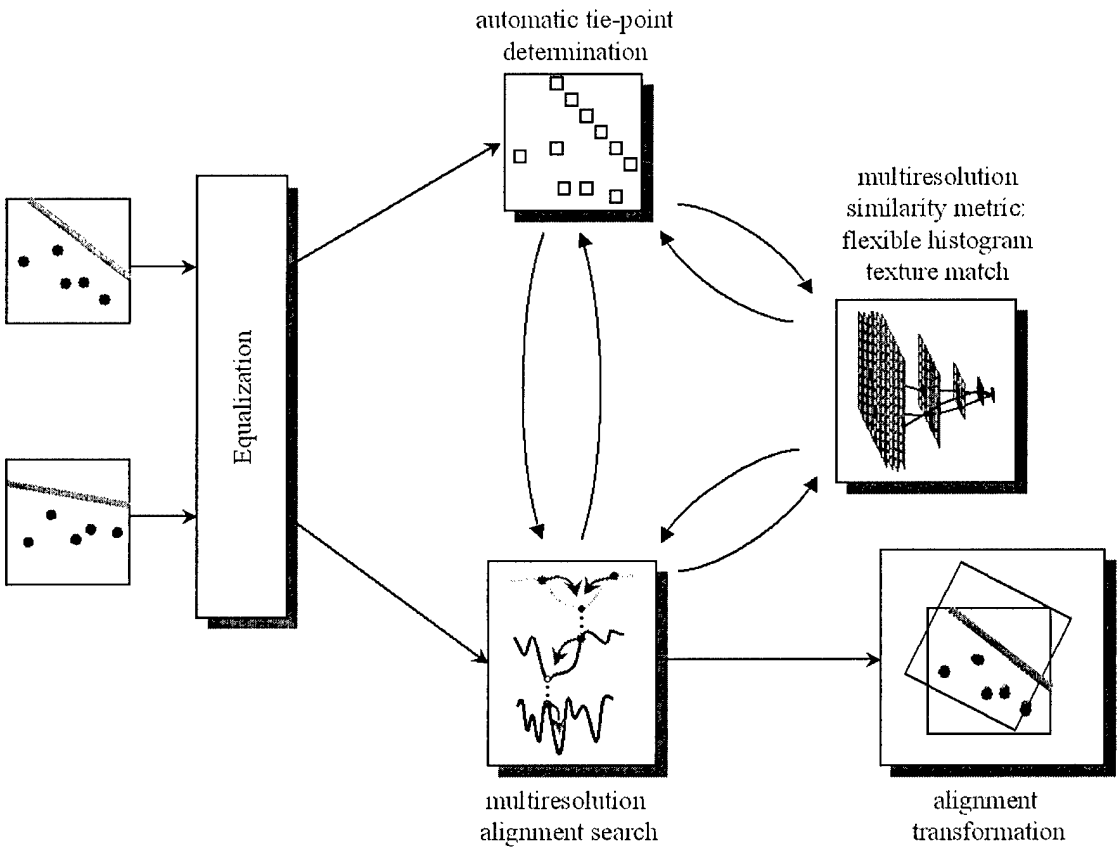


Figure 24: An overview of the image alignment pipeline

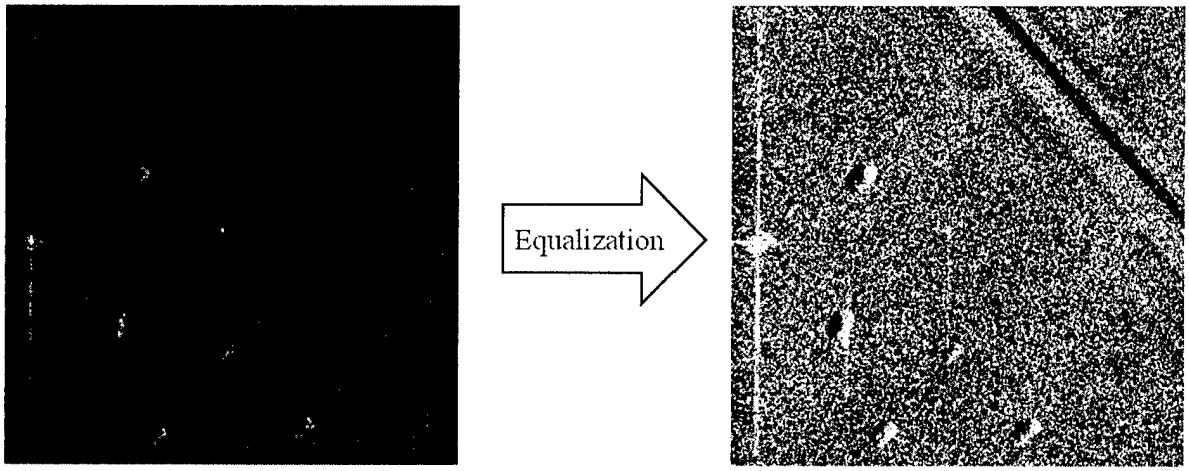


Figure 25: Equalization of the input images enhances the relative importance of low-reflection scene elements

enhance the relative importance of these regions, the input SAR images are preprocessed by dynamically redistributing the pixel brightnesses.

This is done by standard histogram equalization, in which the brightness of each pixel is reassigned by an equalization function $F(\cdot)$ chosen to satisfy the following:

$$\left|\{(x,y)|F(I(x,y)) = V_j\}\right| = K \qquad \forall i,j \tag{38}$$

While simultaneously guaranteeing that if $I(x,y) < I(x',y')$ for any two original pixels, then $F(I(x,y)) < F(I(x',y'))$, i.e. that $F(\cdot)$ is stictly monotonic. Thus a constant number of pixels are histogram equalized to each value V_j, and the brightnesses of pixels are not reordered by the histogram equalization. Though it clearly cannot *increase* the information available in the image, equalization does enhance the effective contrast of elements such as roads, trees, and brick or wood structures, while retaining the highlights and detail on the vehicles.

In Figure 25(a) a SAR scene is shown as a standard log-magnitude image. In this image only the vehicles form high contrast, distinctive regions. After equalization the effective contrast in the road has improved significantly while maintaining high contrast in the vehicle regions. By preprocessing the images in this way the weight of evidence contributed by less reflective objects in the scene can be increased.

5.4 Automatic Tie-point Determination

One method for bringing images into alignment is by choosing several distinctive image regions and carefully registering them. These regions are commonly known as "tie-points" or "landmarks". The quality of a proposed alignment transformation can be determined by evaluating the similarity between the tie-points and the corresponding regions proposed by the transformation. This, counterintuitively, reduces computational cost while simultaneously improving alignment quality.

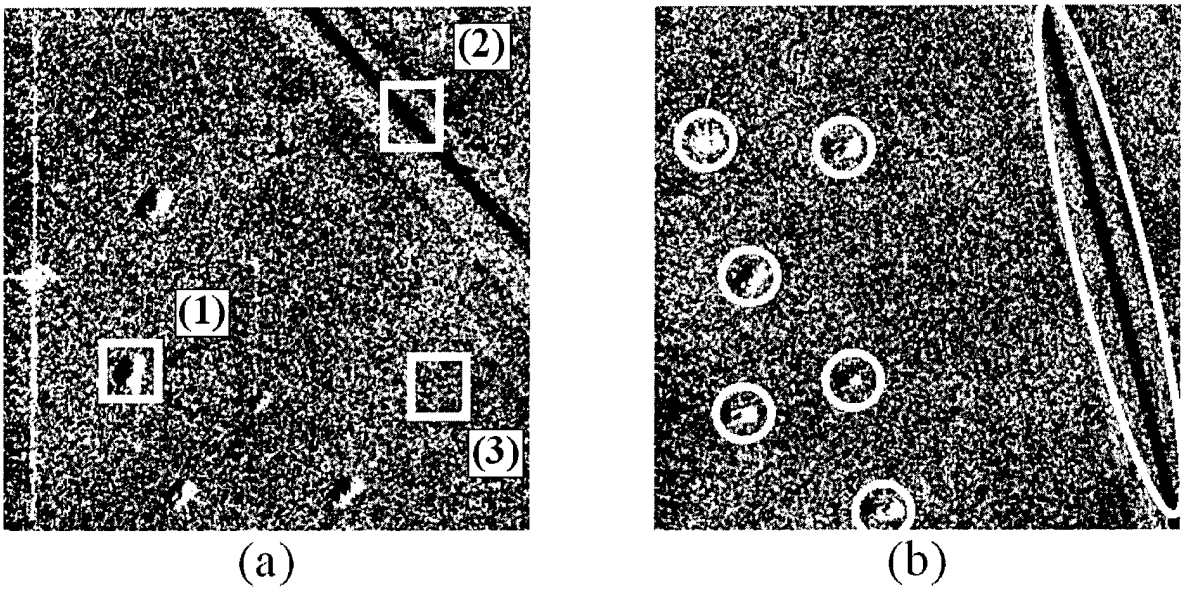


Figure 26: Two SAR images of the same region measured at different aspect angles. If chosen as a tie-point, region (1) constrains alignment to match the circled regions in (b); choosing (2) provides a weaker constraint along region within the oval in (b); while (3) provides no constraint, as it matches with most regions in (b).

Using tie-points has several advantages over the alternative technique of feature matching. Because the nature of SAR imagery, finding features is computationally intensive, and often requires an initial denoising stage[25]. Due to the image speckle and geometry of the SAR imaging process, the features found tend to be unstable and highly variable [26, 25].

The notion of using tie-points to improve measurement in this way is not new, and has been used before[27, 28]. Each of these, however, requires that the tie-points be selected by human operators. We present a method for automatically selecting good tie-points. This technique is highly generalizable and can be used given any particular matching metric for measuring the similarity between two image regions. In the results presented here, we use one such matching metric which has been shown to work well with SAR imagery, the flexible histogram texture matching metric described in a companion paper in these proceedings [29], and elsewhere[24, ?, ?].

In Figure 26(a) three distinct types of image elements have been identified. Localized scene elements – such as the vehicle highlighted in the square (1) – provide the strongest constraints on the possible alignment between two images. Because each localized element is only similar to a few regions in the second image, specifically the region around the true corresponding region and the regions around other similar localized elements, the similarity measured between two images at these localized scene elements is a good indication of the quality of an alignment transformation. Even if one were to use a relatively weak matching metric, such as a CFAR statistic or correlation, the distinctness of such localized scene elements can be used to greatly reduce the number of possible alignment transformations under consideration. Each such local element effectively eliminates all but a small set of possible transformations from consideration. For Figure 26 it rules out any transformation which does not map (1) onto one of the circled regions in image (b). In ideal situations enough such localized elements exist, and are shared by the two images, so that using these regions as tie-points will be sufficient constraints to align the two images closely. Other examples of localized scene elements which are commonly seen are: buildings, isolated trees, or bridges. However, such localized elements tend to be rare, and a general and robust SAR registration algorithm cannot reasonably assume that for all input images, sufficient localized scene elements will exist. Additionally, localized scene elements tend to correspond to objects which could potentially move between the times that two SAR images were acquired. This is particularly evident in the case of vehicles, but is also true to a lesser degree to other local scene elements, such as mobile equipment, supplies or even small buildings.

Regions selected from extended scene elements – such as the section of road highlighted in the square marked (2) in Figure 26a – provide weaker, though still valuable, constraints on the possible alignment transformation. Other examples of such extended scene elements include transitions between different types of terrain, such as tree or water lines, fences or power lines, or large building complexes. Because such extended elements are self-similar, a region in one of these elements will tend to be similar to other regions along the same scene element. Thus, a region chosen from along the road in Figure 26 will match well with any of the regions in the oval in (b). This is true even when using a strong matching metric – such as the flexible histogram texture match[24, ?, ?, 29], or evaluation by a human observer. Such tie-points selected from within extended elements constrain the final solution to a curve or small local region.

Common scene elements – such as the grass in square (3) in Figure 26a – provide virtually no constraint on the possible alignment transformation. Other examples of scene elements which tend to occur too frequently to be useful include the interior regions of farmland, forest, or water. Because the interior of these types of regions are self-similar, any region within them will be similar to any other such region. Thus, even if a such a region matches well under a proposed alignment transformation, little information is gained – because there are a large number of other transformations under which it will match equally well.

By selecting as many localized scene elements and a sufficient number of extended scene elements to use as tie-points, accurate of alignment can be obtained from comparing just these regions. Clearly there

is a significant computational advantage of comparing only those pixels within tie-points regions versus comparing all pixels which overlap under the proposed transformation. Intuitively this corresponds with examining only those regions which are likely to be different if the proposed alignment is inexact.

More importantly, however, limiting the alignment quality estimate to tie-point regions eliminates the additional noise which would otherwise be introduced by the small variations in similarity between the uninformative common scene elements. Even though it is expected that the variation between any two common element regions (i.e. two patches of grass) would be relatively small, the combined effect of all of such regions could introduce sufficient noise to "drown out" the real information from the more localized scene elements.

We present here a formalism for determining tie-point regions in a completely automated fashion. The utility of using a given region as a tie-point is inversely proportional to the frequency with which similar types of regions occur in the input images. Because we can assume that if the images contain views of roughly the same scene it is sufficient to locate distinct points in a single image by comparing them to other regions in the same image.

5.4.1 Tie-Points are High Entropy Regions

Given a statistical model of a SAR imagery we can in principle measure the entropy of a candidate tie-point, t. The entropy of such a patch, $H_{Model}(t)$, is directly proportional to the frequency at which similar patches are expected to appear. The search for distinctive patches is then a search for the image regions with high entropy:

$$\max_t \left[H_{Model}(t)\right] \tag{39}$$

This notion is quite general; there are potentially many possible estimators of tie-point entropy. For example, edges are useful features in imagery precisely because they have high entropy[25]. We have chosen to use the multi-scale statistical models of SAR imagery defined by De Bonet and Viola[24, ?]. These models can be trained directly from SAR imagery. In this case the above entropy can be written as follows:

$$\max_t \left\{E_x\left[H\left(t|x\right)\right]\right\} \tag{40}$$

where x is a patch of SAR imagery drawn at random directly from the image to be registered, and $E_x(\cdot)$ is an expectation taken over all possible image patches.

By manipulating equation (40) we can obtain an expression in terms of mutual information:

$$\begin{aligned}
\arg\max_t \left\{E\left[H\left(t|x\right)\right]\right\} &= \arg\min_t \left\{-E\left[H\left(t|x\right)\right]\right\} && (41)\\
&\overset{(a)}{=} \arg\min_t \left\{E\left[H\left(x\right)\right] - E\left[H\left(t|x\right)\right]\right\} && (42)\\
&\overset{(b)}{=} \arg\min_t \left\{E\left[H\left(x\right) - H\left(t|x\right)\right]\right\} && (43)\\
&\overset{(c)}{=} \arg\min_t \left\{E\left[I\left(x;t\right)\right]\right\} && (44)
\end{aligned}$$

where $I(x;t)$ is the mutual information between x and t. Equality (a) holds because $E[H(x)]$ is independent of t and therefore does not change the result of $argmin_t$; (b) follows from the fact that expectation is linear; and (c) follows from the definition of mutual information. Thus the best tie-point selections are those which have the lowest expected mutual information with other regions in the image.

The output of any image similarity metric, $S(\cdot)$ can be viewed as an approximation of the mutual information between two regions. Thus optimal tie-points can be found from:

$$\arg\min_t \left\{E\left[S\left(x,t\right)\right]\right\} \approx \arg\min_t \left\{E\left[I\left(x;t\right)\right]\right\} \tag{45}$$

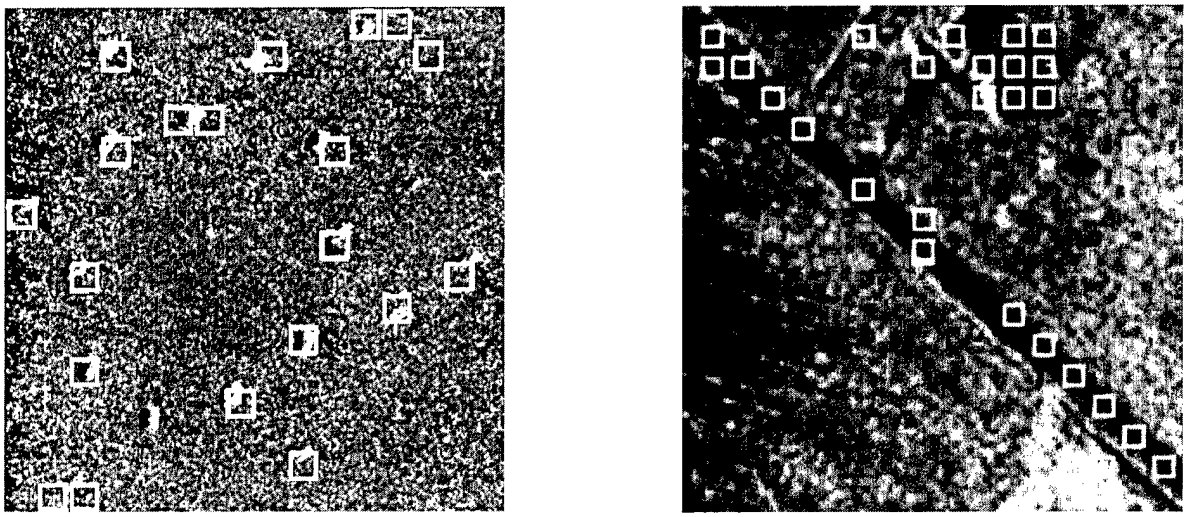


Figure 27: Two typical examples of the tie-points automatically found by this algorithm.

The quality of this approximation is directly related to the quality of the similarity metric. However, if we use the same similarity metric $S(\cdot)$ to determine tie-points as we use to evaluate alignment transformations, then it can be shown equation (45) is optimal with respect to the later measurement.

To reduce computational cost the expectation of $S(\cdot)$ is stochastically approximated. The number of comparisons used in this approximation can be easily tuned to meet specific computational criteria.

In a similar fashion, a set of good set of tie-points $T = t_1, t_2, \ldots, t_n$ can be found by finding the n most distinct regions in the image as measured by equation (45). In practice, however, it is imperative that the tie-points used not only be distinctive, but that they also cover the image, so that at least some tie-points will lie in the overlapping region under any alignment transformation. To guarantee this, any of a number of schemes can be employed. Here we use a scheme which subdives the image into multiple partitions, considers candidate regions from within each partition, and requires that at least one candidate from each partition be chosen as a tie-point.

In later sections we discuss the coarse-to-fine refinement of the alignment estimate, in which sets of tie-points are needed at each resolution. Each set of tie-points need not be recomputed from scratch if we make the following assumption: distinctive regions at some resolution are likely to be at or around the distinctive regions at lower regions. Clearly this assumption is not true in the general case, as it prohibits small objects, or those which are distinct because of high frequency detail, from ever being chosen as tie-points. Empirically, however, they are valid for mm-wave SAR imagery because high frequency detail tends to be unstable due to speckle. Given this assumption, it is sufficient to locate higher resolution tie-points only in those areas which are within a small region surrounding the tie-points used at the previous resolution.

Two typical examples of the tie-points found by this algorithm are shown in Figure 27. In (a), which contains several vehicles, tie-points are selected in regions around each vehicle; thus, evaluating an alignment transformation, to a first approximation, corresponds to measuring how many vehicles line up with other vehicles. The quality of each match only becomes important when fine tuning the alignment transformation.

In (b), fewer localized elements exist and tie-point selections include some of the extended elements in the scene.

5.5 Stochastic Alignment Optimization

Given a set of tie-points in the base image, we can evaluate the quality of a proposed alignment transformation. One could imagine simply computing the quality of every possible alignment and returning the maximum quality alignment. If we were to consider only the class of translations, such an approach might be feasible though still computationally taxing, requiring $O(X \times Y \times N)$ region similarity comparisons for an X by Y image with N tie-points. However, when we even consider adding another dimension (e.g. rotation) this direct approach clearly becomes infeasible.

Instead we employ a directed search through the space of transformations, in an attempt to determine the best transformation. We do this by using simulated annealing, a standard non-linear optimization search technique, at each resolution. We initiate the search with the convergence point reached at the previous (lower) resolution.

In employing a multiresolution gradient based technique, the following two assumptions are made:

- though there may be noise in the objective function – the alignment transformation quality measurement – its maximum will be surrounded by points whose values are also good.
- the objective function will be smoother at lower resolutions, though the maximum at lower resolutions may deviate from the true fine-resolution maximum.

The rationale behind this methodology is shown schematically in the the curves shown on the top of Figure 28

Using the low resolution estimate to initialize the search at the next higher resolution obtains a significant computational benefit in three ways:

- alignment quality evaluations are less computationally expensive at lower resolutions
- as the low resolution objective function surface is smoother, we can use a lower "temperature" in the search
- since low resolution estimates are close to the higher resolution optimum, fewer local optimal are likely to exist between the global maximum and the low resolution estimate than if we were to begin at some random initialization point; this also allows for the use of a lower search temperature.

For SAR imagery these assumptions are in most cases valid. This can be visualized by examining the objective functions shown in Figure 28. In these graphs, alignment quality is shown as a function of translation. At the lowest resolution (a) the surface is smoothest, however, at this resolution each pixel represents an uncertainty of 4 pixels at the finest resolution, so the estimate of $(0, 0)$ is effectively 8 to 12 pixels away from the true alignment. At the next finer resolution (b) the estimate of $(2, 0)$, though farther from the true alignment (which is at $(0, 0)$), it is nevertheless slightly more accurate due to the increased resolution yielding an offset of 6 pixels. At the finest resolution, the global optimum is exact; however, the surface begins to show local optima which could trap a hill climbing procedure. By initiating the search at the estimate obtained from (b) such local optima can be avoided altogether.

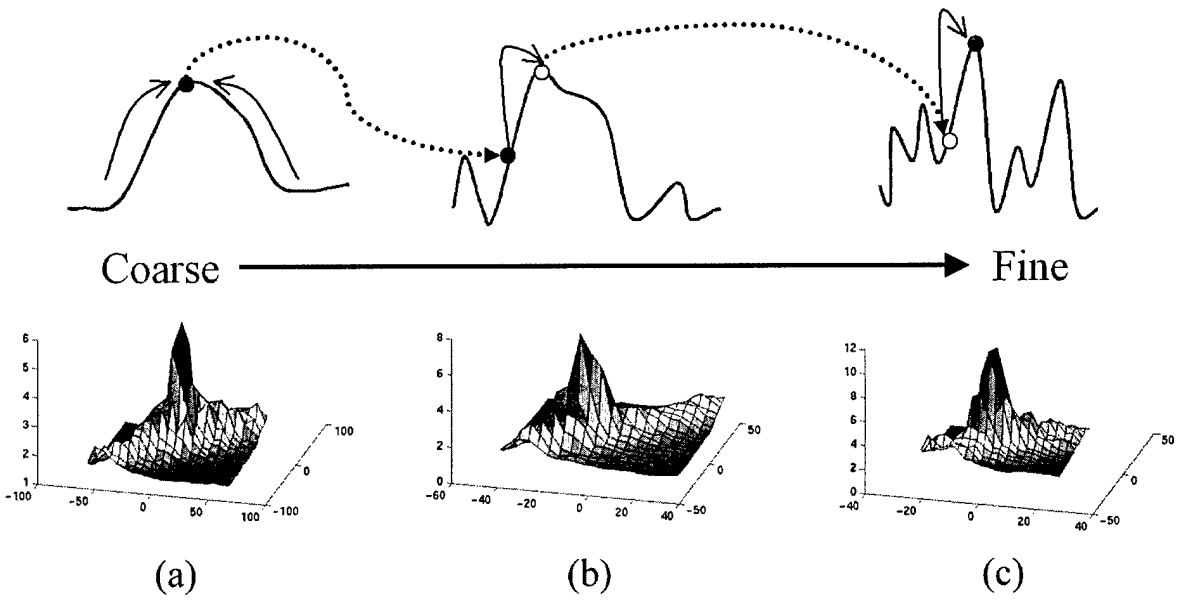


Figure 28: Coarse-to-fine registration is used to improve computational efficiency and to avoid local extrema. A typical example of surface of the alignment quality as a function of translation, at coarse (a), medium (b) and fine (c) scales.

5.6 Overview of Simulated Annealing

Though we do not have space to discuss it in depth here, we present a brief description of the simulated annealing phase of the alignment procedure. A more complete discussion of simulated annealing can be found in [30, 31, 32, 33].

Simulated annealing is a stochastic hill climbing algorithm in which the objective function is explored around a single estimate of the function's optimum. During this exploration, the estimate of the optimum is updated and search begins from this new estimate. In this work we explore the objective function by choosing potential alignment transformations from a probability distribution centered at the current estimate. In this work, new transformations are chosen from a Gaussian distribution about each parameter in the transformation. I.e.:

$$\begin{aligned} x_1' &= x_1 + \eta_1 \qquad \eta_1 \sim \mathcal{N}(0, \sigma_1) \\ x_2' &= x_2 + \eta_2 \qquad \eta_2 \sim \mathcal{N}(0, \sigma_2) \\ &\vdots \\ x_n' &= x_n + \eta_n \qquad \eta_n \sim \mathcal{N}(0, \sigma_n) \end{aligned} \tag{46}$$

where $X = (x_1, x_2, \ldots, x_n)$ are the parameters for the current estimate of the optimal alignment transformation, and $X' = (x_1', x_2', \ldots, x_n')$ the parameters for the new point under consideration.

If the objective score at X' is better than that at the current estimate X, then estimate of the optimum is updated to X'. Additionally, if the objective score at X' is worse than at X then the estimate is updated with probability:

$$\Pr(X \rightarrow X' | S(X) > S(X')) = e^{-\frac{S(X) - S(X')}{\tau}} \tag{47}$$

where $S(X)$ is the objective score at X, and τ is the temperature of the search.

In the case of image alignment $S(X)$ is the combined similarity measurement between all of the tie-points in the base image and their corresponding regions in the second image under the transformation defined by X.

For SAR image alignment, we consider full six dimensional affine transformations. However, in practice most of this transformation consists of a rigid transformation (translation, and rotation) between the two images. We therefore decompose the affine transformation in the following way:

$$\begin{aligned} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} dx \\ dy \end{bmatrix} = & \begin{bmatrix} \cos(r) & \sin(r) \\ -\sin(r) & \cos(r) \end{bmatrix} \begin{bmatrix} m & 0 \\ 0 & n \end{bmatrix} \times \\ & \begin{bmatrix} \cos(s) & \sin(s) \\ -\sin(s) & \cos(s) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} dx \\ dy \end{bmatrix} \end{aligned} \tag{48}$$

in which there is a translational component (dx,dy) a rotation (r) and stretching along a major and minor axis (m, n, along direction defined by s). In decomposing the affine transformation in this way, the simulated annealing search can be directed to explore each of these dimensions with different size steps, as defined by the corresponding $\{\sigma_i\}$ in equation (47)

The temperature parameter τ in equation (47) is annealed according to a fixed schedule for each resolution. The values for the parameters σ's and τ were chosen manually to optimize search time on several pairs of SAR images.

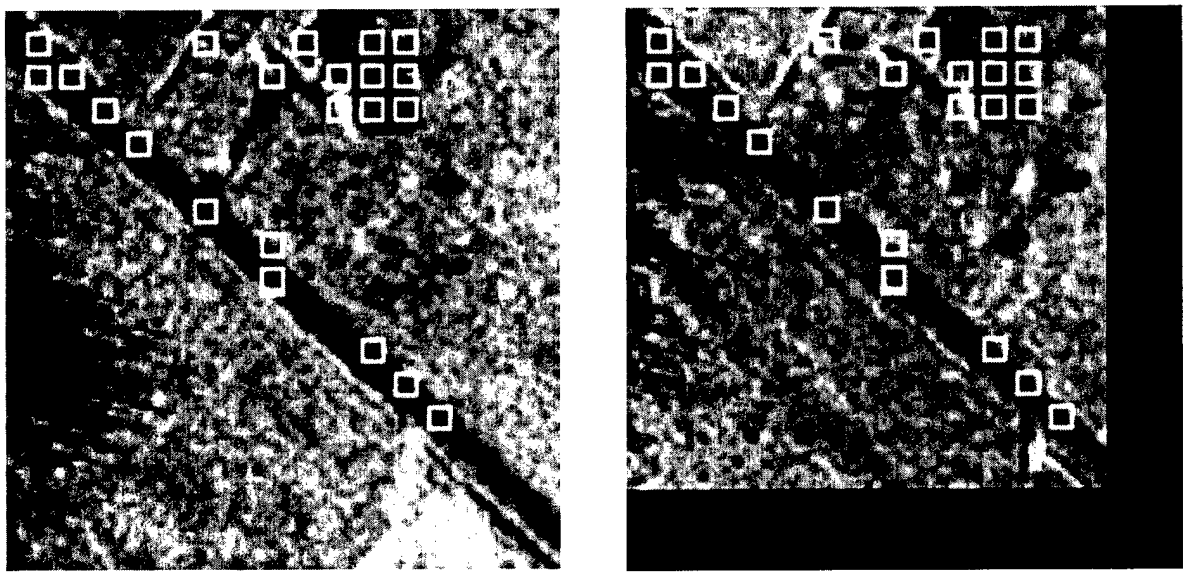


Figure 29: A typical registration

The analysis up to this point has only used the similarity metric $S(\cdot)$ abstractly. Many algorithms have been shown to be good similarity metrics for SAR imagery[27, 28, 6, 5, 34, 35] and would be reasonable candidates for $S(\cdot)$ In this work we use the flexible histogram texture matching metric, which is described in a companion paper in these proceedings[29], and elsewhere [24, ?, ?].

5.7 Results

Several examples of aligned pairs of SAR images are shown in Figures 29 and 30. Automatically selected tie-point regions, highlighted by the white squares, cluster around the distinctive regions in the images.

Though alignment performance is difficult to fully quantify, we have informally been able to make some preliminary measurements of the systems performance. The parameters of the alignment procedure were manually tuned using several image pairs. Alignment was then measured for a collection of 36 different image pairs, with randomly selected initial alignment. Alignment was generally within a few pixels of the optimal affine transformation, as defined by manual alignment. Furthermore, alignment quality was consistent regardless of initial alignment of image pairs.

5.8 Discussion

The two key developments we have presented, which make robust and consistent alignment possible, are the process for the automatic determination of tie-points and the use of coarse-to-fine alignment. Useful tie-points correspond to regions which have a low expected mutual information with other regions in the SAR images. Regions selected in this way correspond to the distinct elements in SAR imagery which are inherently useful in achieving accurate alignment. This formulation for the automatic determination of tie-point regions is extremely general and can can be used by many tie-point based techniques. Using a

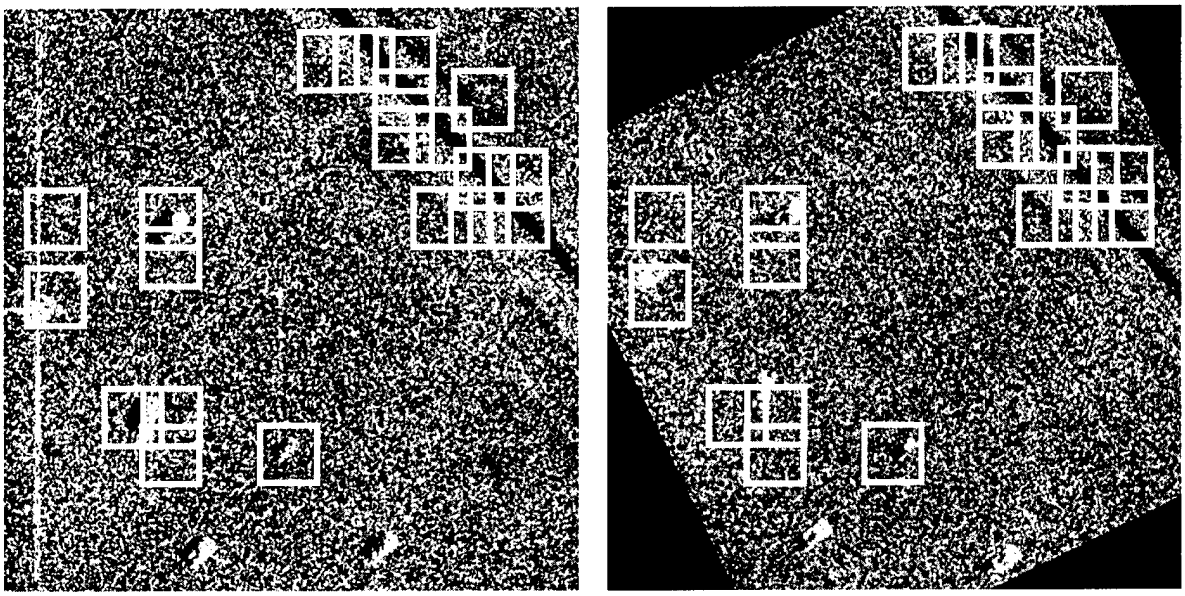


Figure 30: A typical registration

coarse-to-fine progression to refine the estimate of the image alignment improves computational efficiency and results in more robust alignment. At each resolution, simulated annealing is used to determine the optimal alignment, and is greatly aided by the initialization obtained from the alignment at the previous (coarser) resolution.

The alignment of all of the vehicles in Figure 30 is not perfect even though the optimal *affine* alignment has been found indicating that the class of affine transformations are not sufficient to completely align two images. In future research we intend to examine larger classes of transformations to attempt to achieve even tighter registration between images.

Although the flexible histogram texture matching metric performs well on SAR imagery, as seen by the smoothness of the objective functions in Figure 28, in this direct framework it can only make SAR to SAR comparisons. One of the long term objects of image registration is the alignment of data from several sensors into a single reference frame. The mutual information technique used by Viola and Chao[27] presents a method for measuring the mutual information between pixel brightnesses. We hope to extend this work by examining the mutual information between the flexible histogram texture measures obtained from multiple sensors.

6 Information Theoretic Feature Extraction for ATR

Utilizing principles of information theory, non-parametric statistics and machine learning we describe a task-driven feature extraction approach. Specifically, the features preserve information related to the specific estimation problem. Mutual information, motivated by Fano's inequality, is the criterion used for feature extraction. The novelty of our approach is that we optimize mutual information in the feature space (thereby avoiding the curse of dimensionality) and we do so without explicit estimation or modeling of the underlying density. We present experimental results for pose estimation of high-resolution SAR imagery.

6.1 Introduction

Modern ATR system designers are faced with the difficult problem of processing data of increasingly higher dimension. Classical decision approaches are not well suited to high-dimensional data and so dimensionality reduction, or feature extraction, is often performed. This is notionally represented in figure 31 by the function $g([\cdot], \alpha)$ which maps high-dimensional data to a much lower dimension and which has some finite set of parameters, α, which must be set. Popular methods for data driven feature extraction, however, are optimal only in the signal reconstruction sense (e.g. eigenvector methods such as PCA) or make strong assumptions about the underlying data densities (e.g. independent components analysis). It is these shortcomings that we seek to address here.

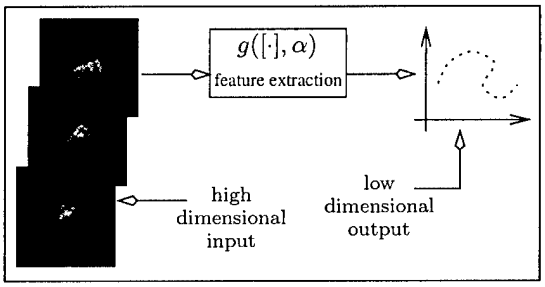


Figure 31: Notional diagram of the feature extraction problem. Our input is high-dimensional data (e.g. SAR imagery) which we wish to map to some low-dimensional representation. We learn the mapping parameters α from data examples.

We discuss an information theoretic approach to feature extraction. Our approach is novel in that it combines a non-parametric density estimator with an information theoretic criterion, namely mutual information. In so doing we formulate a learning approach for features which preserves information related to the parameters of interest.

6.2 Information Theoretic Approach

Our intuition is as follows. High dimensional data, such as images, convey many "bits" of information only some of which we may interested in decoding (e.g. background clutter type, object class, object pose, etc.). Given an estimation task, other pieces of information may not be of interest and can be thought of as

nuisance parameters. It seems only logical that in choosing a feature extraction method we should seek to preserve information about the parameters of interest. This prompts the following questions:

1. Can we find "informative" directions within the high-dimensional input space?
2. Can we learn these directions from data?

The questions are related. What we seek in the first question should guide our choice of a learning criterion in the second. As a first step we can view the feature extraction process as a Markov chain, as in figure 32, beginning with the parameter of interest, θ, proceeding to the observed data, x, and ending in the computed feature, y.

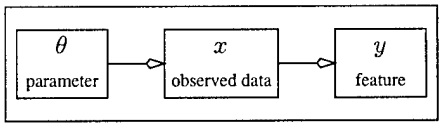


Figure 32: Feature extraction as a Markov process, $p(\theta, x, y) = p(\theta)p(x|\theta)p(y|x)$.

A learning criterion which addresses both questions is mutual information (MI). We learn the parameters, α, of the mapping in figure 31 so as to maximize the MI between θ and y. MI between two random variables is defined in three equivalent ways as [36]

$$\begin{aligned} I(\theta, y) &= H(\theta) + H(y) - H(\theta, y) \\ &= H(\theta) - H(\theta|y) \\ &= H(y) - H(y|\theta) \end{aligned} \tag{49}$$

where $H(z)$ is either the differentiable or discrete entropy of the random variable z (depending on whether z is continuous or discrete). Entropy is defined as

$$H(z) = \begin{cases} -\int_{\Omega_z} \log\left(p(z)\right) p(z) \mathrm{d}z & z \text{ continuous} \\ -\sum \log\left(p(z_i)\right) p(z_i) & z \text{ discrete} \end{cases} \tag{50}$$

Consider the case in which θ is discretely distributed (although x and y need not be). We can lower bound the probability of error in estimating θ as a function of MI by Fano's inequality [36].

$$Pr\{\hat{\theta}(y) \neq \theta\} \geq \frac{H(\theta) - I(\theta, y) - 1}{\log_2(N)} \tag{51}$$

Where $\hat{\theta}$ is the estimator of θ as a function of the feature vector y and N is the number of values which θ takes. Observe that the only degree of freedom in the inequality is the choice of y. Effectively, Fano's inequality relates the inference power (as measured by the lowest potential probability of error) to the MI in the features. Another issue to note is the data processing inequality [36] which states that in a Markov chain:

$$I(\theta, y) \leq I(\theta, x),$$

so that by extracting features we can only destroy information about θ (or at best maintain it). Estimation of continuous parameters can be similarly motivated by the idea of a sufficient statistic. A property of a sufficient statistic is that if $I(\theta, y) = I(\theta, x)$ then y is a sufficient statistic for θ.

These desirable properties of MI are well recognized within the statistical pattern recognition community. What is novel here is that we will present a machine learning approach which incorporates MI as a criterion for setting the mapping parameters, α. We face a key challenge, however, in that entropy and by extension MI is an integral function of joint probability densities which we do not know.

6.3 Approximating Entropy

Examining equation 49 we see that MI is a combination of entropy terms. Therefore if we can approximate entropy (or its gradient) then by extension we can perhaps approximate MI. In previous work we have presented an approach for approximating entropy [37, 38]. For the sake of brevity we summarize the results of that work, the details can be found therein. In estimating entropy and its gradient we make use of the following:

1. the maximum entropy probability density over a finite region is the uniform density, and
2. expanding $p\log(p)$ as a second order Taylor series about the uniform density equates minimizing the integrated squared error between the estimated density and the uniform density to maximizing entropy.

So an approximate criterion for maximizing (or minimizing) entropy is

$$J = \int_{\Omega_u} (\hat{p}(u) - p_u(u))^2 \, \mathrm{d}u \tag{52}$$

where $p_u(u)$ is the uniform density over the finite extent region Ω_u and $\hat{p}(u)$ is a density estimator. For the estimate we use the Parzen density estimate [39] defined as

$$\hat{p}\,(u) = \frac{1}{N} \sum_i \kappa\,(u - y_i, h) \tag{53}$$

where $\kappa([\cdot], h)$ is a Gaussian function with variance h^2 and $\{y_i\}$ are a set of N data samples. In application y_i are the outputs of our mapping function, i.e. $y_i = g(x_i, \alpha)$ where x_i are high-dimensional input data. Additionally, in order to satisfy the finite range extent we choose a multi-layer perceptron (MLP) with squashing nonlinearity as the mapping function. The parameters, α are therefor the weights of the network. As shown in [37, 38] the update term for the networks weights becomes

$$\frac{\partial J}{\partial \alpha} = \frac{1}{N}\sum_i \epsilon_i \frac{\partial}{\partial \alpha} g(x_i, \alpha) \tag{54}$$

$$\epsilon_i = f_r(y_i) - \frac{1}{N}\sum_{j \neq i} \kappa_a \left(y_i - y_j, h\right) \tag{55}$$

$$f_r(y_i)_k \approx \frac{1}{d^M}\prod_{j \neq k}\left(\kappa_1\left(y_{ik} + \frac{d}{2}, h\right) - \kappa_1\left(y_{ik} - \frac{d}{2}, h\right)\right) \tag{56}$$

$$\kappa_a\left(u, h\right) = \kappa(u, h) * \kappa'(u, h) \tag{57}$$

$$= -\frac{\exp\left(-\frac{y^T y}{4h^2}\right)}{\left(2^{M+1}\pi^{M/2}h^{M+2}\right)} y \tag{58}$$

where M is the dimensionality of the feature vector y. Both $f_r(y_i)$ and $\kappa_a(u, h)$ are M-dimensional vector-valued functions and d is the support of the output of the mapping (i.e. a hyper-cube with sides of length d centered at the origin). The notation $f_r(y_i)_k$ refers to the kth element of $f_r(y_i)$ [38].

Despite their appearance, the set of equations has a natural interpretation in the context of maximizing entropy. The gradient update of equation 54 consists of two terms: ϵ_i is the error *direction* term, while $\frac{\partial g(x_i, \alpha)}{\partial \alpha}$ is the network sensitivity term encountered in back propagation training of an MLP. Consequently, the key difference between using MI as the learning criterion versus a supervised learning approach is solely in the computation of the error terms. The error term in equation 55 is comprised of two additional terms. The first, $f_r(y_i)$ directs samples away from the boundary of the output map (consequently preventing saturation). The second, $-k_a(y - y_j, h)$ summed over all samples pairwise repels samples away from each other. The net effect when maximizing entropy is that samples are uniformly distributed in the output space. If, on the other hand, one were minimizing entropy, these terms change sign and have the opposite effect. A combination of maximizing unconditional entropy and minimizing conditional entropy are used for MI.

We should also note that the above algorithm computes an exact gradient to the integral criterion of equation 52 while requiring function evaluations only at the sample points. This is somewhat surprising as the criterion is an integral function over the density.

6.4 Estimation of Pose from Learned Features

We now present the results of a simple experiment. Given SAR images of vehicles, we wish to derive two features (i.e. y is two-dimensional) which convey information about the pose of the vehicle. Examples of these images are shown in figure 31. The images, of dimension 128×128, are taken from the MSTAR public release data base. Specifically we use images of a single vehicle type, T-72 tank. We train on 116 images from one vehicle (s/n s7) sampling about every three degrees of aspect and test on another (s/n 812). We use a single layer perceptron with two outputs. Each of the training images is labeled with an approximate pose angle (relative to the radar platform). We use this information jointly with the images to train the network. Recall there is no desired signal as in the supervised training case, rather we desire only that the feature vector conveys maximal information about the pose of the vehicle as computed from the image. Minor pre-processing is done to each image, namely, normalization to unit energy followed by subtraction of

the mean image over all training samples (also applied to new testing images). As a comparison we use the two largest principal components as a competing set of features.

In applying the technique we maximize entropy in the feature space (i.e. in general samples repel each other) except for samples with poses that are close in which case we minimize entropy (i.e those samples attract each other). Figure 33 shows the resulting feature space using MI (top) as compared to the two largest principle components. Although not unique, the plot at the top is a natural solution to the competing criterion. Furthermore, although not perfect, the mapping generalizes to the testing set quite well. It is not hard to imagine that the features would be useful for estimating the pose of the image, although we will give better evidence of that later. By comparison, although PCA features might useful for something, there is little visual evidence that pose information has been preserved.

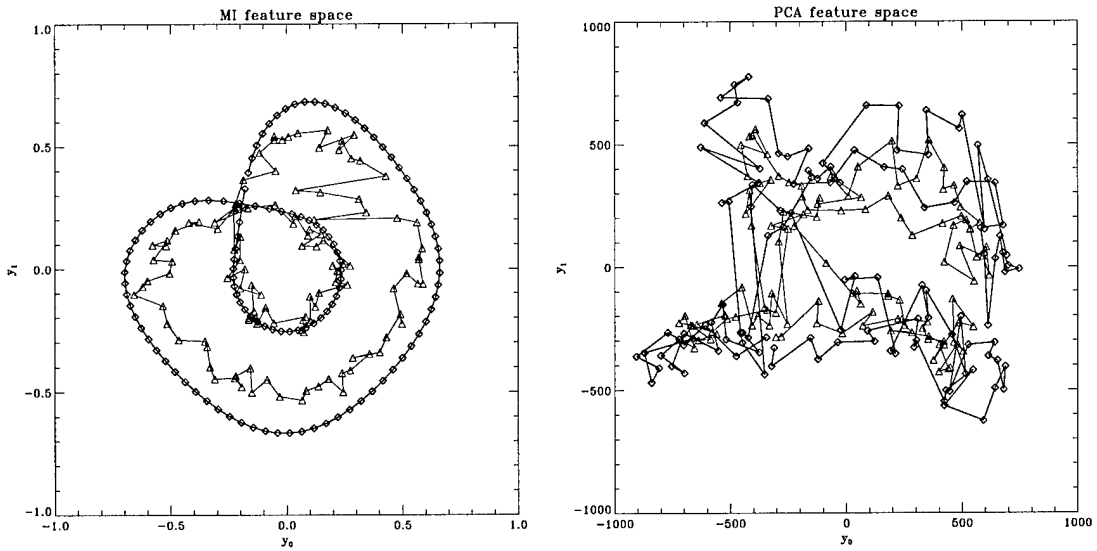


Figure 33: Comparison of feature spaces. MI feature space (top) and PCA feature space (bottom). Training samples are denoted by triangle symbols while testing examples are denoted by diamond symbols. Adjacent pose angles are connected.

As further evidence of the information preserving properties of the approach examine figure 34. Since the input to the single layer network is linear we can examine the subspace of the input projection. That is, we can remove the null space components from the original input image and see which image features remain. We can, of course, do this with the PCA features as well. In the figure we show two examples of a training image (on the left) at two different poses. Alongside these are the back-projected images for both MI (middle) and PCA (right), respectively. A simple psycho-physical test, in which one covers up the original image and tries to determine the vehicle pose by inspecting the back-projected images demonstrates that the MI features yield a slightly better estimate, particularly for the set of images at the top of the figure. Furthermore, leaving the original image uncovered and covering up either the MI image or the PCA image shows that the MI image has slightly better agreement than PCA image. Admittedly this is hardly conclusive and so we present an additional result.

Having computed the features, we are still left with question: how do we use them? Presumably they

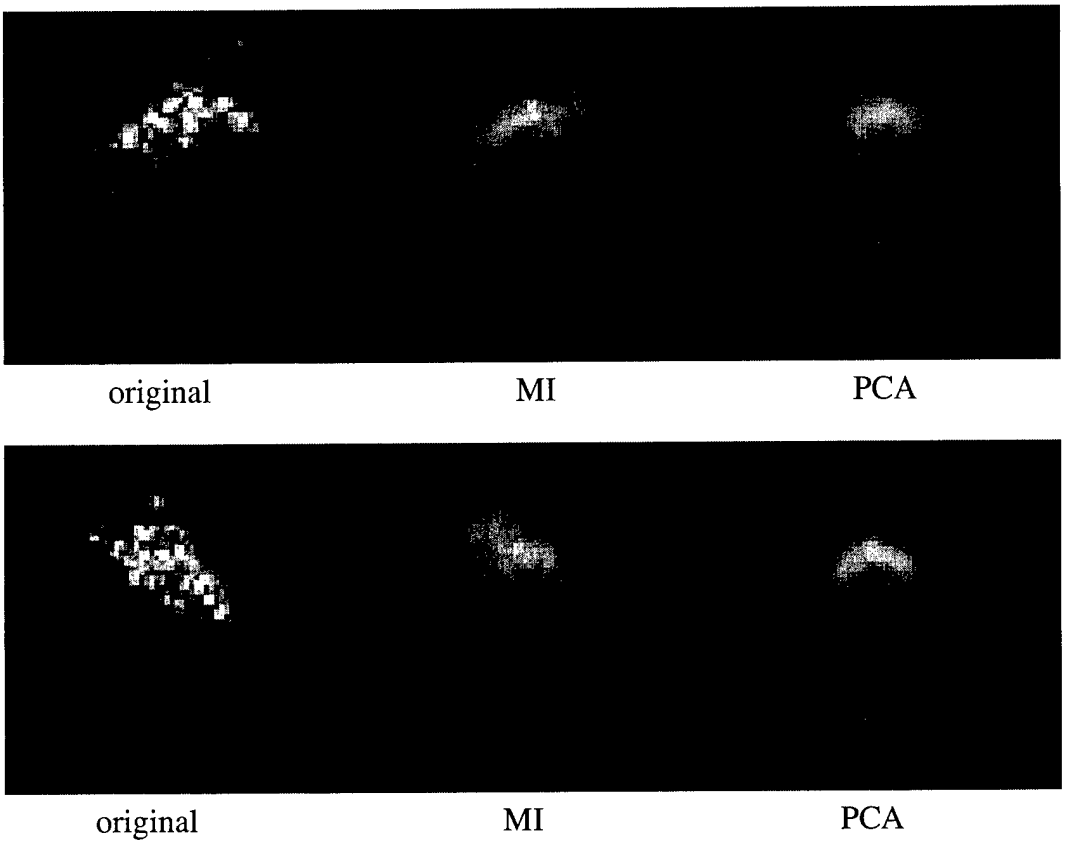


Figure 34: Comparison of back-projected images

convey the information we are interested in, but we must still decode it. One approach is to use the same Parzen density estimator upon which the technique is based to estimate the parameter encoded by the features. Note that this would have been unthinkable in the original input space, but is easily done in the new low dimensional feature space.

Specifically, given a new image and its induced features y_i, we compute

$$\hat{\theta} = \arg\max_{\theta} \hat{p}(\theta \mid y = y_i) = \frac{\hat{p}(\theta, y_i)}{\hat{p}(y_i)} \tag{59}$$

where $\hat{\theta}$ is the MAP estimate of the vehicle pose.

Four examples are illustrated in figures 35-38. We estimate the conditional density above using both sets of features computed from the training samples and compare the results in the figures. We choose four examples, two training and two test images. A training and test example are taken from the ambiguous region of the MI feature space (i.e. where the curve crosses over itself). Presumably this is the region where the MI features will be least useful for estimating pose due to the competing hypotheses. The other two

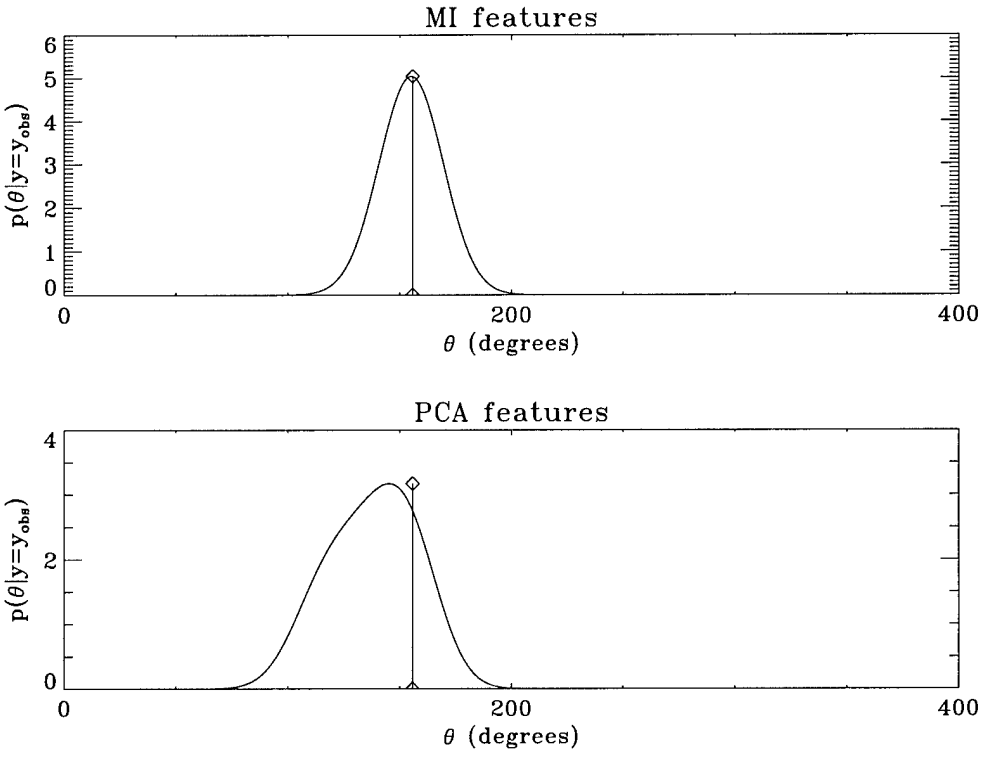


Figure 35: Training image whose projection lies in benign region of MI feature space.

examples are taken from a benign region of the MI feature space. The vertical line indicates the actual pose of the vehicle under test. Unsurprisingly, the MAP estimate using the MI features performs very well on the training sample from the benign region (figure 35), although the PCA features give nearly the same result. In figure 36 the result is quite different. Although the MAP estimate using the MI features is close, as can be seen in the figure another mode in the estimated density is nearly as high. Still, the MI features yield a density which is much less uncertain than the one obtained using the PCA features. The results of figures 37 and 38 are of more interest as they apply to testing data. Once again in the benign case the MI features slightly outperform the PCA features (both in the bias of the estimate and the compactness of the estimated density). Finally the MI features perform much better than the PCA features for the testing sample taken from the ambiguous region.

6.5 Conclusions

We have presented what we believe to be a quite general technique for information theoretic feature extraction. Our choice of PCA features as a comparison was primarily to illustrate the difference in the nature of the feature selection approach, that is, signal representation versus information preservation. Nevertheless,

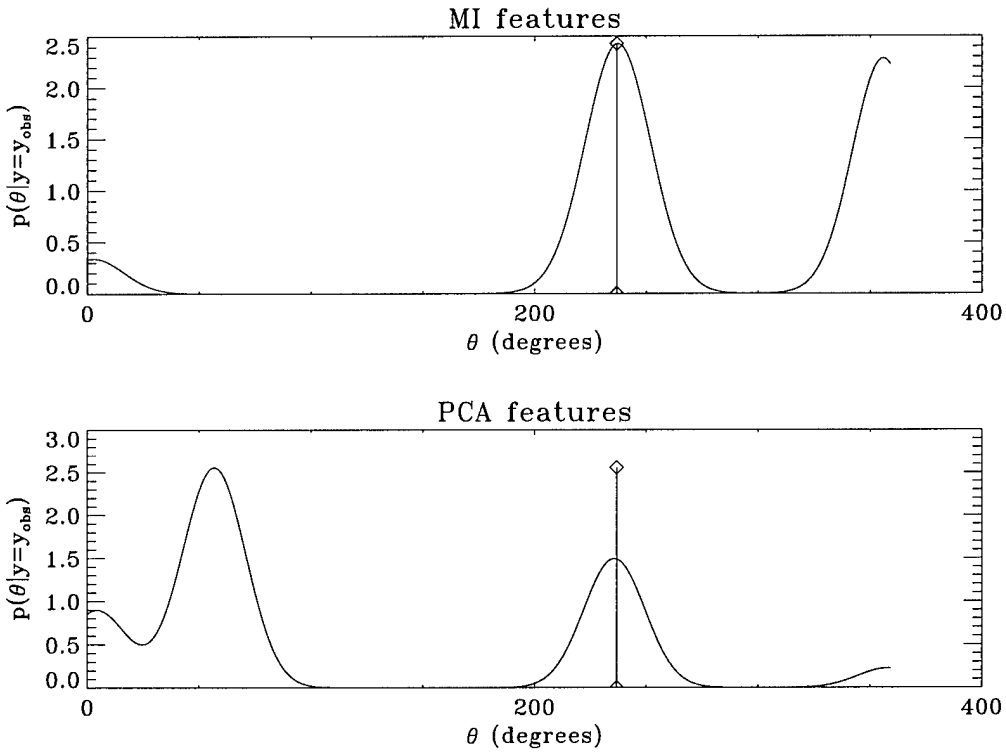


Figure 36: Training sample whose projection lies in ambiguous region of MI feature space.

we feel that these preliminary results are quite promising. Certainly more exhaustive experimentation will have to be done in order to characterize the approach more fully.

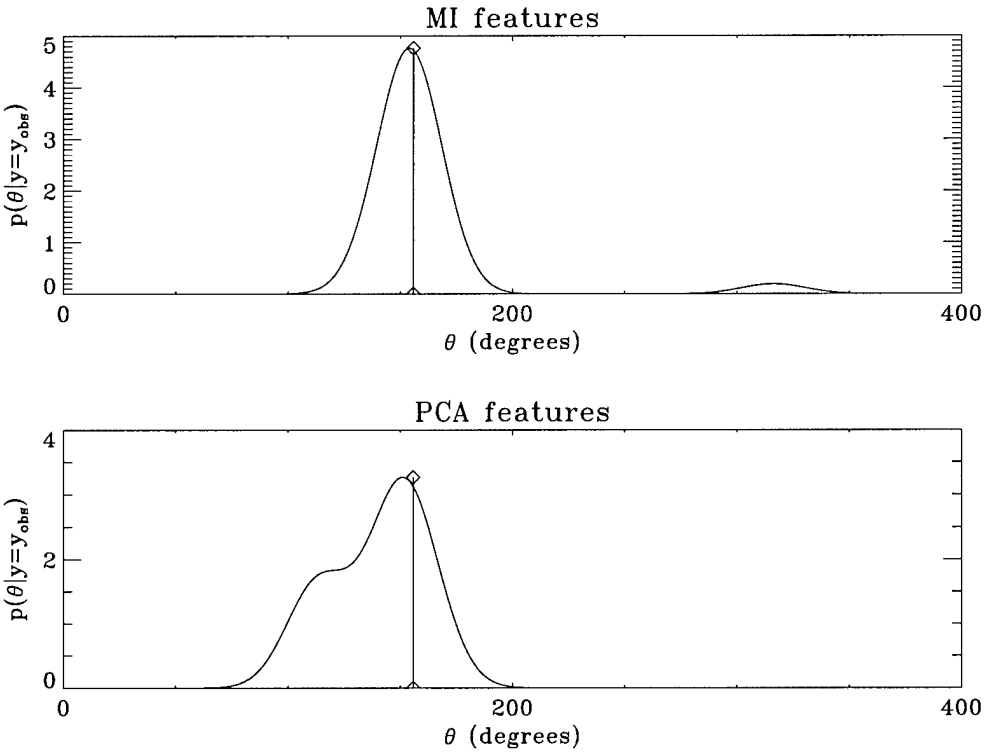


Figure 37: Testing image whose projection lies in benign region of MI feature space.

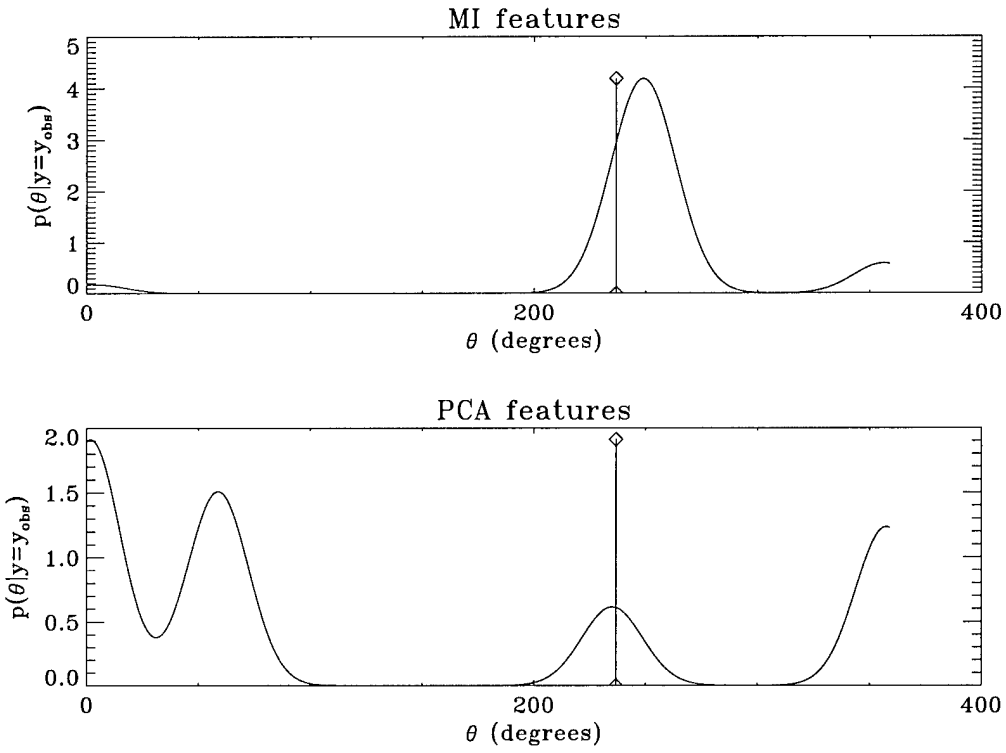


Figure 38: Testing image whose projection lies in ambiguous region of MI feature space.

7 Target model generation from multiple SAR images

An important problem driving much research in the SAR and model-based ATR communities is the generation and modification of target models for ATR system databases. We propose a method for generating or updating 3-D reflector primitive target models. We utilize an existing 2-D extraction algorithm to extract feature locations and classifications (such as scattering primitive type) from each image in a set of SAR data. We formulate the 3-D model generation in terms of a data association problem. We present an iterative algorithm, based on the expectation-maximization (EM) method, to solve the data association problem and yield a maximum likelihood estimate of target feature locations and types from the set of 2-D extracted features. Finally, we present examples and results for sets of simulated SAR imagery.

7.1 Introduction

Recent research into model-based approaches to SAR ATR has yielded significant results. As the sophistication of model-based ATR techniques and model representations grows, however, an increasingly difficult problem is the generation and modification of diverse model libraries necessary for such an approach. Model generation is typically a tedious and difficult task, often requiring detailed descriptions of targets in the form of blueprints or CAD models. Unfortunately, the very targets which are of utmost interest in ATR applications are often those for which little or no detailed prior information is available. An important goal driving much research in the radar and ATR communities is the development of methods for generating reflector primitive target models directly from SAR imagery. Recently, efforts to generate models from a single 1-D radar range profile [40, 41] or a single 2-D SAR image [42, 43] have met with some success. However, the models generated from these data sets are of limited use to most ATR systems because they are not three-dimensional.

Many model-based ATR systems rely on a detailed representation of targets in terms of a small set of canonical reflector primitives such as flat plates, cylinders, tophats, dihedrals, and trihedrals. These reflector primitives enable compact yet physically relevant descriptions of a rich class of targets. A reflector primitive model describes a target in terms of a small set of parameters, specifying information relevant to target recognition and signature prediction, such as the location, type, pose, and size of each of the target's component primitives.

This paper addresses the problem of generating 3-D reflector primitive models for simple targets directly from SAR imagery. Our method relies on a pre-processing step that extracts scattering centers from each 2-D SAR image in the data set, and the expectation-maximization (EM) method to associate extractions implicitly between images. Section 7.2 describes the problem setup and pre-processor; Section 7.3 describes our application of the EM method to the problem; Section 8.4 presents experimental results obtained using XPatch SAR data. Section 8.5 concludes the paper.

7.2 3-D Target Model Extraction

We are given a set of K polarimetric spotlight-mode SAR images (each comprising linear polarizations HH, HV, and VV). Each image k was obtained at an arbitrary look angle defined by center azimuth ϕ_k and depression ψ_k. Perfect ground truth data is available for ϕ_k, ψ_k, and all imaging parameters (such as bandwidth, aperture, range and cross-range locations of each pixel). We wish to extract a 3-D target reflector primitive model directly from these images. This model will describe the target as a collection of discrete scattering primitives. The model will describe the 3-D ground location and discrete type (e.g. , trihedral, tophat) of each target primitive, as well as the radius of curvature for certain curved primitives.

We perform a simple scattering center extraction on each polarimetric vector-intensity SAR image as a pre-processing step, and use the resulting set of K parameterizations as the basis for the 3-D model generation. Examining only the 2-D scattering center extractions instead of the complete data provided by the full set of raw SAR imagery necessarily overlooks information that would be relevant to 3-D model construction; however, this approach greatly simplifies the model generation problem and provides a springboard for ongoing and future research.

7.2.1 2-D Scattering Center Extraction Pre-Processor

The pre-processor used to extract scattering centers from each polarimetric SAR image set provides estimates of the slant-plane locations and discrete types of prominent scattering centers in each polarimetric SAR image. The pre-processor has two distinct steps: first, the peak detector from the MSTAR "extract" module [44] is used to estimate scattering center locations from a noncoherent sum of the three polarimetric images at each look angle (i.e. , the pixelwise polarimetric vector magnitude). The peak detector identifies all local maxima with peak-to-background intensity above a tunable threshold, subject to a proximity constraint; bilinear interpolation is used to yield a subpixel estimate of scattering center location in slant-plane coordinates. The second step is a polarimetric type classification. The extracted polarimetric signature at each extracted peak location is classified as one of M discrete scattering primitive types, using an M-ary type classification procedure based on canonical polarimetric scattering signatures [45]. The pre-processor thus produces a set of K parameterizations from the K polarimetric SAR images, each describing a set of extracted scattering centers in terms of their estimated 2-D slant-plane locations and discrete primitive types.

The front-end pre-processor does not do a perfect job of detecting and classifying scattering centers: there will be missed detections, false alarms, extracted location uncertainty, and incorrect classifications. Characterization of pre-processor performance in terms of these considerations will prove essential to the design of our model generation algorithm, described below.

7.2.2 Definitions and Notation

We now introduce notation that will aid in the development of a model characterizing the front-end 2-D extraction process. (Figure 47, described at the end of this section, illustrates some of the notation presented here.) The set of parameters associated with each 2-D extracted scattering center (i.e. , each combined location and type output from the pre-processor) will be known as a report. The full set of reports for any given SAR image or phase history will be called a frame. All extracted parameters will be indexed by an argument and subscript, denoting the frame and report to which it corresponds. For instance, $y_m(k)$ denotes the estimated slant plane cross-range location y for the mth report extracted from frame k. Finally, $M(k)$ will denote the number of reports (extracted scattering centers) in frame k.

The target model will describe the 3-D location and discrete type of each primitive, and the radius of certain curved primitives. This model must be estimated from the output of the pre-processor, namely, from the K frames of extraction parameters. Each report comprises three parameters: $\alpha_m(k)$, specifying estimated discrete scattering type, and $x_m(k)$ and $y_m(k)$, specifying estimated slant plane down-range and cross-range location. It will be convenient to define the quantity $X_m(k) = [x_m(k)\ y_m(k)]^T$ combining the estimated slant plane down-range and cross-range locations for each report into a single vector quantity. $Z_m(k)$ will denote a single report, i.e. , the full set of parameters associated with any extraction: $Z_m(k) = \{X_m(k), \alpha_m(k)\}$. Further notation will include $Z(k)$ to denote the full set of reports for frame k ($Z(k) = \{Z_m(k)\}_{m=1}^{M(k)}$), and Z^K to denote the full set of reports in all frames ($Z^K = \{Z(k)\}_{k=1}^{K} = \{\{Z_m(k)\}_{m=1}^{M(k)}\}_{k=1}^{K}$).

The set Z^K represents all of the data provided by the front-end extraction process. The 3-D target primitive model must be estimated from Z^K. Let N_t denote the number of primitives comprising the target. We will let Θ_t denote the set of parameters specifying the ground location and discrete type (and radius of curvature, if applicable) of target primitive t (for $t = 1, \ldots, N_t$). The notation Θ_t^X will refer to the ground location of reflector primitive t, i.e. , $\Theta_t^X = [x_t \;\; y_t \;\; z_t]^T$ where x_t, y_t, and z_t are the ground-frame Cartesian coordinates of reflector primitive t. Similarly, Θ_t^α will refer to the discrete type of reflector primitive t, i.e. , to the index specifying which of N_α possible discrete type classes primitive t represents. Likewise, Θ_t^r will refer to the scalar radius of curvature of primitive t (if applicable), with respect to its axis of symmetry. The symbol Θ will denote the complete, all-encompassing parameter vector describing the target model, i.e. , $\Theta = \{\Theta_t\}_{t=1}^{N_t}$. Our goal is to produce an estimate of Θ given Z^K. On occasion we will find it useful to indicate only the discrete types or location parameters of the target model. For this purpose we define $\Theta^X = \{\Theta_t^X\}_{t=1}^{N_t}$, $\Theta^\alpha = \{\Theta_t^\alpha\}_{t=1}^{N_t}$, and $\Theta^r = \{\Theta_t^r\}_{t=1}^{N_t}$.

In order to estimate Θ from Z^K, we must specify how Z^K depends on Θ. To do this, it will be convenient to define an unobservable label parameter for each report $Z_m(k)$ as follows:

$$\lambda_m(k) = \begin{cases} t, & \text{if report } m \text{ in frame } k \text{ corresponds to target primitive } t \\ 0, & \text{if report } m \text{ in frame } k \text{ is spurious (corresponds to no primitive).} \end{cases} \tag{60}$$

Thus $\lambda_m(k)$ is an integer index specifying which of the N_t target primitives, if any, gave rise to report $Z_m(k)$. As before, we introduce notation to refer to sets of these parameters within each frame and throughout all frames. Specifically, let $\lambda(k) = [\lambda_1(k), \cdots, \lambda_{M(k)}(k)]^T$ and $\lambda^K = \{\lambda(k)\}_{k=1}^K$. The set λ^K is unobservable because Z^K does not explicitly contain information about these labels (although it depends on them implicitly). Note that $\lambda(k)$ is subject to certain restrictions: no more than N_t of its elements may be nonzero, it cannot contain the same nonzero index twice, and so on. (Figure 47 contains an example of label parameter assignments.) We will denote the space of all possible $\lambda(k)$ by $\Lambda(k)$; similarly, we will denote the space of all possible λ^K by $\Lambda^K = \Lambda(1) \times \cdots \times \Lambda(K)$.

Finally, it will prove convenient to define several parameters that are explicit functions of $\lambda(k)$. We introduce a primitive detection indicator function $\delta_t(k)$, where

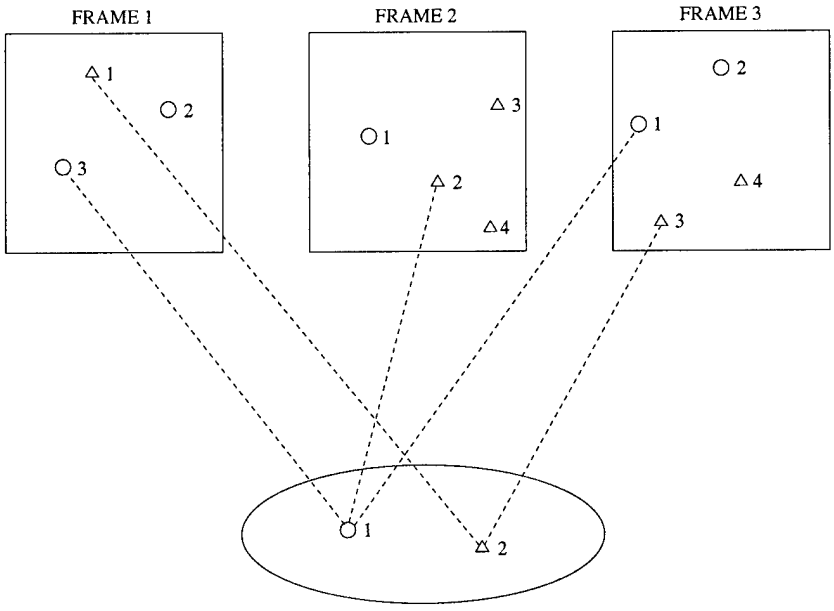
$$\delta_t(k) = \begin{cases} 1, & \text{if } \lambda_m(k) = t \text{ for some } 1 \leq m \leq M(k), \\ 0, & \text{otherwise.} \end{cases} \tag{61}$$

Thus $\delta_t(k)$ indicates whether primitive t generated a report in frame k. A framewise vector of target indicators can be built by concatenating the individual target detection indicators: $\delta(k) = [\delta_1(k), \cdots, \delta_{N_t}(k)]^T$. Finally, we define detection and false alarm totals for each frame: let $D(k) = \sum_{t=1}^{N_t} \delta_t(k)$ and $F(k) = M(k) - D(k) = |\lambda(k)| - D(k)$ such that $D(k)$ is the number of target primitives that generated reports in frame k, and $F(k)$ is the number of spurious reports in frame k.

Figure 47 is a pictorial depiction of some of the notation described in this section. In this figure, there are two target primitives ($N_t = 2$), three frames ($K = 3$), and two possible types ($N_\alpha = 2$). The primitives comprising the target and the reports comprising each frame are numbered in the figure. If the triangle symbol corresponds to type 1, and the circle symbol to type 2, then we have $\Theta_1^\alpha = 2$, $\Theta_2^\alpha = 1$, and other quantities as given in the table on the right of the figure.

7.2.3 Model Generation as a Data Association Problem

Reports generated by reflector primitives appear approximately in locations determined by their projections into the slant plane [46]. Recall that image k was obtained at center azimuth ϕ_k and depression ψ_k. An



FRAME 1	*FRAME 2*	*FRAME 3*
$M(1) = 3$	$M(2) = 4$	$M(3) = 4$
$\alpha_1(1) = 1$ $\alpha_2(1) = 2$ $\alpha_3(1) = 2$	$\alpha_1(2) = 2$ $\alpha_2(2) = 1$ $\alpha_3(2) = 1$ $\alpha_4(2) = 1$	$\alpha_1(3) = 2$ $\alpha_2(3) = 2$ $\alpha_3(3) = 1$ $\alpha_4(3) = 1$
$\lambda_1(1) = 2$ $\lambda_2(1) = 0$ $\lambda_3(1) = 1$	$\lambda_1(2) = 0$ $\lambda_2(2) = 1$ $\lambda_3(2) = 0$ $\lambda_4(2) = 0$	$\lambda_1(3) = 1$ $\lambda_2(3) = 0$ $\lambda_3(3) = 2$ $\lambda_4(3) = 0$
$\delta_1(1) = 1$ $\delta_2(1) = 1$	$\delta_1(2) = 1$ $\delta_2(2) = 0$	$\delta_1(3) = 1$ $\delta_2(3) = 1$
$D(1) = 2$	$D(2) = 1$	$D(3) = 2$
$F(1) = 1$	$F(2) = 3$	$F(3) = 2$

Figure 39: Notation example

approximate model for the extracted location of a scattering center in the slant plane is given by

$$X_m(k) = f\left(\Theta_{\lambda_m(k)}, \psi_k, \phi_k\right) + w_m(k) \tag{62}$$

where $f(\bullet)$ is a projection operator described below, and where $w_m(k)$ is white Gaussian noise (uncorrelated between reports and frames) with mean 0 and covariance R_k. This model does not capture coupling effects between primitives, such as obstruction and multi-bounce scattering, but succeeds in describing the apparent location of primitives in SAR images under many circumstances. The atomic (i.e. , non-coupled) form of $f(\bullet)$ and the whiteness of $w_m(k)$ will enable us to decouple the model estimation problem into N_t independent primitive estimation problems.

Many target primitives produce a SAR response whose apparent specular reflection point remains essentially constant as viewing angle varies: trihedrals, dihedrals, and point scatterers are of this type.[47] Let us denote the set of discrete type indices α for these primitives as I_{fixed}. Other target primitives with curved surfaces, however, produce a response whose apparent location varies with viewing angle; let us denote the set of discrete type parameters for these primitives as I_{curv}. Cylinders, tophats, and spheres are of this latter category.[47] If we restrict I_{curv} to primitives with an axis of symmetry perpendicular to the ground plane, then $f(\bullet)$ takes the following form:

$$f\left(\Theta_{\lambda_m(k)}, \psi_k, \phi_k\right) = \begin{cases} A_k \Theta^X_{\lambda_m(k)}, & \Theta^\alpha_{\lambda_m(k)} \in I_{fixed} \\ A_k \Theta^X_{\lambda_m(k)} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Theta^r_{\lambda_m(k)} \cos(\psi_k), & \Theta^\alpha_{\lambda_m(k)} \in I_{curv} \end{cases} \tag{63}$$

where A_k is the ground-to-slant-plane transformation matrix for look angle (ψ_k, ϕ_k):

$$A_k = \begin{bmatrix} \cos\psi_k \cos\phi_k & \cos\psi_k \sin(\phi)_k & -\sin(\psi)_k \\ -\sin(\phi)_k & \cos\phi_k & 0 \end{bmatrix}. \tag{64}$$

(The form of $f(\bullet)$ for primitives in class I_{curv} captures the fact that the radar responses from these primitives appear to emanate up-range from the center of the primitive, due to their convex surfaces.)

Suppose that λ^K were known exactly. Then estimation of Θ^X would be straightforward via a weighted least-squares approach. Estimation of Θ^α and Θ^r would also be straightforward given λ^K. Unfortunately, λ^K is unknown. If we want to base our estimation on this information, we will need to find a way to estimate λ^K from Z^K, the observable data. The estimation of λ^K is essentially a correspondence or data association problem; a large body of literature exists describing theory and methods for solving data association problems in many contexts and applications. [48, 49, 50] We will utilize the expectation-maximization (EM) method [51], which has been used in numerous applications to facilitate simultaneous consideration of a large set of possible data associations [50]. In order to implement the EM method, described in Section 7.3, we will first need to derive an expression for the pdf $p\left(\lambda^K, Z^K; \Theta\right)$. This entails specifying a model for the behavior of the pre-processor.

7.2.4 Model Generation Front End: the Pre-Processor

Given a number of reasonable assumptions about the 2-D scattering center extractor pre-processor, we will be able to derive an expression for $p\left(\lambda^K, Z^K; \Theta\right)$, the pdf for λ^K and Z^K parameterized by Θ. The whiteness of the noise in (62) means that the $X_m(k)$ are conditionally independent given Θ. We will assume the same of the discrete type indices $\alpha_m(k)$. (This assumption, like the one that allowed us to write (62), is violated for effects like multi-bounce reflections; however, this assumption is reasonable in many circumstances, and

will allow us to treat distinct primitives and distinct frames independently, greatly simplifying the model estimation problem.) Specifically, we will assume that the probability that a primitive of type α_1 is classified as being of type α_2 in an arbitrary frame (given that it generates a report in that frame) is given by the confusion probability $\rho_{\alpha_1\alpha_2}$. Suppose that report $Z_m(k)$ is not spurious; then the prior probability of that report's estimated type being $\alpha_m(k)$ (given that it was detected in frame k) is $\rho_{\Theta^{\alpha}_{\lambda_m(k)}\alpha_m(k)}$. Because this notation is unwieldy at best, we define the shorthand notation $\rho'_{m(k)} = \rho_{\Theta^{\alpha}_{\lambda_m(k)}\alpha_m(k)}$.

The front-end pre-processor has two distinct stages: namely, a peak extraction stage and a classification stage. Since these stages are distinct and memoryless, and because the peak extractor operates on the magnitude of the polarimetric vector, while the classifier utilizes its direction, we will make the reasonable assumption that the type confusion and location uncertainty processes are independent between reports and frames. Finally, we assume that primitive detection and false alarm processes are independent across frames. While this is not true in many cases—consider the case of a primitive with a highly directional response, such as a flat plate or a trihedral—it has the great benefit (when taken with the other independence assumptions we have made) of enabling us to represent $p\left(\lambda^K, Z^K; \Theta\right)$ as a product of independent terms. Specifically, we may write

$$p\left(\lambda^K, Z^K; \Theta\right) = \prod_{k=1}^{K} p\left(\lambda(k), Z(k); \Theta\right) \tag{65}$$

which means that if estimation of λ^K is our goal, it may be accomplished by solving K smaller estimation problems—namely, K decoupled estimations of $p\left(\lambda(k), Z(k); \Theta\right)$.

Experimentation suggests that false alarms may be modeled as a Poisson process (we shall denote its arrival rate by γ_{FA}) with locations uniformly distributed across the sensor volume V and with types equally distributed among the N_α possibilities. Let us denote the probability of detection of a primitive of type α in an arbitrary frame as $P_D(\alpha)$ (a function that takes on only N_α values and is constant over all look angles, by our framewise detection independence assumption above). Given our assumptions in this and previous sections, we can now fully specify $p\left(\lambda(k), Z(k); \Theta\right)$. The full derivation of this quantity is given in Section 7.6; this derivation yields the expression

$$\begin{aligned} p\left(\lambda(k), Z(k); \Theta\right) \;=\; & \frac{e^{-\gamma_{FA}V}\left(\frac{\gamma_{FA}}{N_\alpha}\right)^{F(k)}}{M(k)!}\left(\prod_{t=1}^{N_t}\left(1 - P_D\left(\Theta_t^{\alpha}\right)\right)\right)\left(\prod_{\lambda_m(k)\neq 0}\frac{P_D(\Theta^{\alpha}_{\lambda_m(k)})}{1 - P_D(\Theta^{\alpha}_{\lambda_m(k)})}\right) \\ & \left(\prod_{\lambda_m(k)\neq 0}\frac{\rho'_{m(k)}}{2\pi\left(\det R_k\right)^{1/2}}\exp\left[-\frac{1}{2}\left(f\left(\Theta_{\lambda_m(k)}, \psi_k, \phi_k\right) - X_m(k)\right)^T R_k^{-1}\right.\right. \\ & \left.\left.\left(f\left(\Theta_{\lambda_m(k)}, \psi_k, \phi_k\right) - X_m(k)\right)\right]\right) \end{aligned} \tag{66}$$

where $f(\bullet)$ is the projection operator defined in (89). It is clear that (95) gives a complete model for the probabilistic dependence of $Z(k)$ and $\lambda(k)$ (and thus Z^K and λ^K via (65)) on Θ. We see from the above equation that the dependence of report parameters Z^K and unobservable label parameters λ^K on Θ is parameterized by several attributes of the front-end pre-processor—namely, the false alarm rate γ_{FA}, the probabilities of detection $P_D\left(\Theta_t^{\alpha}\right)$, the type confusion probabilities $\rho'_{m(k)}$, and the extraction location covariances R_k. If these pre-processor parameters are known in advance, then (95) provides a complete

description of the dependence of Z^K and λ^K on Θ; if these pre-processor parameters are unknown, they may be estimated by running the pre-processor on characteristic imagery for which target ground truth data is available.

7.3 EM Formulation of Model Generation

We have cast the target model generation problem in the framework of a data association problem; we have made several simplifying assumptions that allow us to specify the probabilistic dependence of the observed data (Z^K) and the unobservable report-to-primitive associations (λ^K) on the complete target parameter vector Θ in (95). We now turn to the expectation-maximization (EM) method as a tool to produce a target model within this framework. The EM method is an iterative procedure for producing a maximum likelihood estimate of parameters in incomplete-data problems, i.e. , when there is a many-to-one mapping from a postulated set of "complete" data to the set of observed data.[50, 51] In our context, the complete data is the full set of all observations and label parameters, Z^K and λ^K; there are combinatorically many primitive-to-report associations $\lambda^K \in \Lambda^K$ that could give rise to the observations Z^K.

Each iteration of the EM method consists of two steps: an expectation (E) step and a maximization (M) step. The E step takes into account all feasible association possibilities by calculating the expectation over λ^K of the log likelihood of Θ given the observed data Z^K and assuming the previous parameter iterate $\Theta^{[n]}$. (In our context, the E step entails calculation of the report-to-primitive association probabilities for each feasible report-primitive pair.) The M step produces a new iterated estimate, $\Theta^{[n+1]}$, by maximizing the expected log likelihood from the E step over Θ. The E and M steps are alternated until $p(Z^K;\Theta)$ converges. Under relatively weak assumptions, the EM method is guaranteed to converge to at least a local maximum of the likelihood function $p(Z^K;\Theta)$ [51, 52].

Suppose that we have an initial guess $\Theta^{[0]}$ for Θ. The E step of the nth iteration of the EM algorithm requires calculation of the quantity

$$Q\left(\Theta;\Theta^{[n]}\right) = E\left[\log p\left(\lambda^K, Z^K;\Theta\right)\middle| Z^K;\Theta^{[n]}\right] = \sum_{\lambda^K\in\Lambda^K}\left[\log p\left(\lambda^K, Z^K;\Theta\right)\right] p\left(\lambda^K\middle| Z^K;\Theta^{[n]}\right). \tag{67}$$

The M step then requires maximization of $Q\left(\Theta;\Theta^{[n]}\right)$ over Θ. Specifically, the M step determines the next iterate value $\Theta^{[n+1]} = \arg\max_\Theta Q\left(\Theta;\Theta^{[n]}\right)$. We now derive the precise forms the E and M steps for the application of the EM method to the 3-D model generation problem, and discuss the initialization and termination of the algorithm.

7.3.1 Formulation of the E Step

We seek an expression for the expectation $Q\left(\Theta;\Theta^{[n]}\right)$ required in the E step of the EM algorithm. The derivation of Section 7.7 shows that

$$Q\left(\Theta;\Theta^{[n]}\right) = \sum_{t=1}^{N_t} Q_t\left(\Theta_t;\Theta^{[n]}\right) + c_K \tag{68}$$

where

$$Q_t\left(\Theta_t;\Theta^{[n]}\right) = \sum_{k=1}^{K}\Bigg[\log\left(1-P_D\left(\Theta_t^{\alpha}\right)\right)+\sum_{m=1}^{M(k)}\Bigg(\log\left(P_D\left(\Theta_t^{\alpha}\right)\right)-\log\left(1-P_D\left(\Theta_t^{\alpha}\right)\right)+\log\rho'_{m(k)}-$$
$$\frac{1}{2}\left(f\left(\Theta_t,\psi_k,\phi_k\right)-X_m(k)\right)^T R_k^{-1}\left(f\left(\Theta_t,\psi_k,\phi_k\right)-X_m(k)\right)\Bigg)$$
$$\Pr\left(\lambda_m(k)=t\,\middle|\,Z(k);\Theta^{[n]}\right)\Bigg] \tag{69}$$

and c_K is constant with respect to Θ. Note that $Q\left(\Theta;\Theta^{[n]}\right)$ in (97) is separable into N_t terms $Q_t\left(\Theta_t;\Theta^{[n]}\right)$ each depending only on a single primitive, and that each $Q_t\left(\Theta_t;\Theta^{[n]}\right)$ in (69) is further separable into K terms each depending only on a single frame. This is the benefit of our independence assumptions of Section 7.2: the E step can be achieved frame by frame (and, as we will discuss shortly, the M step can be achieved primitive by primitive). This is an encouraging result, because it means that computational complexity of the E step will increase only linearly with K, and not exponentially as might have been expected if the E step had in fact required enumeration of the space Λ^K (which grows exponentially with K). We also see that all the terms in (69) but one have been specified exactly: we have not yet described how to obtain the report-to-primitive association probabilities $\Pr\left(\lambda_m(k)=t\,\middle|\,Z(k);\Theta^{[n]}\right)$. Completion of the E step thus requires us to calculate this probability for all (m,t) pairs in frame k. Bayes' rule allows us to write

$$\Pr\left(\lambda_m(k)=t\,\middle|\,Z(k);\Theta^{[n]}\right)=\frac{\sum_{\lambda_m(k)=t}p\left(\lambda(k),Z(k);\Theta^{[n]}\right)}{\sum_{\lambda(k)\in\Lambda(k)}p\left(\lambda(k),Z(k);\Theta^{[n]}\right)}. \tag{70}$$

For any $Z(k)$, $\lambda(K)$, and $\Theta^{[n]}$, (95) gives $p\left(\lambda(k),Z(k);\Theta^{[n]}\right)$. This appears to provide the means for computing $\Pr\left(\lambda_m(k)=t\,\middle|\,Z(k);\Theta^{[n]}\right)$. Unfortunately, calculation of this probability by way of (70) is generally intractable due to the enormous size of $\Lambda(k)$.

Examining (95), we see that the Gaussian exponential term ensures that $\Pr\left(\lambda_m(k)=t\,\middle|\,Z(k);\Theta^{[n]}\right)\approx 0$ except when $f\left(\Theta_{\lambda_m(k)},\psi_k,\phi_k\right)$ is relatively near $X_m(k)$ for all $m=1,\ldots,M(k)$. That is, the association of a measurement with a distant target primitive is extremely unlikely. This suggests that the size of $\Lambda(k)$ can be reduced by excluding all distant associations. One way to do this is by gating. Gating is a procedure often used in data association problems [48] to reduce the size of the feasible association set by including only those associations that meet a proximity criterion. In this context, gating could be implemented by including in $\Lambda(k)$ only those $\lambda(k)$ that satisfy the criterion $\left\|X_m(k)-f\left(\Theta_{\lambda_m(k)},\psi_k,\phi_k\right)\right\|_2\le r_{gate},\ m=1,\ldots,M(k)$. (Proximity can be quantified by defining r_{gate} as a function of R_k.) Gating greatly simplifies the E step calculation by allowing us to dismiss out of hand a majority of all possible associations as being infeasible, and considering only those associations that are relatively likely. Gating thus has the important effect of reducing the required enumeration and sets of calculation from the billions to thousands, enabling us to implement the E step. The price of gating is the possibility (albeit unlikely, if the gate radius is large enough) that a report generated by the gated primitive falls outside of the gate region and is thus neglected.

7.3.2 Formulation of the M step

We now examine how to maximize $Q\left(\Theta;\Theta^{[n]}\right)$ in (97) to obtain iterated estimates of the target parameters. First of all, because c_K is constant with respect to Θ, it does not affect the maximization and thus can

be neglected as far as the M step is concerned. Additionally, because our reflector primitive independence assumptions enabled us to write the non-constant portion of $Q(\Theta;\Theta^{[n]})$ as a sum of terms of the form $Q_t(\Theta_t;\Theta^{[n]})$, the maximization is independent for each primitive. This is good news, because it indicates that the M step will require not a joint maximization of N_t parameter sets, but rather N_t independent maximizations.

Each $Q_t(\Theta_t;\Theta^{[n]})$ must be maximized over continuous parameters (Θ_t^X and possibly Θ_t^r) and a discrete parameter (Θ_t^α). Because Θ_t^α is chosen from a discrete (and typically small) set, we can achieve the maximization of $Q_t(\Theta_t;\Theta^{[n]})$ by maximizing over Θ_t^X and Θ_t^r with Θ_t^α fixed at each of its N_α possible values, then picking the largest of the N_α resulting quantities. Define $Q_{t,\alpha}(\Theta_t;\Theta^{[n]}) = Q_t(\Theta_t;\Theta^{[n]})\big|_{\Theta_t^\alpha=\alpha}$ for each $\alpha = 1,\ldots,N_\alpha$. Let $\hat{\Theta}_{t,\alpha}^X$ and $\hat{\Theta}_{t,\alpha}^r$ be the location and radius parameters that maximize $Q_{t,\alpha}(\Theta_t;\Theta^{[n]})$for each α. Then the type $\hat{\Theta}_t^\alpha$ that maximizes $Q_t(\Theta_t;\Theta^{[n]})$ is given by

$$\hat{\Theta}_t^\alpha = \arg\max_\alpha Q_{t,\alpha}(\Theta_t;\Theta^{[n]})\big|_{\Theta_t^X=\hat{\Theta}_{t,\alpha}^X,\ \Theta_t^r=\hat{\Theta}_{t,\alpha}^r} \tag{71}$$

and the maximizing location and radius are given by

$$\hat{\Theta}_t^X = \hat{\Theta}_{t,\hat{\Theta}_t^\alpha}^X \tag{72}$$

and

$$\hat{\Theta}_t^r = \hat{\Theta}_{t,\hat{\Theta}_t^\alpha}^r \tag{73}$$

respectively. Thus we can maximize $Q_t(\Theta_t;\Theta^{[n]})$ by performing N_α maximizations over Θ_t^X and Θ_t^r. We now show how these maximizations may be achieved.

In the case when $\Theta_t^\alpha \in I_{fixed}$, we must maximize $Q_{t,\alpha}(\Theta_t;\Theta^{[n]})$ over Θ_t^X only. In this case, (69) becomes

$$\begin{aligned} Q_{t,\alpha}(\Theta_t;\Theta^{[n]}) &= \sum_{k=1}^{K}\left(\sum_{m=1}^{M(k)}\left[c_\alpha - \frac{1}{2}(A_k\Theta_t^X - X_m(k))^T R_k^{-1}(A_k\Theta_t^X - X_m(k))\right]\right. \\ &\qquad \left.\Pr\left(\lambda_m(k)=t\,\middle|\,Z(k);\Theta^{[n]}\right)\right) \end{aligned} \tag{74}$$

where c_α is constant with Θ_t^α fixed, and the remaining term is quadratic in Θ_t^X. Then straightforward calculations show that the maximizing Θ_t^X for iteration $n+1$ is given by

$$\begin{aligned} \Theta_t^{X[n+1]} &= \left[\sum_{k=1}^{K} A_k^T R_k^{-1} A_k \cdot \Pr\left(\delta_t(k)=1\,\middle|\,Z(k);\Theta^{[n]}\right)\right]^{-1} \times \\ &\qquad \left[\sum_{k=1}^{K}\sum_{m=1}^{M(k)} A_k^T R_k^{-1} X_m(k) \cdot \Pr\left(\lambda_m(k)=t\,\middle|\,Z(k);\Theta^{[n]}\right)\right] \end{aligned} \tag{75}$$

where the inverse explicit in the first term exists except in degenerate cases (i.e. , except when the rows of all $\{A_k\}_{k=1}^K$ fail to form a basis for $\Re^3$).

When $\Theta_t^\alpha \in I_{curv}$, we must maximize $Q_t\left(\Theta_t;\Theta^{[n]}\right)$ over Θ_t^X and Θ_t^r simultaneously. In this case, (69) becomes

$$Q_{t,\alpha}\left(\Theta_t;\Theta^{[n]}\right) = \sum_{k=1}^{K}\sum_{m=1}^{M(k)}\left[c_\alpha - \frac{1}{2}\left(A_k\Theta_t^X - X_m(k) - \begin{bmatrix}1\\0\end{bmatrix}\Theta_t^r\cos(\psi_k)\right)^T R_k^{-1}\cdot \left(A_k\Theta_t^X - X_m(k) - \begin{bmatrix}1\\0\end{bmatrix}\Theta_t^r\cos(\psi_k)\right)\right]\Pr\left(\lambda_m(k)=t\,\middle|\,Z(k);\Theta^{[n]}\right). \quad (76)$$

Proceeding by setting partial derivatives $\frac{\partial}{\partial\Theta_t^X}\left[Q_{t,\alpha}\left(\Theta_t;\Theta^{[n]}\right)\right]$ and $\frac{\partial}{\partial\Theta_t^r}\left[Q_{t,\alpha}\left(\Theta_t;\Theta^{[n]}\right)\right]$ to zero, and checking the second derivative to ensure maximization, we arrive at a necessary and sufficient condition for maximization of $Q_{t,\alpha}\left(\Theta_t;\Theta^{[n]}\right)$:

$$\begin{bmatrix}\Theta_t^{X[n+1]}\\ \Theta_t^{r[n+1]}\end{bmatrix} = \left(\sum_{k=1}^{K}\begin{bmatrix} A_k^TR_k^{-1}A_k & -A_k^TR_k^{-1}\begin{bmatrix}1\\0\end{bmatrix}\cos\psi_k \\ -\begin{bmatrix}1 & 0\end{bmatrix}R_k^{-1}A_k\cos\psi_k & \begin{bmatrix}1 & 0\end{bmatrix}R_k^{-1}\begin{bmatrix}1\\0\end{bmatrix}\cos^2\psi_k\end{bmatrix}\Pr\left(\delta_t(k)=1\,\middle|\,Z(k);\Theta^{[n]}\right)\right)^{-1}\cdot \sum_{k=1}^{K}\sum_{m=1}^{M(k)}\left(\begin{bmatrix} A_k^TR_k^{-1}X_m(k)\\ -\begin{bmatrix}1 & 0\end{bmatrix}R_k^{-1}\cos\psi_k X_m(k)\end{bmatrix}\Pr\left(\lambda_m(k)=t\,\middle|\,Z(k);\Theta^{[n]}\right)\right) \quad (77)$$

which is essentially an augmented version of (75). Again, the inverse implicit in the first term of (77) will exist except in degenerate cases.

We see that (75) and (77) both take the form of a weighted least-squares solution to the parameter estimation problem, where the weights are determined by the probabilities of association. Additionally, we see that in either case, the maximization of $Q_t\left(\Theta_t;\Theta^{[n]}\right)$ over Θ_t requires only $\mathcal{O}(N_\alpha\sum_{k=1}^{K}M(k))$ operations; thus the M step in each iteration will require only $\mathcal{O}(N_tN_\alpha\sum_{k=1}^{K}M(k))$ operations. The E step, on the other hand, requires the calculation of $p\left(\lambda(k),Z(k);\Theta^{[n]}\right)$ for all points in the gating-reduced Λ^K space. Although gating reduces the space to a manageable size, the calculations are still much more burdensome than the relatively trivial matrix multiplications and inversions required for the M step. Thus the E step will be responsible for the bulk of the computational burden of the algorithm.

7.3.3 Initialization and Termination

We have specified a procedure for producing a sequence of estimates of the target parameters given a previous iterate. We must also describe the initialization and termination criteria. The original proponents of the EM algorithm [51] showed that at each iteration the likelihood term will increase or remain constant, i.e. , $p\left(Z^K;\Theta^{[n+1]}\right)\geq p\left(Z^K;\Theta^{[n]}\right)$, and thus in principle the likelihood $p\left(Z^K;\Theta^{[n]}\right)$could be monitored for convergence. However, since neither the E nor M step of the algorithm requires explicit calculation of $p\left(Z^K;\Theta^{[n]}\right)$, in practice it is more convenient to monitor the estimate iterates $\Theta^{[n]}$ and halt when locations, radii, and types have all reached apparent convergence.

The initialization of the EM method with an initial guess $\Theta^{[0]}$ is a more sophisticated issue. If a prior model of the target is available, this provides a natural iteration. If no prior model is available, then the

Table 4: Imaging parameters

imaging parameter	*value*
image size	32×32
range resolution	0.3 m
cross-range resolution	0.3 m
range pixel spacing	0.2 m
cross-range pixel spacing	0.2 m
center frequency	9.6 GHz
squint angle	90°
sidelobe weighting	−35 dB Kaiser

initialization must be generated by other means. In this paper, we assume that we are provided with an initialization $\Theta^{[0]}$ that differs from truth according to some uncertainty statistics, via a prior model or some other initialization oracle.

7.4 Experimental Results

In this section we present results from the application of our algorithm to synthetic SAR imagery generated by the XPatch electromagnetic simulation software package, based on a facetization model of a simple target. The imaging parameters used as inputs to XPatch are given in Table 4. Clutter was modeled as a polarimetric K-distributed process with parameters (covariance and alpha) set to values typically observed for grassy terrain [53].

As described in Section 7.2, the MSTAR peak detector [44] and a polarimetric type classifier [45] form the front-end pre-processor. Because probabilities of detection, false alarm rates, type confusion probabilities, and location extraction covariances required to calculate quantities needed in the EM iteration are not known *a priori*, they were estimated from characteristic imagery containing all types of interest as a preliminary step to the experiments presented here. The experiments in this section were performed on a target containing only two primitives: a trihedral and a tophat. The estimated pre-processor statistics used for all experiments in this section are given in Table 5, where trihedral is defined as type 1 and tophat is defined as type 2. (These parameters were obtained with a pre-processor extraction signal-to-background threshold of 6.7 dB, and a proximity constraint of 3 pixels.) Note that although a trihedral generally has a brighter specular response than a tophat, its probability of detection is generally smaller because it is visible over a narrower aspect region. Additionally, because a trihedral exhibits a triple-bounce response over only part of this region (it produces a double- or single-bounce response when illuminated far enough from its specular angle) there is a higher probability of confusing a trihedral as a tophat than vice-versa.

7.4.1 Experiment 1

In this section we examine the performance of the algorithm on a simple target comprising two primitives: a one-foot square-plate trihedral at ground coordinates $[x_t \;\; y_t \;\; z_t] = [30'' \;\; 30'' \;\; 0'']$, and a tophat with cylinder radius of six inches and base radius of twelve inches centered at $[-34.24'' \;\; -34.24'' \;\; 0'']$. The estimated pre-processor statistics required for the EM algorithm are given in Table 5. For this experiment, 72 SAR images were generated by XPatch: 36 of these were equally spaced at 10° azimuth intervals from 5° to 355° at depression angle 30°; the other 36 were generated at 45° depression at the same azimuths.

Table 5: Estimated pre-processor statistics

statistic	*notation*	*estimated value*
probabilities of detection	$P_D(1) \quad P_D(2)$	0.24 0.93
normalized false alarm rate	γ_{FA}/V	1.9
(extracted location covariance)$^{1/2}$	$R_k^{1/2}$	$0.88'' \times I$
type confusion matrix	$\{\rho_{ij}\}$	0.86 0.14 0.03 0.97

As described in Section 7.3.3, we assume that we are given an initialization $\Theta^{[0]}$ for the EM iteration. In this experiment, we produce an initialization by perturbing the true target model (described above) as follows. The model order (N_t) of the initialization is the same as that of the true target. The initial guess for the location of each primitive, $\Theta_t^{X[0]}$, is Gaussian distributed with mean Θ_t^X and covariance $[3'' \times I]^2$, such that the error in each direction is uncorrelated, has zero mean, and has a standard deviation of three inches. The initial guess for the type of each primitive, $\Theta_t^{\alpha[0]}$, is equally distributed among the $N_\alpha = 2$ types, i.e. , the initial guess for the type of each primitive has probability 1/2 of being correct. For any primitives with $\Theta_t^{\alpha[0]} = 2$ (tophat), the initial guess for the radius of the primitive, $\Theta_t^{r[0]}$, is zero. The gate radius used for this experiment was 2 m.

Figure 40 is a scatter plot of the estimated x and y locations (estimated z locations not shown) and estimated types of each primitive for 500 trials of this experiment. (For each trial, a new random initialization $\Theta^{[0]}$ and random clutter were generated.) Triangle symbols correspond to an estimated type of trihedral; circles correspond to an estimated type of tophat. We see that in all cases but one, the estimated types of each primitive are correct. Furthermore, all but one of the estimated tophat locations, and all but 11 of the estimated trihedral locations, are clustered around the true locations of these primitives. The clustering of the estimated trihedral locations is looser than that of the tophat; this is to be expected, since the trihedral produces a response over a narrower aspect region than the tophat and hence we have fewer measurements of its position (roughly one-quarter as many, according to Table 5). The elongation of the trihedral cluster corresponds to the specular orientation of the trihedral. The mean and error covariance of the clustered location estimates for each primitive (including the radius estimate for the tophat) are given in Table 6. (All units are in inches; the outlying location estimates have been omitted from this calculation.) The off-diagonal terms for the trihedral square-root error covariance matrix correspond to the elongated trihedral location cluster in Figure 40. The large correlation between the z-location error and the radius error in the tophat square-root error covariance matrix are due to the fact that an increase in radius and a corresponding increase in z-coordinate are projectionally indistinguishable when viewed from a depression angle of 45°, and nearly indistinguishable for depression angles near 45°. (We postulate that this is also responsible at least in part for the bias in the z-location and radius estimates.) Figure 41 is a histogram of the radius estimates for the tophat.

7.4.2 Experiment 2

In this section we examine the performance of the algorithm on the same target model and under circumstances identical to those in the previous experiment, with one difference: the location initialization for each primitive now has covariance $[6'' \times I]^2$. Figure 42 is a scatter plot of the estimated x and y locations and types of each primitive for 500 trials of this experiment. Comparing to Figure 40, we see that the larger variance of the location initialization degrades performance by reducing the fraction of estimates that cluster

Table 6: Experiment 1 statistics

trihedral	
fraction in cluster	97.8%
$\text{mean}\left(\hat{\Theta}_t^X\right)$	$\begin{bmatrix} 29.92 \\ 29.92 \\ 0.04 \end{bmatrix}$
$\left[\text{cov}\left(\hat{\Theta}_t^X\right)\right]^{1/2}$	$\begin{bmatrix} 0.28 & 0.14 & -0.24 \\ 0.14 & 0.28 & -0.25 \\ -0.24 & -0.25 & 0.56 \end{bmatrix}$
tophat	
fraction in cluster	99.8%
$\text{mean}\left(\begin{bmatrix} \hat{\Theta}_t^X \\ \hat{\Theta}_t^r \end{bmatrix}\right)$	$\begin{bmatrix} -34.21 \\ -34.22 \\ 0.23 \\ 5.76 \end{bmatrix}$
$\left[\text{cov}\left(\begin{bmatrix} \hat{\Theta}_t^X \\ \hat{\Theta}_t^r \end{bmatrix}\right)\right]^{1/2}$	$\begin{bmatrix} 0.14 & -0.00 & -0.00 & -0.00 \\ -0.00 & 0.13 & 0.00 & -0.01 \\ -0.00 & 0.00 & 0.53 & -0.38 \\ -0.00 & -0.01 & -0.38 & 0.44 \end{bmatrix}$

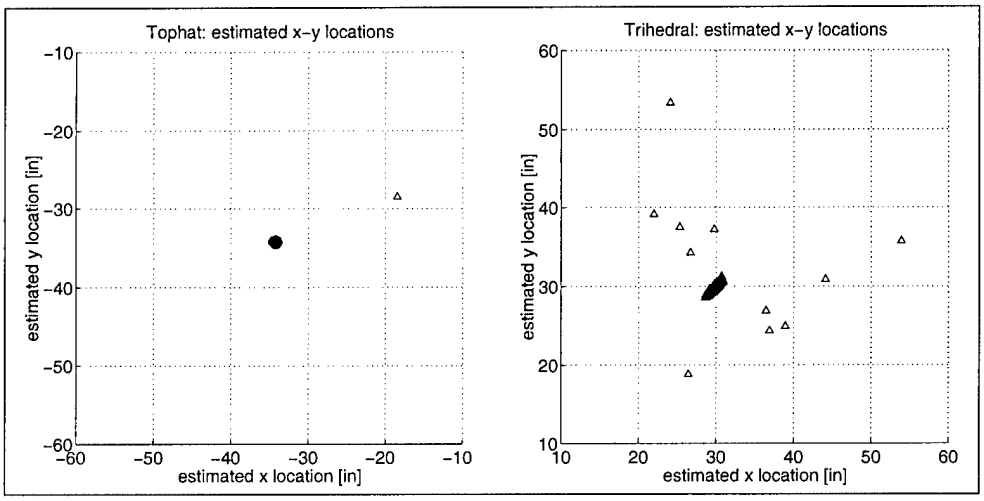


Figure 40: Experiment 1: Estimated locations and types of primitives

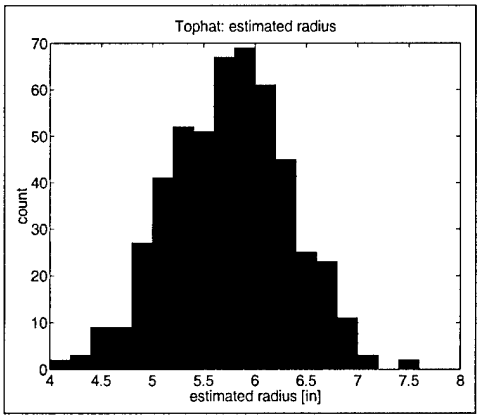


Figure 41: Experiment 1: Estimated radius of tophat

Table 7: Experiment 2 statistics

trihedral	
fraction in cluster	84.0%
$\text{mean}\left(\hat{\Theta}_t^X\right)$	$\begin{bmatrix} 29.94 \\ 29.95 \\ -0.02 \end{bmatrix}$
$\left[\text{cov}\left(\hat{\Theta}_t^X\right)\right]^{1/2}$	$\begin{bmatrix} 0.30 & 0.15 & -0.27 \\ 0.15 & 0.29 & -0.27 \\ -0.27 & -0.27 & 0.59 \end{bmatrix}$
tophat	
fraction in cluster	99.2%
$\text{mean}\left(\begin{bmatrix} \hat{\Theta}_t^X \\ \hat{\Theta}_t^r \end{bmatrix}\right)$	$\begin{bmatrix} -34.21 \\ -34.22 \\ 0.21 \\ 5.77 \end{bmatrix}$
$\left[\text{cov}\left(\begin{bmatrix} \hat{\Theta}_t^X \\ \hat{\Theta}_t^r \end{bmatrix}\right)\right]^{1/2}$	$\begin{bmatrix} 0.13 & -0.01 & -0.00 & -0.00 \\ -0.01 & 0.14 & 0.00 & -0.01 \\ -0.00 & 0.00 & 0.56 & -0.41 \\ -0.00 & -0.01 & -0.41 & 0.45 \end{bmatrix}$

around the true location. There are now four outliers in the tophat location estimate (compared to one in experiment 1) and 80 in the trihedral location estimate (compared to 11 in experiment 1). Estimation of types is not adversely affected: the only misclassified primitives are the outlying tophats. The mean and error covariance of the clustered location estimates for each primitive (including the radius estimate for the tophat) are given in Table 7. (Again, units are in inches, and the outlying estimates have been omitted from the calculation.) Comparing Tables 6 and 7, we see that although the larger location initialization covariance results in fewer location estimates in each cluster, the statistics for those that within the cluster are nearly identical. This suggests that as long as the initialization is "sufficiently good," the resulting error in the location estimate will not depend strongly on the precise initialization location. If the initialization is poor, however, then convergence toward the true location is unlikely. Figure 43 is a histogram of the radius estimates for the tophat.

7.5 Summary and Future Work

We have presented an iterative algorithm for producing a 3-D target model from a collection of SAR images. This model describes the target in terms of a collection of reflector primitives. We cast the estimation problem as a data association problem. We made several simplifying assumptions to decouple the underlying target model estimation problem into smaller reflector primitive estimation problems, and to model the report-to-

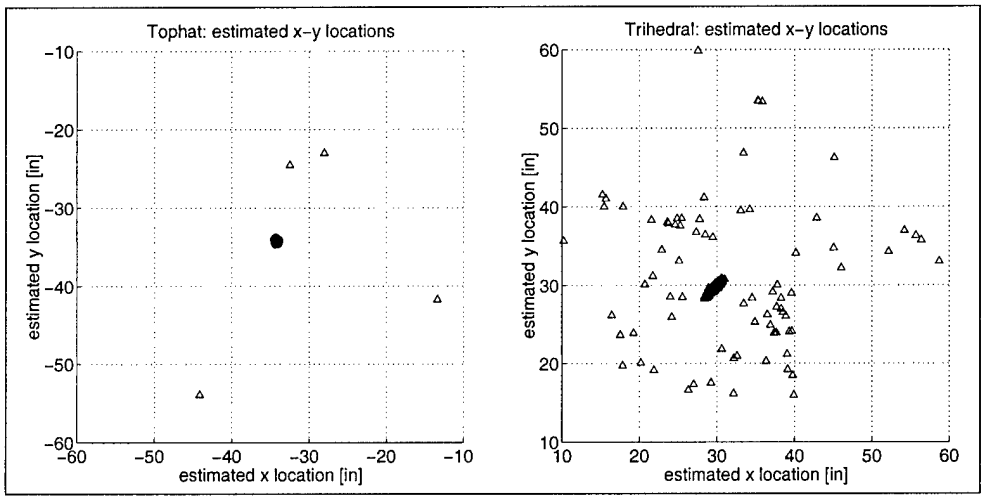


Figure 42: Experiment 2: Estimated locations and types of primitives

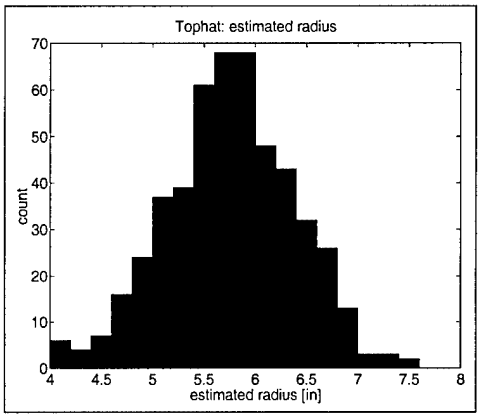


Figure 43: Experiment 2: Estimated radius of tophat

primitive associations as independent between frames. We utilized the EM method to produce a target model specifying primitive locations and types (and radii for curved primitives with axis of symmetry perpendicular to the ground plane) by considering all feasible 2-D scattering center extraction associations across images. We presented experimental results for the application of our algorithm to a simple target.

Direct extensions to this research we hope to undertake in the near future include investigation into initialization procedures for the model generation iteration, model order estimation, and the incorporation of other physically relevant features (such as amplitude information) to enhance the current algorithm and to enable estimation of additional target information such as primitive pose and size. We will eventually relax some of the assumptions utilized here in favor of less restrictive, more physically faithful ones enabling us to capture the spatial scattering dependence of primitives and to accommodate non-discrete primitive types. Other, more fundamental issues we wish to examine as research progresses involve the theoretical limits of model estimation accuracy given a specific choice of reflector primitive parameterization, and how observability and robustness issues might influence the model estimation process.

7.6 Derivation of $p\left(\lambda(k), Z(k); \Theta\right)$

We seek an expression for $p\left(\lambda(k), Z(k); \Theta\right)$ based on the formulation and assumptions of Section 7.2. We can write

$$p\left(\lambda(k), Z(k); \Theta\right) = p(X(k)|\lambda(k); \Theta) \cdot p(\alpha(k)|\lambda(k); \Theta) \cdot p(\lambda(k)|\delta(k), F(k); \Theta) \cdot p(\delta(k), F(k); \Theta) \tag{78}$$

because $X(k)$ and $\alpha(k)$ were assumed conditionally independent given Θ and because $\delta(k)$ and $F(k)$ are functions of $\lambda(k)$. We will now specify $p\left(\lambda(k), Z(k); \Theta\right)$ by specifying models for the components of the preceding equation.

We have already specified the model for $p(X(k)|\lambda(k); \Theta)$ in the case when reports correspond to target primitives (i.e. , are not false alarms) by (62) and (89). We have also assumed that false alarms are uniformly distributed across the sensor volume V. Thus we may write

$$p(X(k)|\lambda(k); \Theta) = \left(\frac{1}{V}\right)^{F(k)} \prod_{\lambda_m(k) \neq 0} \frac{\exp\left[-\frac{1}{2}\left(f\left(\Theta_{\lambda_m(k)}, \psi_k, \phi_k\right) - X_m(k)\right)^T R_k^{-1}\left(f\left(\Theta_{\lambda_m(k)}, \psi_k, \phi_k\right) - X_m(k)\right)\right]}{2\pi\left(\det R_k\right)^{1/2}}. \tag{79}$$

We have also specified the model for $p(\alpha(k)|\lambda(k); \Theta)$ when reports correspond to target primitives by the definition of the confusion term $\rho_{\alpha_1 \alpha_2}$. Furthermore, type parameter estimates of spurious reports were assumed to be uniformly distributed across the N_α allowable types. This allows us to write

$$p(\alpha(k)|\lambda(k); \Theta) = \left(\frac{1}{N_\alpha}\right)^{F(k)} \prod_{\lambda_m(k) \neq 0} \rho'_{m(k)}. \tag{80}$$

We now must describe $p(\lambda(k)|\delta(k), F(k); \Theta)$ and $p(\delta(k), F(k); \Theta)$. Describing the first of these is very simple. Given $F(k)$ and $\delta(k)$, we can compute $D(k)$ and $M(k)$. Furthermore, $\lambda(k)$ contains exactly $F(k)$ zero entries, and the remaining $D(k)$ entries are some permutation of the positions of the nonzero entries of $\delta(k)$. Because we have assumed that there is no preferential or systematic way of ordering the $\lambda_m(k)$ within $\lambda(k)$ for each frame, each permutation is equally likely. Thus we may write

$$p(\lambda(k)|\delta(k), F(k); \Theta) = \left[\frac{(F(k) + D(k))!}{F(k)!D(k)!} D(k)!\right]^{-1} = \frac{F(k)!}{M(k)!}. \tag{81}$$

We have assumed that each detection or false alarm is conditionally independent from any other detection or false alarm given Θ. Recall also the assumptions that $F(k)$ is determined by a Poisson process with mean $\gamma_{FA}V$, and that the probability of detection for a primitive of type α is $P_D(\alpha)$. All these assumptions allow us to write

$$p(\delta(k), F(k); \Theta) = \frac{e^{-\gamma_{FA}V}(\gamma_{FA}V)^{F(k)}}{F(k)!} \cdot \prod_{t=1}^{N_t}(1 - P_D(\Theta_t^\alpha)) \cdot \prod_{\lambda_m(k)\neq 0} \frac{P_D(\Theta^\alpha_{\lambda_m(k)})}{1 - P_D(\Theta^\alpha_{\lambda_m(k)})} \tag{82}$$

where $\delta(k)$ is implicitly restricted to be a N_t-vector with binary entries and $F(k)$ to be a non-negative integer.

We can now combine the preceding equations according to (78) to yield a single expression for $p(\lambda(k), Z(k); \Theta)$. After cancelling some terms and rearranging others, this yields the expression given in (95).

7.7 Derivation of $Q(\Theta; \Theta^{[n]})$

We seek an expression for $Q(\Theta; \Theta^{[n]})$ in order to describe the E step of the EM algorithm as presented in Section 7.3.1. Utilizing the independence of report parameters between frames and exchanging the order of summation in (95), we can write

$$Q(\Theta; \Theta^{[n]}) = \sum_{k=1}^{K}\left[\sum_{\lambda^K \in \Lambda^K} \log p(\lambda(k), Z(k); \Theta)\right] \cdot p\left(\lambda(k) \,|Z(k); \Theta^{[n]}\right) \cdot p\left(\lambda^K \setminus \lambda(k) \,\middle|\, Z^K \setminus Z(k); \Theta^{[n]}\right) \tag{83}$$

where the notation "$\lambda^K \setminus \lambda(k)$" is shorthand for "$\lambda(1), \ldots, \lambda(k-1), \lambda(k+1), \ldots, \lambda(K)$" (an analogous definition applies to "$Z^K \setminus Z(k)$"). It then follows that

$$Q(\Theta; \Theta^{[n]}) = \sum_{k=1}^{K}\left[\sum_{\lambda(k)\in\Lambda(k)} \left[\log p(\lambda(k), Z(k); \Theta)\right] \cdot p\left(\lambda(k) \,|Z(k); \Theta^{[n]}\right)\right]. \tag{84}$$

Replacing the symbolic term $\log[p(\lambda(k), Z(k); \Theta)]$ in the above expression with its value according to (95), we obtain

$$\begin{aligned} Q(\Theta; \Theta^{[n]}) \;=\; & \sum_{k=1}^{K}\sum_{\lambda(k)\in\Lambda(k)} \left\{-\gamma_{FA}V + F(k)\log\left(\frac{\gamma_{FA}}{N_\alpha}\right) - \log(M(k)!) + \sum_{t=1}^{N_t}\log(1 - P_D(\Theta_t^\alpha)) + \right. \\ & \sum_{\lambda_m(k)\neq 0}\left[-\frac{1}{2}\left(f(\Theta_{\lambda_m(k)}, \psi_k, \phi_k) - X_m(k)\right)^T R_k^{-1}\left(f(\Theta_{\lambda_m(k)}, \psi_k, \phi_k) - X_m(k)\right) + \right. \\ & \left.\left. \log\frac{P_D(\Theta^\alpha_{\lambda_m(k)})}{1 - P_D(\Theta^\alpha_{\lambda_m(k)})} - \log 2\pi(\det R_k)^{1/2} + \log\rho'_{m(k)}\right]\right\} \cdot p\left(\lambda(k) \,|Z(k); \Theta^{[n]}\right) \end{aligned} \tag{85}$$

Now, to simplify this expression, let us define an indicator function for each $t = 0, \ldots, N_t$:

$$\chi_{\{\lambda_m(k)=t\}} = \begin{cases} 1, & \text{if } \lambda_m(k) = t \\ 0, & \text{otherwise.} \end{cases} \tag{86}$$

Then note that for any expression $\zeta_m(k)$, we have

$$\sum_{\lambda(k)\in\Lambda(k)}\sum_{\lambda_m(k)\neq 0}\zeta_m(k)\cdot p\left(\lambda(k)\left|Z(k);\Theta^{[n]}\right.\right) = \sum_{\lambda(k)\in\Lambda(k)}\sum_{m=1}^{M(k)}\left[\sum_{t=1}^{N_t}\chi_{\{\lambda_m(k)=t\}}\right]\cdot\zeta_m(k)\cdot p\left(\lambda(k)\left|Z(k);\Theta^{[n]}\right.\right)$$
$$= \sum_{m=1}^{M(k)}\sum_{t=1}^{N_t}\zeta_m(k)|_{\lambda_m(k)=t}\cdot\Pr\left(\lambda_m(k)=t\left|Z(k);\Theta^{[n]}\right.\right) \quad (87)$$

which allows us to rewrite (85) as

$$Q\left(\Theta;\Theta^{[n]}\right) = \sum_{t=1}^{N_t}\left\{\sum_{k=1}^{K}\left[\log\left(1-P_D\left(\Theta_t^{\alpha}\right)\right)+\sum_{m=1}^{M(k)}\left(\log\left(P_D\left(\Theta_t^{\alpha}\right)\right)-\log\left(1-P_D\left(\Theta_t^{\alpha}\right)\right)+\log\rho'_{m(k)}-\right.\right.$$
$$\left.\left.\frac{1}{2}\left(f\left(\Theta_t,\psi_k,\phi_k\right)-X_m(k)\right)^T R_k^{-1}\left(f\left(\Theta_t,\psi_k,\phi_k\right)-X_m(k)\right)\right)\Pr\left(\lambda_m(k)=t\left|Z(k);\Theta^{[n]}\right.\right)\right]\right\}+$$
$$\sum_{k=1}^{K}\left[-\gamma_{FA}V-\log(M(k)!)+\log\left(\frac{\gamma_{FA}}{N_\alpha}\right)\cdot\sum_{m=1}^{M(k)}\Pr\left(\lambda_m(k)=0\left|Z(k);\Theta^{[n]}\right.\right)-\right.$$
$$\left.\sum_{m=1}^{M(k)}\sum_{t=1}^{N_t}\log\left(2\pi\left(\det R_k\right)^{1/2}\right)\cdot\Pr\left(\lambda_m(k)=t\left|Z(k);\Theta^{[n]}\right.\right)\right]. \quad (88)$$

Defining $Q_t\left(\Theta_t;\Theta^{[n]}\right)$ to be the argument of the "$t=1$ to N_t" summation (enclosed in the curly braces in the first two lines of (88)) and c_K to be the remaining portion of (88) (which is constant with respect to Θ), we arrive at the expression for $Q\left(\Theta;\Theta^{[n]}\right)$ given in (97).

8 An Expectation-Maximization Approach to Target Model Generation from Multiple SAR Images

A key issue in the development and deployment of model-based automatic target recognition (ATR) systems is the generation of target models to populate the ATR database. Model generation is typically a formidable task, often requiring detailed descriptions of targets in the form of blueprints or CAD models. Recently, efforts to generate models from a single 1-D radar range profile or a single 2-D synthetic aperture radar (SAR) image have met with some success. However, the models generated from these data sets are of limited use to most ATR systems because they are not three-dimensional. We propose a method for generating a 3-D target model directly from multiple SAR images of a target obtained at arbitrary viewing angles. This 3-D model is a parameterized description of the target in terms of its component reflector primitives. We pose the model generation problem as a parametric estimation problem based on information extracted from the SAR images. We accomplish this parametric estimation in the context of data association using the expectation-maximization (EM) method. Although we develop our method in the context of a specific data extraction technique and target parameterization scheme, our underlying framework is general enough to accommodate different choices. We present results demonstrating the utility of our method.

8.1 Introduction

In recent years there has been a surge of interest in model-based automatic target recognition (ATR) algorithms for use with synthetic aperture radar (SAR) imaging systems. The broad utility of SAR as an imaging methodology is well-known, and SAR imaging techniques and systems have been extensively documented [54, 46]. The effectiveness of SAR as a remote sensing tool has motivated research into the development of model-based ATR systems [55, 56]. Model-based ATR systems identify targets by comparing image features to classification hypotheses generated from a database of physical target models. The generation of target models to populate this database is a problem that is central to the implementation of any model-based ATR system [55].

In this paper we present a framework for producing a three-dimensional target model from multiple SAR images of a target. Our models consist of spatial collections of reflector primitives such as cylinders, tophats, dihedrals, and trihedrals [56, 47]. Reflector primitive models offer compact representations of many targets, highlighting and parameterizing observable features in terms of information of direct use to ATR, including primitive locations, types, poses, sizes, and possibly other information relevant to describing the scattering signatures of man-made targets. Reflector primitives couple physical relevance to predictive utility in ATR, facilitating the model manipulation and component articulation required to form classification hypotheses. Our reflector primitive models describe each component primitive with a small number of parameters, namely, a discrete index identifying the scatterer as one of a small number of canonical scattering types, and several continuous parameters including location and pose, completely describing the scattering behavior of that type of primitive.

Our framework entails estimation of the number of scatterers and their descriptive parameters based on the observed set of SAR images. In principle the optimal way to do this is to use all of the available imagery to perform the parameter estimation directly. Note that the explicit inclusion of location as one of the parameters describing each primitive implies that the model estimation procedure must deal with establishing a *correspondence* between each postulated primitive and the observed scattering responses in all of the SAR images. In principle the optimal way to do *this* is to use all of the SAR images directly to establish these correspondences at the same time that the parameters of each primitive are estimated.

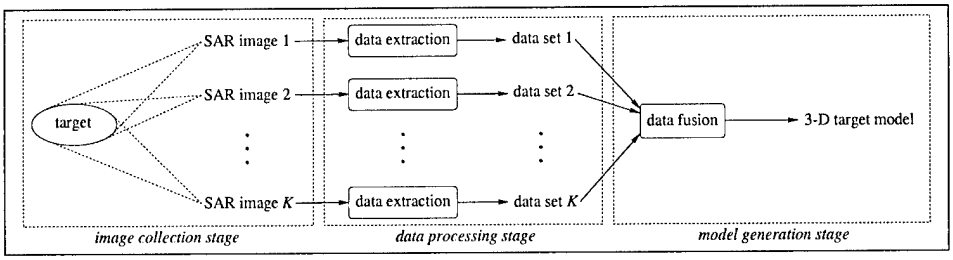


Figure 44: Target model generation block diagram

However, because of the complexity of such a task, the fact that our ultimate objective is a low-dimensional description of the target as a set of primitives, and the fact that model-based ATR systems already operate in this manner, we propose to view the estimation problem as a two-step procedure. Each SAR image is first compressed into a set of augmented detections consisting of relevant information about significant scattering responses in each image, including location and other data extracted from the individual images or phase histories. These compressed representations are then fused in order to estimate the 3-D locations and characteristics of the target primitives. This framework offers great flexibility in the choice of a compression scheme, with possibilities ranging from fine-grained extractions in which the compression of each SAR image involves keeping a great many basis functions that capture most of the energy in the raw image, to more coarse-grained representations in which only a small number of dominant scatterers are kept from each image, with only a few parameters describing each response. In order to introduce our framework and to highlight representations commonly used in ATR, we focus here on a parameterization at the coarser end of this spectrum. This choice also highlights the importance of the correspondence problem mentioned previously.

Much of this paper represents a continuation of work first presented at last year's conference [57]. The research reported here represents a significant expansion of this earlier work. In particular, the model generation algorithm described here estimates not only primitive location and type, but also primitive pose and amplitude; another significant advance is the inclusion of unsupervised initialization and model order estimation stages. Our new framework also incorporates a broader set of scattering primitives than previously considered. In the next section we present our formulation of the target model estimation problem, and in Section 8.3 we describe our application of the expectation-maximization (EM) method to its solution. In Section 8.4 we present experimental results illustrating the performance of our algorithm. Section 8.5 concludes the paper.

8.2 Problem Formulation

A block diagram representation of our approach to 3-D target model estimation is depicted in Figure 44. A target is observed through a set of K SAR images. Each of these images corresponds to a particular viewing geometry, as illustrated in Figure 45: each image k is characterized by a line-of-sight vector from the center of the synthetic aperture to the center of the target region being imaged. The azimuth ϕ_k and elevation ψ_k defining this line-of-sight vector in terms of a fixed ground frame of reference are arbitrary; we assume each image has been formed at a squint angle of 90°. (Extension of our approach to allow arbitrary squint angles is straightforward.) The synthetic aperture along the platform motion vector and the line-of-sight vector define the *slant plane*, the imaging plane for the SAR image [54].

As indicated in Figure 44, each of the K SAR images is processed to extract a set of observed features

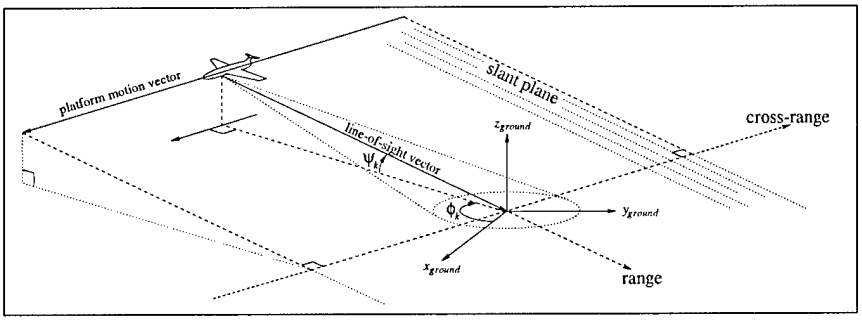


Figure 45: Imaging geometry and the slant plane for image k

which are then fused in order to produce a 3-D target model; the feature extraction considered in this paper is described in Section 8.2.2. Our fundamental focus is the design of the data fusion module in Figure 44. This requires specification of exactly what we wish to estimate (i.e. , the parameterization of our target models) and how the features serving as input to this module are related to the quantities to be estimated. The latter step involves modeling both the SAR image collection process and the subsequent data processing that produces the observables on which the fusion module will operate. We first describe the notation and basic assumptions defining the problem and then present the measurement model that relates observables and target parameters.

8.2.1 Target Models: Assumptions and Notation

Our target models consist of collections of reflector primitives, each of which is described by a small set of parameters that completely specify the scattering behavior of such a primitive given any imaging geometry. As we indicated previously, in this paper we will restrict attention to a comparatively constrained set of primitives, each of which can be completely described for our purposes by a short vector of parameters. In particular, a target model will be specified by the number of primitives N comprising the target and a vector of parameters $\boldsymbol{\theta}_i$ associated with each component primitive $i = 1, \ldots, N$. In general, we can express this vector as $\boldsymbol{\theta}_i = [\theta_i^t, \theta_i^X, \boldsymbol{\theta}_i^s]$, where θ_i^t is an integer index designating the primitive as one of n_t canonical primitive types, θ_i^X is the 3-D location of the primitive, and $\boldsymbol{\theta}_i^s$ is a generic vector parameter corresponding to a set of continuous-valued parameters that, along with θ_i^t and θ_i^X, completely specify the log-amplitude scattering response or *radar cross section* (RCS) of the primitive from any viewing angle [47]. We will denote the log-RCS observed from a primitive parameterized by $\boldsymbol{\theta}_i$ and viewed from elevation ψ and azimuth ϕ as $A(\boldsymbol{\theta}_i, \psi, \phi)$, which we typically quote in dBsm.

For this paper, we will constrain the set of scattering types to a small class of idealized primitives consisting of trihedrals, tophats, dihedrals, and cylinders (so that $n_t = 4$), depicted in Figure 46; we assign type indices 1 through 4 to these primitives, respectively. For these primitives, $\boldsymbol{\theta}_i^s$ will consist of at most three parameters: an overall base amplitude θ_i^a related to the physical size of the scatterer, a pose θ_i^p indicating the orientation of the scatterer, and a radius of curvature θ_i^r for radially symmetric primitives including tophats and cylinders. Each primitive's location θ_i^X is defined to correspond to the origin of the primitive's local axes in Figure 46; primitive pose indicates the orientation of these axes with respect to the fixed ground-based coordinate system in terms of three Euler angles [58]. Primitive pose and the absolute viewing angle of

image k together define a relative viewing elevation $\psi'_{i,k}$ and azimuth $\phi'_{i,k}$ for each primitive, as depicted in Figure 46.

The complete vector $\boldsymbol{\theta}_i$ provides a concise yet accurate description of a primitive's appearance in an arbitrary SAR image. Location θ_i^X and radius of curvature θ_i^r (for those primitives for which it is defined), along with viewing angle, determine the apparent location of the primitive in the slant plane [54]. In particular, as described elsewhere[57], we can model the location of primitive i in image k as a projection of θ_i^X into the slant plane, with an uprange offset for radially symmetric curved primitives:

$$\boldsymbol{\pi}_k(\boldsymbol{\theta}_i) = \begin{cases} H_k \boldsymbol{\theta}_i^X, & \theta_i^t \in T_{\text{fixed}}, \\ H_k \boldsymbol{\theta}_i^X - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \theta_i^r \cos\psi'_{i,k}, & \theta_i^t \in T_{\text{curved}}, \end{cases} \tag{89}$$

where $\psi'_{i,k}$ is as pictured in Figure 46, where T_{curved} is the set of type indices for radially symmetric primitives (i.e. , tophats and cylinders) and T_{fixed} is the set of the others (i.e. , dihedrals and trihedrals), and where H_k is the 2×3 ground-to-slant-plane transformation matrix for image k [57].[11] The other components of $\boldsymbol{\theta}_i$ determine other features of the observed response: discrete type θ_i^t specifies the basic dependence of the response on viewing angle [47] (and, if polarimetric measurements are made, the polarimetric signature vector [59]); pose θ_i^p orients this response by rotating it to correspond to the orientation of the primitive; base amplitude θ_i^a scales the response intensity according to the physical size of the primitive. In particular, physical optics provides expressions for the RCS of each primitive as the product of a size-dependent amplitude term and a unique type-dependent shaping function capturing the dependence of RCS on relative viewing angle and possibly size [47]. Our log-RCS models are based on these physical optics results and take the form

$$A(\boldsymbol{\theta}_i, \psi_k, \phi_k) = \theta_i^a + S_{\theta_i^t}(\psi'_{i,k}, \phi'_{i,k}), \tag{90}$$

where $S_{\theta_i^t}(\bullet)$ is the physical optics log-shaping function describing the variation in scattering response in terms of viewing angle for all primitives of type θ_i^t, and where θ_i^a encapsulates the fundamental size dependence described by physical optics[47]. For each primitive, $S_{\theta_i^t}(\psi'_{i,k}, \phi'_{i,k})$ is scaled to give a maximum response of 0 dBsm, so that θ_i^a will correspond to the maximum amplitude of the primitive response. A difficulty in the specification of the $S_{\theta_i^t}(\bullet)$ as in (90) is the fact that each primitive's physical optics shaping function depends on its dimensions. This dependence is most pronounced for the dihedral and cylinder, which exhibit sinc-like elevation responses depending on b and h, respectively. Additionally, the trihedral, dihedral, and tophat responses, each of which comprises both single- and multiple-bounce response mechanisms, depend on primitive dimensions to determine the relative phase between each component response. Although $\boldsymbol{\theta}_i^s$ could be enhanced to include primitive dimension and the shaping functions of (90) expanded to model these dependences, we maintain our chosen parameterization in the interest of presenting a basic framework with which to demonstrate model construction, and which can be broadened as necessary. Our models for the dihedral and cylinder shaping function models are constructed using empirically chosen nominal values for b and h, respectively, and our trihedral, dihedral, and tophat models combine the highest- and lower-bounce response mechanisms via a noncoherent sum without regard to the size-dependent relative phase.

[11] The model of (89), while accurate for most primitives viewed from most angles, will be inaccurate for trihedrals and dihedrals viewed at angles at which the lower-bounce reflection mechanisms dominate and the apparent specular reflection point does not correspond to θ_i^X (e.g. , when a single- or double-bounce response is observed from a trihedral).

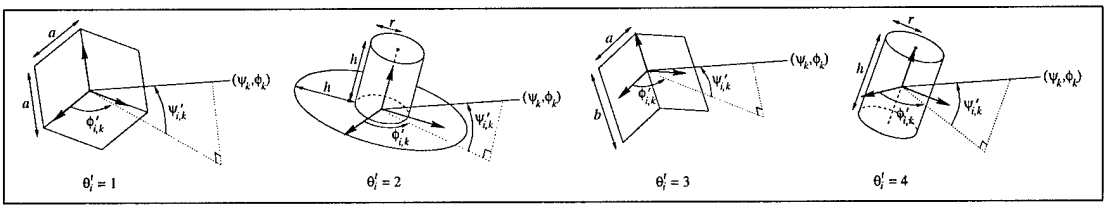


Figure 46: Reflector primitives: trihedral, tophat, dihedral, and cylinder. Relative elevation $\psi'_{i,k}$ and azimuth $\phi'_{i,k}$ are determined by the absolute viewing elevation ψ_k and azimuth ϕ_k and the pose of primitive, indicated by the orientation of its local axes; primitive dimensions relevant to physical optics RCS models are indicated.

Before proceeding, we introduce notation $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N]$. (Note that $\boldsymbol{\theta}$ implicitly specifies the model order N.) Target model generation in our framework is thus estimation of the vector $\boldsymbol{\theta}$. We will model $\boldsymbol{\theta}$ (and thus N) as unknown parameters about which no information is available other than that provided by the SAR images.

8.2.2 Observable Features for Model Generation

We assume that we have multiple spotlight-mode SAR images [54, 46] of the target, formed at arbitrary viewing angles as depicted in Figure 45. Each of these images is polarimetric, comprising linear polarizations HH, HV, and VV. Furthermore, we assume that all SAR imaging parameters (such as bandwidth, aperture width, range and cross-range locations of each pixel center, and azimuth and depression to the target center) are known, and can be related to the absolute ground-based frame of reference. Such information could be provided, for instance, by geolocation or global positioning measurements taken as the images are collected, coupled with accurate ranging and positioning of the target.

As previously described in the introduction and in conjunction with Figure 44, we compress the full set of raw SAR imagery by extracting information from each image prior to the model generation stage. For this purpose we utilize a simple peak-extraction routine, similar to that of the MSTAR "extract" module [44] and described in detail in a previous work[57]. This processor extracts an arbitrary number of intensity peaks M_k from each image k, and describes each peak $j = 1, \ldots, M_k$ in terms of three parameters: a 2-D slant-plane range/cross-range location $\mathbf{X}_{k,j}$, a discrete polarimetric-signature type index $t_{k,j}$, and a scalar log-amplitude $a_{k,j}$. Location and amplitude are obtained using a simple subpixel-interpolation procedure, and polarimetric signature type is obtained via a generalized likelihood ratio test to distinguish between odd-bounce and even-bounce responses[59]. The extracted type is thus a binary variable; we designate an odd-bounce classification as $t_{k,j} = 1$ and an even-bounce classification as $t_{k,j} = 2$. Note that because trihedrals and cylinders are predominantly odd-bounce scatterers, and dihedrals and tophats predominantly even-bounce scatterers, discrimination between trihedrals and cylinders, or between dihedrals and tophats, will be based predominantly on location and amplitude information. The effect of the indistinguishable type measurements for these primitive classes is investigated in Section 8.4.

For convenient reference, the three-parameter location/amplitude/type description of the jth peak of image k will be called a *report* and denoted by $\mathbf{Z}_{k,j}$. At times it will be convenient to refer to the collection of reports within a single image or across images. For these purposes we define notation for all reports in a single image, $\mathbf{Z}_k = [\mathbf{Z}_{k,1}, \ldots, \mathbf{Z}_{k,M_k}]$, and notation for all reports in all images, $\mathbf{Z} = [\mathbf{Z}_1, \ldots, \mathbf{Z}_K]$.

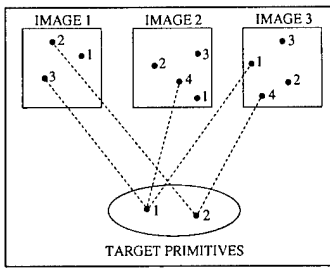


IMAGE 1	*IMAGE 2*	*IMAGE 3*
$M_1 = 3$	$M_2 = 4$	$M_3 = 4$
$\lambda_{1,1} = 0$ $\lambda_{1,2} = 2$ $\lambda_{1,3} = 1$	$\lambda_{2,1} = 0$ $\lambda_{2,2} = 0$ $\lambda_{2,3} = 0$ $\lambda_{2,4} = 1$	$\lambda_{3,1} = 1$ $\lambda_{3,2} = 0$ $\lambda_{3,3} = 0$ $\lambda_{3,4} = 2$
$F_1 = 1$	$F_2 = 3$	$F_3 = 2$

Figure 47: Notation example

8.2.3 Measurement Model

In this section we describe the probabilistic model relating features extracted by the data processor to the target parameters that must be estimated from those features. The uncertainties in the extracted features come at two levels of granularity, one coarse and one fine. The coarse-level uncertainty involves the identity of each measurement: given a set of reports extracted from a single SAR image and a set of target primitives, there is no way of knowing with certainty which reports correspond to which primitives. The fine-level uncertainty involves the stochastic nature of the elements of $\mathbf{Z}_{k,j}$, even given the report's proper correspondence. Compounding the coarse-level uncertainty is the fact that, like any detector, the data processor is subject to missed detections and false alarms, so in general there will not be exhaustive correspondence between the sets of reports and target primitives. To formalize the coarse-level uncertainty we will introduce a vector of hidden parameters $\boldsymbol{\lambda}$ that describes the correspondences between reports and target primitives in concrete terms. In particular, we define a label parameter describing the identity of each report $\mathbf{Z}_{k,j}$ as follows:

$$\lambda_{k,j} = \begin{cases} i, & \text{if report } \mathbf{Z}_{k,j} \text{ corresponds to target primitive } i, \\ 0, & \text{if report } \mathbf{Z}_{k,j} \text{ is spurious (corresponds to no primitive).} \end{cases}$$

We also define F_k to be the number of false alarms in image k, i.e. , the number of $\lambda_{k,j}$ which equal 0 for a given k. Figure 47 presents an illustrative example of the notation and concepts encapsulated in $\lambda_{k,j}$. This figure depicts a scenario involving two target primitives ($N = 2$) and three images ($K = 3$).

It will be convenient to define a vector $\boldsymbol{\lambda}_k$ collecting the label parameters for all of the reports in image k: $\boldsymbol{\lambda}_k = [\lambda_{k,1}, \cdots, \lambda_{k,M_k}]$. The vector $\boldsymbol{\lambda}$ introduced above can be formally defined as $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_K]$. Given knowledge of $\boldsymbol{\lambda}$, the uncertainty remaining in $\mathbf{Z}$ is the distribution of the components of each report $\mathbf{Z}_{k,j}$; this is our fine-level uncertainty. Characterization of the fine-level uncertainty can be done conditionally, and the measurement model can be specified as

$$p(\boldsymbol{\lambda}, \mathbf{Z}|\boldsymbol{\theta}) = p(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\theta})\, p(\boldsymbol{\lambda}|\boldsymbol{\theta}), \tag{91}$$

a product of the fine-level probability density function (pdf) and the coarse-level probability mass function (pmf).[12]

We make a number of general assumptions about the relationship of $\boldsymbol{\lambda}$ and $\mathbf{Z}$ to $\boldsymbol{\theta}$ that will facilitate the specification of a measurement model. The first three of these concern the coarse-level uncertainty expressed

[12]Throughout this paper we will describe discrete random variables and vectors such as $\boldsymbol{\lambda}$ by their pmfs, and continuous random variables and vectors such as $\mathbf{Z}$ by their pdfs, using the same notation $p(\bullet)$ in both cases.

by $p(\boldsymbol{\lambda}|\boldsymbol{\theta})$; the remaining two concern the remaining fine-scale uncertainty expressed by $p(\mathbf{Z}|\boldsymbol{\lambda},\boldsymbol{\theta})$. All five of these assumptions are largely justifiable on simple physical grounds, and are standard in a variety of data association contexts [48]. These assumptions are as follows:

Assumption 1 *False alarms are independent from image to image and do not depend on* $\boldsymbol{\theta}$.

Assumption 2 *The detectability of the ith primitive in any image depends only on* $\boldsymbol{\theta}_i$ *and on the viewing angle of the image; furthermore, missed detections are conditionally independent from image to image and from report to report given* $\boldsymbol{\theta}$ *and are also independent of false alarms.*

Assumption 3 *Any primitive generates at most one report in each image, and any report is attributable to at most one primitive.*

Assumption 4 *Reports in a single image and between images are conditionally independent given* $\boldsymbol{\theta}$ *and* $\boldsymbol{\lambda}$, *whether they are detections or false alarms.*

Assumption 5 *The component measurements* $\mathbf{X}_{k,j}$, $a_{k,j}$, *and* $t_{k,j}$ *comprising each report are conditionally independent given* $\boldsymbol{\theta}$ *and* $\boldsymbol{\lambda}$, *whether the report is a detection or a false alarm.*

Together, Assumptions 1, 2, and 3 imply the conditional independence of the label parameter vectors for each image:

$$p(\boldsymbol{\lambda}|\boldsymbol{\theta}) = \prod_{k=1}^{K} p(\boldsymbol{\lambda}_k|\boldsymbol{\theta}). \tag{92}$$

Similarly, Assumptions 4 and 5 imply that $p(\mathbf{Z}|\boldsymbol{\lambda},\boldsymbol{\theta})$ can be factored as

$$p(\mathbf{Z}|\boldsymbol{\lambda},\boldsymbol{\theta}) = \prod_{k=1}^{K} p(\mathbf{Z}_k|\boldsymbol{\lambda}_k,\boldsymbol{\theta}) = \prod_{k=1}^{K} \left(\prod_{j=1}^{M_k} p(\mathbf{X}_{k,j}|\lambda_{k,j},\boldsymbol{\theta}) p(a_{k,j}|\lambda_{k,j},\boldsymbol{\theta}) p(t_{k,j}|\lambda_{k,j},\boldsymbol{\theta}) \right). \tag{93}$$

Although there are situations in which these assumptions will fail—for instance, obstruction will violate Assumption 2, multiple-bounce reflections will violate Assumption 3, and phenomena that fall outside of the chosen parameterization of Sections 8.2.1 and 8.2.2 could compromise Assumption 4—these assumptions are largely realistic and greatly facilitate the specification of a measurement model, which now requires only specification of the terms on the right-hand sides of (92) and (93).

Specification of $p(\boldsymbol{\lambda}_k|\boldsymbol{\theta})$ as required by (92) is almost completely determined by Assumption 3 and the constraints it imposes on $\boldsymbol{\lambda}_k$: no more than N of its elements may be nonzero, it cannot contain the same nonzero index twice, and so on. We complete $p(\boldsymbol{\lambda}_k|\boldsymbol{\theta})$ by assuming a standard Poisson false-alarm model and defining a probability-of-detection function that depends only on a primitive's amplitude in any image k. In particular, we write $P_{D'_{k,i}} \equiv P_D(A(\boldsymbol{\theta}_i, \psi_k, \phi_k))$, where $P_D(\bullet)$ is a function that we assume is empirically estimated by running the processor on characteristic imagery. It can then be shown [57] that

$$p(\boldsymbol{\lambda}_k|\boldsymbol{\theta}) = \frac{e^{-\gamma_{FA}V}(\gamma_{FA}V)^{F_k}}{M_k!} \cdot \prod_{i=1}^{N} (1 - P_{D'_{k,i}}) \cdot \prod_{j:\lambda_{k,j}\neq 0} \frac{P_{D'_{k,\lambda_{k,j}}}}{1 - P_{D'_{k,\lambda_{k,j}}}}, \tag{94}$$

where V denotes the sensor volume and where γ_{FA} is the false-alarm rate, a parameter that we assume is empirically estimated in the same manner as $P_D(\bullet)$.

Completing the specification of our model now requires only the specification of the densities for $\mathbf{X}_{k,j}$, $a_{k,j}$, and $t_{k,j}$ in (93). We model the continuous location parameter for a report corresponding to primitive i as a Gaussian with mean $\boldsymbol{\pi}_k(\boldsymbol{\theta}_i)$ and some covariance R; we assume that false alarms are uniformly distributed throughout the SAR images. Likewise, we model the amplitude of a report corresponding to the ith primitive in image k as a Gaussian with mean $A(\boldsymbol{\theta}_i, \psi_k, \phi_k)$ and some variance σ_a^2; false alarms have a separate amplitude distribution denoted by $p_{FA}(\bullet)$. (The parameters R, σ_a^2, and $p_{FA}(\bullet)$ can be estimated from characteristic imagery.) Finally, to model the type extraction process, we assume the availability of an $n_t \times 2$ confusion matrix $\{\rho\}$, where $\rho_{i,j}$ is the probability that the data processor classifies a primitive of type i as having polarimetric signature type j, given that the primitive is detected. As with our other assumed parameters, $\{\rho\}$ can be estimated by processing training data. To simplify notation in subsequent expressions, we write $\rho'_{k,j} \equiv p\left(t_{k,j} \,|\lambda_{k,j}, \boldsymbol{\theta}\right) = \rho_{\theta^t_{\lambda_{k,j}}, t_{k,j}}$ for any detection (i.e. , when $\lambda_{k,j} \neq 0$). We assume false alarms are equally likely to be classified as either polarimetric type.

We now have all the required components to specify $p(\mathbf{Z}|\boldsymbol{\lambda}, \boldsymbol{\theta})$ as in (93); this can in turn be combined with $p(\boldsymbol{\lambda}|\boldsymbol{\theta})$ according to (91) to yield a complete measurement model that can be factored into K product terms, one for each image. In particular, $p(\boldsymbol{\lambda}, \mathbf{Z}|\boldsymbol{\theta}) = \prod_{k=1}^{K} p(\boldsymbol{\lambda}_k, \mathbf{Z}_k|\boldsymbol{\theta})$, where

$$p(\boldsymbol{\lambda}_k, \mathbf{Z}_k|\boldsymbol{\theta}) = \frac{e^{-\gamma_{FA}V}\left(\frac{\gamma_{FA}}{2}\right)^{F_k}}{M_k!} \cdot \prod_{i=1}^{N}(1 - P_{D'_{k,i}}) \cdot \prod_{j:\lambda_{k,j}=0} p_{FA}(a_{k,j}) \cdot \prod_{j:\lambda_{k,j}\neq 0} \frac{P_{D'_{k,\lambda_{k,j}}}}{1 - P_{D'_{k,\lambda_{k,j}}}} \cdot \rho'_{k,j} \cdot$$

$$\prod_{j:\lambda_{k,j}\neq 0} \frac{\exp\left(-\frac{1}{2\sigma_a^2}(a_{k,j} - A(\boldsymbol{\theta}_{\lambda_{k,j}}, \psi_k, \phi_k))^2 - \frac{1}{2}\left(\boldsymbol{\pi}_k(\boldsymbol{\theta}_{\lambda_{k,j}}) - \mathbf{X}_{k,j}\right)^T R^{-1}\left(\boldsymbol{\pi}_k(\boldsymbol{\theta}_{\lambda_{k,j}}) - \mathbf{X}_{k,j}\right)\right)}{(2\pi)^{3/2}\sigma_a(\det R)^{1/2}}.$$

8.3 A Data Association Approach to Model Generation

The measurement model specified in Section 8.2.3 relies on the introduction of a vector of unobservable label parameters $\boldsymbol{\lambda}$ describing the origin of each report. This vector provides not only a convenient device for the specification of a measurement model, but also a conceptual foothold for the estimation of the target parameters. Specifically, if these label parameters were observable—if report data could be associated across images—estimation of $\boldsymbol{\theta}$ would be straightforward. This suggests approaching model generation by way of the underlying data association problem. There is a large body of literature describing theory and methods for solving data association problems in various contexts and applications [48, 50]. The chief difficulty facing almost all data association problems, including the one described here, is the combinatorial proliferation of possible correspondences. One way to manage the combinatorial explosion of possibilities is to dismiss as infeasible a majority of associations corresponding to extremely unlikely events; we will utilize a technique known as gating, to be described later, for this purpose [48]. Even with such a simplification, however, the remaining data association problem is still formidable and requires a powerful tool for solution. The tool we apply is the expectation-maximization (EM) method [50, 51]. In the following section we briefly describe the EM method, and in subsequent sections describe its application to the problem of model generation in the framework we have constructed.

8.3.1 The Expectation-Maximization Method

The EM method is an iterative procedure for producing a maximum likelihood (ML) estimate of parameters when there is a many-to-one mapping from a postulated set of "complete" data to the set of observed data

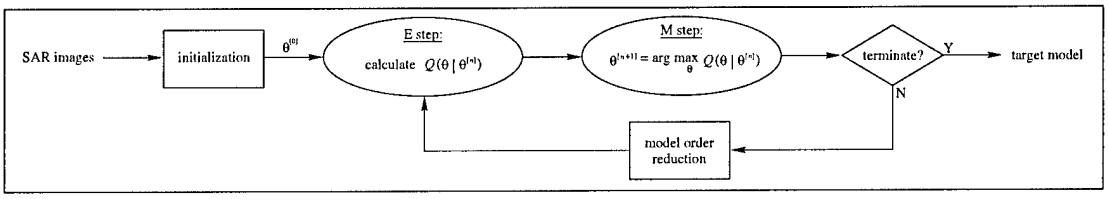


Figure 48: Expectation-maximization (EM) method block diagram

[50]. In data association problems, the set of complete data comprises the observed data and the vector of associations—$\mathbf{Z}$ and $\boldsymbol{\lambda}$ in our context. Each iteration of the basic EM method consists of two steps: an expectation (E) step and a maximization (M) step. The E step averages the log-likelihood of the complete data over all feasible association vectors given the observed data and the latest parameter estimate iterate. The result is an expected log-likelihood that is a function of the true parameter vector $\boldsymbol{\theta}$. The M step then maximizes this expected log-likelihood with respect to the parameter vector. This yields an estimate of $\boldsymbol{\theta}$ for the current iteration that may be used to recompute the expected log-likelihood in the next iteration's E step. Under relatively mild conditions, the EM method is guaranteed to converge to at least a local maximum of the likelihood function of the observed data [51, 52].

In our context, the EM method proceeds as follows. Let $\boldsymbol{\theta}^{[n-1]}$ be the estimate of $\boldsymbol{\theta}$ produced by the M step in iteration $n-1$. The E step of the nth iteration requires calculation of the expected log-likelihood

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{[n-1]}) = E\left[\log p(\boldsymbol{\lambda}, \mathbf{Z}|\boldsymbol{\theta}) \,\middle|\, \mathbf{Z}, \boldsymbol{\theta}^{[n-1]}\right] = \sum_{\boldsymbol{\lambda}\in\Lambda}\Big[\log p(\boldsymbol{\lambda}, \mathbf{Z}|\boldsymbol{\theta})\Big] p(\boldsymbol{\lambda}|\mathbf{Z}, \boldsymbol{\theta}^{[n-1]}), \tag{95}$$

where Λ is the set of all possible $\boldsymbol{\lambda}$. The M step then requires maximization of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{[n-1]})$ over $\boldsymbol{\theta}$. Specifically, the M step determines the nth iterate value:

$$\boldsymbol{\theta}^{[n]} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{[n-1]}). \tag{96}$$

We describe the implementation of the E and M steps for our problem in Sections 8.3.2 and 8.3.3, respectively. Because the EM method is an iterative procedure, it requires an initialization $\boldsymbol{\theta}^{[0]}$ and a criterion for termination. We describe these components of the algorithm, as well as a modification that enables adaptive selection of model order as the algorithm progresses, in Section 8.3.4. Figure 48 depicts a block diagram of the complete algorithm.

8.3.2 Implementation of the E Step

It can be shown[60] that the expected log-likelihood to be calculated in the E step as in (95) can be expressed as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{[n]}) = \sum_{i=1}^{N} Q_i(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{[n]}) + C_K = \sum_{i=1}^{N}\sum_{k=1}^{K} Q_{i,k}(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{[n]}) + C_K, \tag{97}$$

where $Q_i(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{[n]}) = \sum_{k=1}^{K} Q_{i,k}(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{[n]})$ and

$$Q_{i,k}(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{[n]}) = \sum_{j=1}^{M_k} \Pr(\lambda_{k,j} = i|\mathbf{Z}_k, \boldsymbol{\theta}^{[n]}) \left[\log \frac{P_{D'_{k,i}}}{1 - P_{D'_{k,i}}} + \log \rho'_{k,j} - \frac{1}{2} \left(\mathbf{X}_{k,j} - \boldsymbol{\pi}_k(\boldsymbol{\theta}_i)\right)^T R^{-1} \left(\mathbf{X}_{k,j} - \boldsymbol{\pi}_k(\boldsymbol{\theta}_i)\right) - \frac{1}{2\sigma_a^2} \left(a_{k,j} - A(\boldsymbol{\theta}_i, \psi_k, \phi_k)\right)^2 \right] + \log(1 - P_{D'_{k,i}}). \quad (98)$$

In other words, the expected log-likelihood separates into NK terms, each depending only on a single target primitive and the reports in a single image. This decoupling of the expected log-likelihood is a consequence of our independence assumptions of Section 8.2.3. (A similar decomposition will be possible in the M step.) This is an encouraging result, because it means that the computational complexity of the E step will increase only linearly with K and N.

Examining (98), we see that the computation of the E step uses quantities specified previously, as well as the report-to-primitive association probabilities $\Pr(\lambda_{k,j} = i|\mathbf{Z}_k, \boldsymbol{\theta}^{[n]})$. In theory these probabilities can be calculated via Bayes' rule. In practice, however, this computation is typically intractable even for problems of modest size.[13] To overcome this difficulty we use a common and easily justifiable simplification known as *gating* [48]. Specifically, complete enumeration of the set of possible $\boldsymbol{\lambda}_k$ entails consideration of all possible associations, even very unlikely ones in which measured locations $\mathbf{X}_{k,j}$ are associated with target primitives that project to points in the slant plane far from $\mathbf{X}_{k,j}$. Gating is a method for excluding such unlikely pairings from consideration by adaptively defining the set of feasible associations to be the much smaller set of $\boldsymbol{\lambda}_k$ that correspond to associations between reports and primitives that are believed to be "close enough," i.e. , for which $\left\|\mathbf{X}_{k,j} - \boldsymbol{\pi}_k\left(\boldsymbol{\theta}_{\lambda_{k,j}}\right)\right\|_2 \leq r_{\text{gate}}$, $j = 1, \ldots, M_k$. Typically r_{gate} is taken as a small multiple of $(\text{trace}(R))^{1/2}$.

8.3.3 Implementation of the M step

The M step requires maximization of the E step's expected log-likelihood $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{[n]})$ with respect to $\boldsymbol{\theta}$ as in (96). The separation of this expected log-likelihood into independent terms for each primitive in (97) implies that this maximization may be achieved independently for each primitive. In particular, the M step requires N independent maximizations, each of a single $Q_i(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{[n]})$ over $\boldsymbol{\theta}_i$. Since $\boldsymbol{\theta}_i$ includes both continuous parameters (θ_i^X, θ_i^p, θ_i^a, and possibly θ_i^r) and a discrete parameter (θ_i^t), we are faced with a hybrid maximization problem for each primitive, with the discrete parameter limited to a small, finite space of n_t elements. We thus maximize $Q_i(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{[n]})$ by performing n_t separate trial maximizations, one for each possible value of θ_i^t. Examination of (98) reveals that each trial maximization is nontrivial: there is a complicated relationship between $Q_i(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{[n]})$ and the set of continuous parameters. Specifically, the pose, location, and radius terms are coupled due to $\boldsymbol{\pi}_k(\boldsymbol{\theta}_i)$, and the pose and base amplitude are coupled due to $P_{D'_{k,i}}$ and $A(\boldsymbol{\theta}_i, \psi_k, \phi_k)$.

Consider the following approximate maximization over the continuous parameters with θ_i^t fixed, equivalent to a single-iteration coordinate ascent: first, maximize $Q_i(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{[n]})$ over pose while fixing amplitude, location, and radius at their maximizing values from the previous iteration; this can be accomplished with a coarse-to-fine search over pose. Second, perform a line search to maximize over amplitude with pose fixed at the value just obtained, and with location and radius fixed at their values from the previous iteration. Finally,

[13] For the multiple-primitive example of Section 8.4, the set of possible association vectors numbers in the hundreds; for a problem involving as few as a dozen primitives, the cardinality increases to billions.

maximize $Q_i(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{[n]})$ over location and radius while fixing pose and amplitude at the values just obtained; this can be done in closed form due to the quadratic dependence of $Q_i(\boldsymbol{\theta}_i|\boldsymbol{\theta}_i^{[n]})$ on location and radius. This type of partitioned maximization is known as "expectation-conditional maximization" (ECM) and is sufficient to ensure eventual convergence of the EM method to a maximum of the likelihood function under the same conditions as an algorithm that achieves a true maximum at each M step [50]. If not for the pose search, the computational burden of the M step would generally be insignificant compared to that of the E step. As it is, however, the M step greatly exceeds the E step in execution time.

8.3.4 Initialization, Termination, and Model Order Estimation

The block diagram of Figure 48 depicts three stages in addition to the E and M steps described above: a test for termination, an initialization procedure, and a model order reduction stage. Specification of a termination criterion is straightforward. Rather than directly monitoring $p(\mathbf{Z}|\boldsymbol{\theta}^{[n]})$ for convergence, we adopt the computationally simpler and widely used procedure of monitoring the estimates $\boldsymbol{\theta}^{[n]}$ themselves. Once the estimates of θ_i^t produced by the M step remain fixed between iterations and the changes in the continuous parameter estimates all drop below specified thresholds, the iteration is terminated and the final iteration's $\boldsymbol{\theta}^{[n]}$ is used as the final estimate of $\boldsymbol{\theta}$.

Our initialization and model order reduction stages are somewhat more involved, and are fully described elsewhere[60]. Together these components enable adaptive model order selection as the iteration progresses. Our model order adjustment stage is capable only of *reducing* the model order or leaving it unchanged at the conclusion of each EM iteration. This imposes the important guideline that the initialization should be biased toward overestimating N: any overfit can be corrected in subsequent iterations by the model order reduction stage, but any underfit is permanent. Briefly, the initialization procedure groups reports from separate images into clusters based on $\mathbf{X}_{k,j}$ using an iterative chi-squared-statistic-based agglomerative clustering algorithm; reports that can be well-explained as projections from a single point in $\mathbb{R}^3$ are grouped into a single cluster. Each cluster produced by this agglomerative method is used to initialize a single primitive in $\boldsymbol{\theta}^{[0]}$ according to means similar to those described in Section 8.3.3. We implement this initialization so that it tends to overestimate model order N, in accordance with the above observation regarding the model order reduction stage. The model order reduction stage examines the set of report-to-primitive association probabilities calculated in the E step, and removes any primitives which do not seem to correspond to enough reports, i.e. , whose primitive-correspondence probabilities sum to a near-zero quantity. In this way the initialization and model order reduction stages enable adaptive estimation of N as the EM iteration progresses.

8.4 Results

In this section we present results of the application of our algorithm to synthetic SAR imagery generated by XPatch, an electromagnetic simulation package capable of accurately simulating arbitrary electromagnetic scattering measurements obtained by interrogating a facetization-model target with radiation [61]. We use the XPatch-T module of the package to produce image chips at a range and cross-range resolution of 0.3 m, a range and cross-range pixel spacing of 0.2 m, and a center frequency of 9.6 GHz, and use a -35 dB Kaiser sidelobe weighting function for image formation from phase history data. XPatch produces an image of a target in the absence of natural clutter; we model clutter as an additive K-distributed process independent for each pixel, with grassy-terrain parameters [53].

Recall that the measurement model described in Section 8.2.3 is parameterized by several quantities that must be specified in advance. The quantities we use for the experiments in this section are given in

quantity	notation	value
location covariance	R	$(5.0\text{ cm})^2 \times I$
normalized false alarm rate	γ_{FA}	0.0023 / pixel
type confusion matrix	$\{\rho\}$	$\begin{bmatrix} 0.78 & 0.01 & 0.17 & 0.87 \\ 0.22 & 0.99 & 0.83 & 0.13 \end{bmatrix}^T$
amplitude variance	σ_a^2	$(5\text{ dBsm})^2$

Table 8: Measurement model parameters

Table 8. The location covariance, false alarm rate, and confusion matrix given in Table 8 are average values compiled by processing a set of training images, each containing a single primitive in the grassy-terrain clutter environment. The probability-of-detection function and false-alarm amplitude pdf are histograms compiled from the training results. The amplitude variance term is a heuristic value chosen with the intention of capturing some of the variability in primitive responses encountered in the real world that would be difficult to model in a training set (e.g. , geometrical deviations or perturbations from ideality). Finally, the primitive dimensions used to construct primitive scattering models as described in Section 8.2.1 are equal to the true primitive dimensions in all examples presented here; there is a slight degradation in performance when these dimensions are mismatched [60].

For each target described below, we generated a superset of 2736 XPatch images—one for each point on a 2.5° elevation/azimuth grid with elevations from 5° to 50° and azimuths from 0° to 357.5°. For each Monte Carlo run in the trials described below, we selected a random subset of 180 images (if equally spaced on the 2.5° grid, a set of 180 images would form a 10° grid in azimuth and elevation) and corrupted each image with independent K-distributed clutter as described above.

8.4.1 Single-Primitive Targets

Our first set of experiments details the performance of the algorithm on four targets, each consisting of a single primitive. This is useful in establishing a benchmark for the algorithm's performance on more complex targets: results similar to those obtained for a single-primitive target would indicate that the algorithm is performing the data association successfully. Each of the four targets in this section corresponds to a single primitive (a unique type for each target) located at ground coordinates [12″ 0″ 6″]. The trihedral and tophat are oriented with their bases parallel to the ground plane, the trihedral rotated to give a maximum specular response at azimuth 0°; the dihedral and cylinder are oriented so that a maximum specular response is obtained at elevation 25° and azimuth 0°. Each primitive is sized to give a maximum specular RCS of 10 dBsm: the trihedral plate dimension is 4.99″, the tophat has radius 7.24″ and height 14.48″, the dihedral is composed of 5.53″ square plates, and the cylinder has radius 6.97″ and height 20.88″. A total of 100 Monte Carlo runs were performed for each single-primitive target.

Tables 9 and 10 present the performance of the algorithm on these four targets. The "P_{det}" column of Table 9 lists the fraction of runs in which an estimate was produced for the primitive, i.e. , in which it was captured by the initialization stage and survived the model order reduction stage through convergence of the EM iteration. Note that the values in this column reflect the relative observability of the primitive types: trihedrals and tophats have broad angular responses, whereas dihedrals and cylinders have narrow responses largely confined to a single plane[47]. Additionally, the relative heights of the cylinder and dihedral indicate

	P_{det}	confusion				pose az/el rms	pose rot rms
trihedral	1.000	[1.000	0.000	0.000	0.000]	2.331	7.799
tophat	1.000	[0.000	1.000	0.000	0.000]	1.413	n/a
dihedral	0.590	[0.000	0.000	1.000	0.000]	11.778	7.289
cylinder	0.350	[0.000	0.000	0.086	0.914]	0.301	n/a

Table 9: Single-primitive model order, confusion, and pose statistics

that the dihedral response will be broader than the cylinder response [47]; this accounts in part for the better detection of the dihedral.[14]

Of note in Table 9 is the excellent type classification performance of the algorithm: in almost every trial in which the primitive is detected, its type is correctly identified. This suggests that the limited type information provided by the even-bounce/odd-bounce discriminator in the data extraction stage, as discussed in Section 8.2.2, is not an impediment to type estimation. Table 9 also displays statistics from the pose estimation. In general, three Euler angles, corresponding essentially to elevation, azimuth, and rotation, are required to define the pose of a primitive; due to rotational symmetry, two angles suffice for the tophat and cylinder. The "pose az/el rms" column lists the joint root-mean-squared (rms) error in the azimuth and elevation angles, specified as the angular separation in degrees between two points on a sphere. The "pose rot rms" column presents the rms error in the rotation angle for those primitives to which it applies. On the whole, the pose results are encouraging, considering the average spacing between images (10° in elevation and azimuth). The dihedral pose errors, which are larger than those observed for the other primitive types, are attributable to the fact that there is an elevation/azimuth/rotation vector along which direction dihedral responses look very similar [60].

Table 10 displays statistics from the amplitude and location estimation. In the absence of other effects, we would observe a small negative bias in the amplitude estimates, as is observed for the tophat and cylinder; this is attributable to the frequency windowing inherent in the SAR imaging process [54, 46]. In particular, a primitive's brightness in an image is affected by its location in the slant plane relative to the pixel centers. Unless a primitive projects directly onto a pixel center, it will appear dimmer than its RCS would indicate. In the case of the dihedral and trihedral, this effect is counteracted by a slight mismatch between $S_{\theta_i^t}(\bullet)$ and the true scattering responses, induced by forming $S_{\theta_i^t}(\bullet)$ as a noncoherent sum of all the response mechanisms without regard to their relative phases as indicated in Section 8.2.1. Because of the small dimensions of the trihedral and dihedral in this example, their lower-bounce responses are relatively broad and have a noticeable effect on the amplitude estimate in the form of a positive bias.

The Table 10 location estimation statistics are presented in inches. The dihedral and trihedral location estimates exhibit a bias attributable to the influence of the lower-bounce responses from these primitives (e.g. , the double- and single-bounce responses from the trihedral), which do not appear to emanate from the same point as the highest-bounce response and thus violate (89), as described in Section 8.2.1. This effect can be corrected by a post-processing step that re-estimates location using only the highest-bounce reports, as indicated by the estimated primitive pose [60]. Note that the location estimates of the trihedral and tophat have much smaller covariances than that of the dihedral or cylinder; this is due to the relatively narrow responses and resulting low observability of the latter two primitives. Another effect of the narrow responses of the dihedral and cylinder is the large uncertainty along the axis perpendicular to their specular

[14] Also contributing to the better detection of the dihedral as compared to the cylinder is the detrimental effect of the cylinder curvature on the behavior of the initialization stage[60].

	base amplitude			[location] or location radius		
primitive	truth	mean	std dev	truth	mean	(covariance)$^{1/2}$
trihedral	10	10.185	0.408	12 0 6	13.588 0.020 4.015	0.670 0.011 −0.405 0.011 0.340 0.019 −0.405 0.019 0.832
tophat	10	8.590	0.166	12 0 6 7.24	12.055 0.001 6.199 7.111	0.086 −0.003 0.005 −0.001 −0.003 0.097 0.003 −0.002 0.005 0.003 0.386 −0.284 −0.001 −0.002 −0.284 0.291
dihedral	10	9.591	1.293	12 0 6	13.560 −0.060 2.390	1.142 −0.056 −1.613 −0.056 0.447 0.065 −1.613 0.065 3.882
cylinder	10	8.795	0.959	12 0 6 6.97	11.709 −0.028 7.446 6.743	0.670 0.142 −0.207 −0.369 0.142 0.840 −0.333 −0.174 −0.207 −0.333 5.562 −1.142 −0.369 −0.174 −1.142 1.165

Table 10: Single-primitive base amplitude, location, and radius statistics

	P_{det}	confusion	pose az/el rms	pose rot rms
trihedral	1.000	[1.000 0.000 0.000 0.000]	2.693	12.134
tophat	1.000	[0.000 1.000 0.000 0.000]	1.861	n/a
dihedral	0.940	[0.000 0.000 1.000 0.000]	9.746	4.670
cylinder	0.280	[0.000 0.000 0.000 1.000]	0.371	n/a

Table 11: Multiple-primitive model order, confusion, and pose statistics

plane. Note also that due to layover effects, there is significant coupling between the radius and location errors for the tophat and cylinder.

8.4.2 A Multiple-Primitive Target

This section details the performance of our algorithm on a target incorporating one of each of the four primitive types. These primitives are located at four corners of a square, centered at coordinates [±18″ ±18″ 0″], as indicated in Table 11. The cylinder and tophat are the same size as those tested in the previous section; the dihedral and trihedral are larger in order to demonstrate the dependence of the algorithm performance on primitive amplitude. Specifically, the dihedral is composed of 13.11″ square plates to give a maximum RCS of 25 dBsm, and the trihedral is composed of 11.02″ square plates to give a maximum RCS of 23.75 dBsm. The results from 50 Monte Carlo trials are portrayed in Tables 11 and 12 in a format identical to that of Tables 9 and 10.

Examining statistics for the tophat and cylinder in Tables 11 and 12, we see that they are on the whole slightly worse than, but still comparable to, those in Tables 9 and 10. (The slight improvement in the cylinder confusion and location covariance is likely due to the reduced statistical significance stemming

	base amplitude			[location] or location radius		
primitive	truth	mean	std dev	truth	mean	(covariance)$^{1/2}$
trihedral	23.75	22.512	0.511	$\begin{bmatrix} -18 \\ -18 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -19.465 \\ -19.408 \\ -2.637 \end{bmatrix}$	$\begin{bmatrix} 0.361 & 0.122 & 0.307 \\ 0.122 & 0.327 & 0.290 \\ 0.307 & 0.290 & 0.753 \end{bmatrix}$
tophat	10	8.879	0.230	$\begin{bmatrix} -18 \\ 18 \\ 0 \\ 7.24 \end{bmatrix}$	$\begin{bmatrix} -18.070 \\ 17.999 \\ 0.244 \\ 7.082 \end{bmatrix}$	$\begin{bmatrix} 0.101 & 0.001 & 0.004 & 0.008 \\ 0.001 & 0.107 & -0.012 & 0.002 \\ 0.004 & -0.012 & 0.442 & -0.282 \\ 0.008 & 0.002 & -0.282 & 0.298 \end{bmatrix}$
dihedral	25	23.626	1.879	$\begin{bmatrix} 18 \\ 18 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 18.705 \\ 18.670 \\ -2.496 \end{bmatrix}$	$\begin{bmatrix} 0.670 & 0.267 & -0.886 \\ 0.267 & 0.767 & -1.036 \\ -0.886 & -1.036 & 3.698 \end{bmatrix}$
cylinder	10	9.000	1.017	$\begin{bmatrix} 18 \\ -18 \\ 0 \\ 6.97 \end{bmatrix}$	$\begin{bmatrix} 18.119 \\ -18.119 \\ -1.324 \\ 7.196 \end{bmatrix}$	$\begin{bmatrix} 0.491 & -0.173 & -0.343 & -0.045 \\ -0.173 & 0.582 & 0.354 & 0.098 \\ -0.343 & -0.354 & 4.743 & -1.188 \\ -0.045 & 0.098 & -1.188 & 0.896 \end{bmatrix}$

Table 12: Multiple-primitive base amplitude, location, and radius statistics

from the small number of trials in which the cylinder was successfully detected.) This is an indication that the algorithm is generally successful in performing the data association that is implicitly required to solve the model generation problem. Comparing the statistics for the dihedral and trihedral in Tables 11 and 12 to those observed in Tables 9 and 10, several things are apparent. First, and most markedly, the dihedral detection performance has improved drastically for the brighter primitive in this target. Second, the dihedral and trihedral amplitude estimates no longer exhibit the positive bias observed in Table 10, and instead exhibit roughly the same negative bias as the cylinder and tophat. This is because the larger-sized trihedral and dihedral of this experiment have less pronounced lower-bounce responses, and thus the positive bias described in Section 8.4.1 disappears. Third, these primitives still exhibit a location bias, as explained in Section 8.4.1. Fourth, due to the increased observability of these primitives, we obtain smaller location covariances. Finally, the dihedral pose errors have been reduced while the trihedral pose errors have increased; this is because increasing the trihedral amplitude from 10 dBsm has less marginal benefit to the primitive's observability than increasing the dihedral amplitude from the same value, because a 10-dBsm trihedral is already broadly observable. Overall comparison of the results of this experiment to those of the previous section suggests that algorithm performance for simple multiple-primitive targets is similar to the performance when the underlying data association problem is trivial. This is an encouraging result.

8.5 Summary

We have presented an iterative algorithm for producing a 3-D reflector-primitive target model from a collection of SAR images. We simplified the model generation problem by considering only a compressed version of the full set of available data, namely, the locations, amplitudes, and polarimetric signature types of the peaks in each image. This compression, along with the reflector primitive parameterization, enabled us to pose the model generation problem as a data association or fusion problem. We constructed a measurement

model with two major components to relate target primitive parameters to observed data: the first component was a correspondence model describing the detection of target components and false alarms; the second component was a conditional measurement model describing the fine characteristics of the data given the previous component's correspondences. We then showed how the EM method can be applied to the problem; in addition to the E and M steps of the standard EM method, we indicated two modifications that enable adaptive selection of model order as the EM iteration progresses. The algorithm we presented, while tailored to our specific measurement model, is adaptable to a broad class of measurement models. We presented experiments demonstrating the performance of the algorithm on several targets. We demonstrated that dihedrals and especially cylinders are more difficult to detect than either tophats or trihedrals, due to the wider angular response of the latter two primitives. We demonstrated that the performance of the algorithm on a target consisting of a handful of primitives compares favorably to its performance on individual primitives, suggesting that it is successfully solving the implicit data association problem underlying the target model generation problem.

9 References

[1] J. S. De Bonet. Multiresolution sampling procedure for analysis and synthesis of texture images. In *Computer Graphics*. ACM SIGGRAPH, 1997.

[2] J. S. De Bonet and P. Viola. Texture recognition using a non-parametric multi-scale statistical model. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, 1998.

[3] Vince Velten, Timothy Ross, John Mossing Stephen Worrell, and Michael Bryant. Standard SAR ATRevaluation experiments using the MSTAR public release data set. In Edmund G. Zelnio, editor, *Algrorithms for Synthetic Aperture Radar Imagery V*, volume 3370, pages ?–?, 1998.

[4] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons, 1973.

[5] M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky. Modeling and estimation of multiresolution stochastic processes. *IEEE Transactions on Information Theory*, 38(2):766–784, 1992.

[6] W.W. Irving, A.S. Willsky, and L.M. Novak. A multiresolution approach to discriminating targets from clutter in SAR imagery. *Proc. SPIE*, 2487, 1995.

[7] Nikola S. Subotic, Leslie M. Collins, Michael F. Reiley, and John D. Gorman. A multiresolution generalized likelihood ratio detection approach to target screening in synthetic aperture radar data. *Proc. SPIE*, 2487, 1995.

[8] Eero P. Simoncelli, William T. Freeman, Edward H. Adelson, and David J. Heeger. Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38(2):587–607, March 1992.

[9] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[10] S. C. Zhu, Y. Wu, and D. Mumford. Filters random fields and maximum entropy (FRAME): To a unified theory for texture modeling. *To appear in Int'l Journal of Computer Vision.*

[11] J. S. De Bonet and P. Viola. A non-parametric multi-scale statistical model for natural images. In *Advances in Neural Information Processing*, volume 10, 1997.

[12] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[13] 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.

[14] 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.

[15] 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.

[16] M. McClure and L. Carin. Matching Pursuits with a Wave-Based Dictionary. *IEEE Trans. on Sig. Proc.*, 45(12):2912–2927, Dec. 1997.

[17] L. Potter and R. Moses. Attributed Scattering Centers for SAR ATR. *IEEE Trans. on Image Proc.*, 5(1):79–91, Jan. 1997.

[18] M. Rosenblatt. Remarks on Some Nonparametric Estimates of a Density Function. *Annals of Mathematical Statistics*, 27(3):832–837, Sep. 1956.

[19] 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.

[20] 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.

[21] 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.

[22] H. Chiang, R. Moses, and L. Potter. Model-based bayesian feature matching with application to synthetic aperture radar target recognition. *to appear in Pattern Recognition*, 2000.

[23] 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.

[24] J. S. De Bonet. Novel statistical multiresolution techniques for image synthesis, discrimination, and recognition. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, May 1997.

[25] P.M. Dare and I. J. Dowman. An automated procedure for regisering SAR and optical imagery. volume 2958, pages 140–151, 1996.

[26] A. Chao and J. S. De Bonet. SAR image registration using information maximization. ALPHATECH, Inc., August 1997.

[27] P. Viola and A. Chao. Multiple sensor image alignment by maximization of mutual information. Massachusetts Institute of Technology and ALPHATECH, Inc., 1995.

[28] Paul A. Viola. *Alignment by Maximization of Mutual Information.* PhD thesis, Massachusetts Institute of Technology, 1995. MIT AI Laboratory TR 1548.

[29] J. S. De Bonet, P. Viola, and J. Fisher. Flexible histograms: A multiresolution target discrimination model. *Proc. SPIE*, 1998.

[30] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1091, 1953.

[31] McCalla. *Introduction to Numerical Methods and FORTRAN Programming.* John Wiley and Sons, New York, 1967.

[32] S. Kirkpatrick, C.D. Gelatt Jr, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[33] L. Ingber. Simulated annealing: Practice versus theory. *Mathl. Comput. Modelling*, 18(11):29–57, 1993.

[34] L. M. Novak, G. J. Owirka, and C. M. Netishen. Performance of a high-resolution polarimetric SAR automatic target recognition system. *The Lincoln Laboratory J.*, 6(1):11–24, 1993.

[35] V. Larson, L. M. Novak, and C. Stewart. Joint spatial-polarimetric whitening filter to improve SAR target detection performance for spatially distributed targets. *SPIE Conf. on Alg. for SAR Imagery*, April 1994.

[36] T. Cover and J. Thomas. *Elements of Information Theory.* John Wiley & Sons, New York, 1991.

[37] J.W. Fisher and J.C. Principe. Entropy manipulation of arbitrary nonlinear mappings. In J.C. Principe, editor, *Proc. IEEE Workshop, Neural Networks for Signal Processing VII*, pages 14–23, 1997.

[38] J.W. Fisher and J.C. Principe. A methodology for information theoretic feature extraction. In A. Stuberud, editor, *Proceedings of the IEEE International Joint Conference on Neural Networks*, pages ?–?, 1998.

[39] E. Parzen. On estimation of a probability density function and mode. *Ann. of Math Stats.*, 33:1065–1076, 1962.

[40] L. C. Potter, D.-M. Chiang, R. Carrière, and M. J. Gerry. A GTD-based parametric model for radar scattering. *IEEE Trans. Ant. Prop.*, 43:1058–1068, 1995.

[41] M. McClure and L. Carin. Matching pursuits with a wave-based dictionary. *IEEE Trans. Sig. Proc.*, 45:2912–2927, 1997.

[42] J. J. Sacchini, W. M. Steedly, and R. L. Moses. Two-dimensional Prony modeling and parameter estimation. *IEEE Trans. Sig. Proc.*, 41:3127–3137, 1993.

[43] L. C. Potter and R. L. Moses. Attributed scattering centers for SAR ATR. *IEEE Trans. Im. Proc.*, 5:79–91, 1997.

[44] Science Applications International Corporation, Tucson, AZ. *Software User's Manual for the Feature Extraction Module (FES) of MSTAR v3.0.*

[45] E. Ertin and L. C. Potter. Polarimetric classification of scattering centers. In E. G. Zelnio and R. J. Douglass, editors, *Algorithms for SAR Imagery III*, volume 2757, pages 206–216, 1996.

[46] W. G. Carrara, R. S. Goodman, and R. M. Majewski. *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms.* Artech House, Norwood, MA, 1995.

[47] G. T. Ruck, D. E. Barrick, W. D. Stuart, and C. K. Krichbaum. *Radar Cross Section Handbook.* Plenum Press, New York, 1970.

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

[49] A. K. Jain. *Algorithms for Clustering Data.* Prentice-Hall, Englewood Cliffs, NJ, 1988.

[50] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions.* John Wiley & Sons, New York, 1997.

[51] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Royal Stat. Soc. B*, 39:1–38, 1977.

[52] C. Wu. On the convergence properties of the EM algorithm. *Ann. Statistics*, 11:95–103, 1983.

[53] L. M. Novak, M. C. Burl, and W. W. Irving. Optimal polarimetric processing for enhanced target detection. *IEEE Trans. Aero. Elec. Sys.*, 29:234–244, 1993.

[54] J. C. Curlander. *Synthetic Aperture Radar: Systems and Signal Processing.* John Wiley & Sons, New York, 1991.

[55] D. E. Dudgeon and R. T. Lacoss. An overview of automatic target recognition. *Lincoln Laboratory Journal*, 6(1):3–10, 1993.

[56] E. R. Keydel and S. W. Lee. Signature prediction for model-based automatic target recognition. In E. G. Zelnio and R. J. Douglass, editors, *Algorithms for Synthetic Aperture Radar Imagery III*, volume 2757 of *Proc. SPIE*, pages 306–317, 1996.

[57] J. A. Richards, J. W. Fisher, and A. S. Willsky. Target model generation from multiple SAR images. In E. G. Zelnio, editor, *Algorithms for SAR Imagery VI*, volume 3721, pages 598–611, 1999.

[58] E. C. G. Sudarshan and N. Mukunda. *Classical Dynamics: A Modern Perspective.* John Wiley & Sons, New York, 1974.

[59] F. T. Ulaby and C. Elachi, editors. *Radar Polarimetry for Geoscience Applications.* Artech House, Norwood, MA, 1990.

[60] J. A. Richards. *Target Model Generation from Multiple Synthetic Aperture Radar Images.* PhD thesis, M.I.T., to appear.

[61] M. Hazlett, D. J. Andersch, S. W. Lee, H. Ling, and C. L. Yu. XPatch: a high-frequency electromagnetic scattering prediction code using shooting and bouncing rays. In W. R. Watkins and D. Clement, editors, *Targets and Backgrounds: Characterization and Representation*, volume 2469, pages 266–275, 1995.

Bibliography

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

M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky. Modeling and estimation of multiresolution stochastic processes. *IEEE Transactions on Information Theory*, 38(2):766–784, 1992.

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.

J. S. De Bonet and P. Viola. A non-parametric multi-scale statistical model for natural images. In *Advances in Neural Information Processing*, volume 10, 1997.

J. S. De Bonet and P. Viola. Texture recognition using a non-parametric multi-scale statistical model. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, 1998.

W. G. Carrara, R. S. Goodman, and R. M. Majewski. *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms.* Artech House, Norwood, MA, 1995.

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.

A. Chao and J. S. De Bonet. SAR image registration using information maximization. ALPHATECH, Inc., August 1997.

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.

H. Chiang, R. Moses, and L. Potter. Model-based bayesian feature matching with application to synthetic aperture radar target recognition. *to appear in Pattern Recognition*, 2000.

T.M. Cover and J.A. Thomas. *Elements of Information Theory.* John Wiley & Sons, 1991.

J. C. Curlander. *Synthetic Aperture Radar: Systems and Signal Processing.* John Wiley & Sons, New York, 1991.

P.M. Dare and I. J. Dowman. An automated procedure for regisering SAR and optical imagery. volume 2958, pages 140–151, 1996.

J. S. De Bonet. Multiresolution sampling procedure for analysis and synthesis of texture images. In *Computer Graphics.* ACM SIGGRAPH, 1997.

J. S. De Bonet. Novel statistical multiresolution techniques for image synthesis, discrimination, and recognition. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, May 1997.

J. S. De Bonet, P. Viola, and J. Fisher. Flexible histograms: A multiresolution target discrimination model. *Proc. SPIE*, 1998.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Royal Stat. Soc. B*, 39:1–38, 1977.

R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis.* John Wiley and Sons, 1973.

D. E. Dudgeon and R. T. Lacoss. An overview of automatic target recognition. *Lincoln Laboratory Journal*, 6(1):3–10, 1993.

E. Ertin and L. C. Potter. Polarimetric classification of scattering centers. In E. G. Zelnio and R. J. Douglass, editors, *Algorithms for SAR Imagery III*, volume 2757, pages 206–216, 1996.

J.W. Fisher and J.C. Principe. Entropy manipulation of arbitrary nonlinear mappings. In J.C. Principe, editor, *Proc. IEEE Workshop, Neural Networks for Signal Processing VII*, pages 14–23, 1997.

J.W. Fisher and J.C. Principe. A methodology for information theoretic feature extraction. In A. Stuberud, editor, *Proceedings of the IEEE International Joint Conference on Neural Networks*, pages ?–?, 1998.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

M. Hazlett, D. J. Andersch, S. W. Lee, H. Ling, and C. L. Yu. XPatch: a high-frequency electromagnetic scattering prediction code using shooting and bouncing rays. In W. R. Watkins and D. Clement, editors, *Targets and Backgrounds: Characterization and Representation*, volume 2469, pages 266–275, 1995.

L. Ingber. Simulated annealing: Practice versus theory. *Mathl. Comput. Modelling*, 18(11):29–57, 1993.

W.W. Irving, A.S. Willsky, and L.M. Novak. A multiresolution approach to discriminating targets from clutter in SAR imagery. *Proc. SPIE*, 2487, 1995.

A. K. Jain. *Algorithms for Clustering Data.* Prentice-Hall, Englewood Cliffs, NJ, 1988.

E. R. Keydel and S. W. Lee. Signature prediction for model-based automatic target recognition. In E. G. Zelnio and R. J. Douglass, editors, *Algorithms for Synthetic Aperture Radar Imagery III*, volume 2757 of *Proc. SPIE*, pages 306–317, 1996.

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.

S. Kirkpatrick, C.D. Gelatt Jr, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

V. Larson, L. M. Novak, and C. Stewart. Joint spatial-polarimetric whitening filter to improve SAR target detection performance for spatially distributed targets. *SPIE Conf. on Alg. for SAR Imagery*, April 1994.

McCalla. *Introduction to Numerical Methods and FORTRAN Programming.* John Wiley and Sons, New York, 1967.

M. McClure and L. Carin. Matching pursuits with a wave-based dictionary. *IEEE Trans. Sig. Proc.*, 45:2912–2927, 1997.

G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions.* John Wiley & Sons, New York, 1997.

N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1091, 1953.

L. M. Novak, M. C. Burl, and W. W. Irving. Optimal polarimetric processing for enhanced target detection. *IEEE Trans. Aero. Elec. Sys.*, 29:234–244, 1993.

L. M. Novak, G. J. Owirka, and C. M. Netishen. Performance of a high-resolution polarimetric SAR automatic target recognition system. *The Lincoln Laboratory J.*, 6(1):11–24, 1993.

E. Parzen. On estimation of a probability density function and mode. *Ann. of Math Stats.*, 33:1065–1076, 1962.

L. C. Potter, D.-M. Chiang, R. Carrière, and M. J. Gerry. A GTD-based parametric model for radar scattering. *IEEE Trans. Ant. Prop.*, 43:1058–1068, 1995.

L. C. Potter and R. L. Moses. Attributed scattering centers for SAR ATR. *IEEE Trans. Im. Proc.*, 5:79–91, 1997.

J. A. Richards. *Target Model Generation from Multiple Synthetic Aperture Radar Images.* PhD thesis, M.I.T., to appear.

J. A. Richards, J. W. Fisher, and A. S. Willsky. Target model generation from multiple SAR images. In E. G. Zelnio, editor, *Algorithms for SAR Imagery VI*, volume 3721, pages 598–611, 1999.

M. Rosenblatt. Remarks on Some Nonparametric Estimates of a Density Function. *Annals of Mathematical Statistics*, 27(3):832–837, Sep. 1956.

G. T. Ruck, D. E. Barrick, W. D. Stuart, and C. K. Krichbaum. *Radar Cross Section Handbook.* Plenum Press, New York, 1970.

J. J. Sacchini, W. M. Steedly, and R. L. Moses. Two-dimensional Prony modeling and parameter estimation. *IEEE Trans. Sig. Proc.*, 41:3127–3137, 1993.

Science Applications International Corporation, Tucson, AZ. *Software User's Manual for the Feature Extraction Module (FES) of MSTAR v3.0.*

Eero P. Simoncelli, William T. Freeman, Edward H. Adelson, and David J. Heeger. Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38(2):587–607, March 1992.

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.

Nikola S. Subotic, Leslie M. Collins, Michael F. Reiley, and John D. Gorman. A multiresolution generalized likelihood ratio detection approach to target screening in synthetic aperture radar data. *Proc. SPIE*, 2487, 1995.

E. C. G. Sudarshan and N. Mukunda. *Classical Dynamics: A Modern Perspective.* John Wiley & Sons, New York, 1974.

F. T. Ulaby and C. Elachi, editors. *Radar Polarimetry for Geoscience Applications.* Artech House, Norwood, MA, 1990.

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.

P. Viola and A. Chao. Multiple sensor image alignment by maximization of mutual information. Massachusetts Institute of Technology and ALPHATECH, Inc., 1995.

Paul A. Viola. *Alignment by Maximization of Mutual Information.* PhD thesis, Massachusetts Institute of Technology, 1995. MIT AI Laboratory TR 1548.

C. Wu. On the convergence properties of the EM algorithm. *Ann. Statistics*, 11:95–103, 1983.

S. C. Zhu, Y. Wu, and D. Mumford. Filters random fields and maximum entropy (FRAME): To a unified theory for texture modeling. *To appear in Int'l Journal of Computer Vision.*

LIST OF ACRONYMS

ACRONYM	DESCRIPTION
ATD/R	Automatic target detection/recognition
ATR	Automatic target recognition
CFAR	Constant false alarm rate
DARPA	Defense Advanced Research Projects Agency
ERIM	Environmental Research Institute of Michigan
GLLR	Generalized log-likelihood ratio
GTD	Geometric thoery of diffraction
KL	Kullback-Leibler
LLR	log-likelihood ratio
MI	Mutual information
MLP	Multi-layer perceptron
MSTAR	Moving and stationary target acquisition and recognition
PCA	Principal components analysis
RMS	Root mean square
ROC	Receiver operating characteristic
ROI	Region(s) of interest
SAR	Synthetic aperture radar
SBIR	Small business innovative research
WLS	Weighted least squares