

Model-Based Classification of Radar Images

Hung-Chih Chiang, Randolph L. Moses, *Senior Member, IEEE*, and Lee C. Potter, *Senior Member, IEEE*

Abstract—A Bayesian approach is presented for model-based classification of images with application to synthetic-aperture radar. Posterior probabilities are computed for candidate hypotheses using physical features estimated from sensor data along with features predicted from these hypotheses. The likelihood scoring allows propagation of uncertainty arising in both the sensor data and object models. The Bayesian classification, including the determination of a correspondence between unordered random features, is shown to be tractable, yielding a classification algorithm, a method for estimating error rates, and a tool for evaluating performance sensitivity. The radar image features used for classification are point locations with an associated vector of physical attributes; the attributed features are adopted from a parametric model of high-frequency radar scattering. With the emergence of wideband sensor technology, these physical features expand interpretation of radar imagery to access the frequency- and aspect-dependent scattering information carried in the image phase.

Index Terms—Model-based classification, parametric modeling, point correspondence, radar image analysis.

I. INTRODUCTION

SYNTHETIC-aperture radar (SAR) provides all-weather, day-or-night remote sensing for mapping, search-and-rescue, mine detection, and target recognition [1]. SAR data processing entails forming an image from measured radar backscatter returns, followed by processing to detect and recognize targets from the formed image. Current SAR processing practice decouples the image formation from the decision task for which the imagery is ultimately intended.

In this paper, a Bayesian-model-based imaging and decision approach is presented for classification of radar images. The approach provides a structured, implementable, scalable means for managing complexity of the hypothesis set and bypassing the complexity of joint distributions on image pixels. Model-based classification, or pattern matching, combines uncertainty in both the object class models and the sensor data to compute posterior probabilities of hypotheses. The Bayesian formalism allows clear and explicit disclosure of all assumptions. The pattern matching permits tractable performance estimation and provides robustness against environments previously not measured, and hence not available for construction of image templates.

Manuscript received September 1, 1999; revised April 6, 2000. This work was supported by the U.S. Air Force Materiel Command under Contract F33615-97-1020.

H.-C. Chiang was with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210 USA. He is now with Etrend Electronics, Inc., Kaohsiung, Taiwan 80778, R.O.C. (e-mail: chiang@etrend.com.tw).

R. L. Moses and L. C. Potter are with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: moses.2@osu.edu; potter.36@osu.edu).

Communicated by J. A. O'Sullivan, Guest Editor.
Publisher Item Identifier S 0018-9448(00)06070-3.

A. Problem Complexity

Classification of radar images, like many image inference tasks, is characterized by a complex hypothesis space. The hypothesis set consists of M classes, or objects; typical cases are $8 \leq M \leq 32$. The complexity arises in that each object may be observed in a variety of poses, configurations, articulations, and environments, thereby resulting in an intractable density function for the radar image conditioned solely on the object class. To manage the complexity, object classes are each expressed as a mixture density of subclasses. Each subclass is defined by a deterministic description of object pose, configuration, articulation, occlusion, sensor orientation, etc. Additional variability within subclasses is modeled stochastically to account for object differences due, for example, to manufacturing variations or wear. The number of enumerated subclasses explodes exponentially; a typical application might dictate 10^{12} states for each hypothesis class [2]. Moreover, an application may dictate that many more than M decision classes be formed by defining sets of individual subclasses, H_{ij} ; e.g., the configuration of an object may be an important distinguishing characteristic.

Likewise, the classification of radar images is characterized by a high-dimensional observation space defying a direct random model. The observation, a collection of pixels, is a vector in R^N . A typical case is a 128-by-128 array of complex-valued pixels, yielding $N = 2^{15}$. A joint density on the N pixel values, when conditioned on a hypothesis H_{ij} , is non-Gaussian and may be multimodal [7]. For example, a simple Gaussian uncertainty on the location of a scattering mechanism leads to non-Gaussian image pixel uncertainties. Further, pixel values exhibit strong correlation due to the coherent combination of scattered energy from an object's constituent parts. Multiple reflections or large conducting surfaces can result in large distances between correlated pixels, and hence seemingly arbitrary correlation matrices.

B. Model-Based Classification

To proceed when confronting a large hypothesis space and complex image density functions, we adopt a model-based classification approach. First, a physically based feature set provides a simple, constructive alternative to joint densities on pixels for expressing uncertainty in the target and the sensor. The extraction of features is performed by statistical estimation using the physics-based parametric model of sensor data and specification of the image formation procedure. Second, a coarse-to-fine staged classification strategy is used to efficiently search the hypothesis space. Third, the sensor data model is combined with object models to predict features conditioned on a hypothesis. The on-line prediction of features eliminates the need for a prohibitively large catalog of image templates.

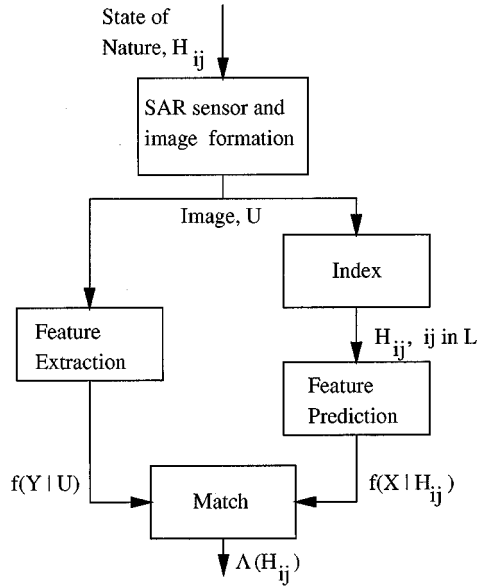


Fig. 1. A model-based approach to classification.

The model-based approach is depicted in Fig. 1. A state of nature is characterized by the hypothesis of an object class H_i which is further specified by one of finitely many subclasses H_{ij} . The SAR measurement resulting from a sensor and an image-formation algorithm provides an image U . Along the left branch in the figure, a Feature Extraction stage serves to compress the image and assign uncertainties to features. For SAR imaging, a sensor data model derived from high-frequency approximation to scattering physics provides a parametric family of densities for estimating features. Parameters are estimated from imagery and used as low-dimensional surrogates for sufficient statistics; each feature is a location together with a vector of attributes. The feature uncertainty is given as a density function $f(Y | U)$ and acknowledges the sensitivity of parameter estimates Y to noisy sensor data given the image data U .

Along the right branch in the figure, complexity of the hypothesis space is addressed in a coarse-to-fine approach. An Index stage provides a list of candidate subclass hypotheses H_{ij} , $ij \in L$ based on a coarse partitioning of the hypothesis space. Evaluation of the candidate hypotheses then proceeds using a model for the observations. A Feature Prediction stage computes a predicted feature set by combining the sensor data model from the Feature Extraction stage and a computer-aided design (CAD) representation of a hypothesis H_{ij} . The feature set X has an associated uncertainty, acknowledging error in the modeling and variation among objects in the subclass. The uncertainty is expressed as a density $f(X | H_{ij})$. Importantly, the use of physically motivated features facilitates compatibility of extracted and predicted feature sets.

Finally, the predicted and extracted feature sets are combined in a Match stage to compute a posterior probability of a candidate hypothesis $\Lambda(H_{ij})$. The top hypotheses, and their likelihoods, are reported as the output of the classification system. Computation of the likelihood scores requires a correspondence between the unordered lists of extracted and predicted features and an integration over feature uncertainty. Further, the likeli-

hood must incorporate the possibilities of missing or spurious features in the predicted or extracted feature lists. The matching task can be viewed as a probabilistic graph match of fully connected, attributed graphs with deletions and insertions of nodes.

C. Contributions and Organization

This paper presents a Bayesian formalism for model-based classification. We demonstrate that the resulting hypothesis testing algorithm, including the feature correspondence, is tractable, even for problem sizes encountered in SAR target recognition. In addition, the paper adopts a physics-based model for extracting features from SAR images; the features use the phase in complex-valued SAR images to infer the frequency- and aspect-dependent scattering behavior of objects. Recent advances in technology yield sensor bandwidths exceeding 20% of the center frequency; for such systems, the proposed feature sets provide much greater information than does processing motivated by a narrowband point-scattering model.

Detailed construction of the Index and Feature Prediction stages is not considered here; these stages are discussed in [3] and [7]. An adaptive refinement of the candidate hypothesis list from the Index stage is considered in [6]. Moreover, a Feature Prediction stage that faithfully simulates frequency- and aspect-dependent scattering behavior is currently under development [7].

The paper is organized as follows. In Section II we present a parametric model for radar sensor data, as required in the Feature Extraction and Feature Prediction stages. Maximum-likelihood estimation of parameters from images computed using sensor data is discussed; also, parameter uncertainty, the definition of image resolution, and the Fisher information in image phase are addressed. Section III presents a Bayesian computation of a hypothesis likelihood given sets of extracted and predicted features. In particular, the problem of determining a feature correspondence is addressed.

In Section IV, synthetic classification results are computed using class means estimated from a measured set of X-band radar images for ten objects. The simulation results illustrate four points: 1) the Bayes approach to model-based classification, including feature correspondence, is tractable; 2) classification using the Bayes classifier permits estimation of the optimal error rate, given the assumed priors and feature uncertainties; 3) classification using the Bayes classifier allows designers to explore the performance effects of sensor parameters, such as bandwidth; and iv) classification using the Bayes classifier provides a simulation tool to investigate sensitivity of the estimated error rate to the assumed priors and feature uncertainties.

II. A PHYSICAL MODEL FOR SENSOR DATA AND FEATURE EXTRACTION

In this section we address the problem of feature extraction. We adopt a parametric model describing the sensor data, develop a feature estimation algorithm, and discuss feature uncertainty both for extraction and feature prediction. The model we employ is based on high-frequency approximation of electromagnetic scattering [9], [11] and represents the object of interest

as a set of scattering centers. The scattering centers are described by attributes that characterize the scattering center geometry and orientation. The attributed scattering centers are used as features for both the prediction and extraction stages in Fig. 1. The scattering model provides a method of constructing and succinctly representing hypotheses from CAD representations of class objects. Additionally, the model allows feature extraction to be cast as a parameter estimation problem.

For a Bayesian classifier, uncertainty must be characterized for both predicted and extracted feature sets. Because the proposed features relate directly to physical components in a CAD representation, uncertainty in predicted features can be estimated from uncertainty in the CAD model. This is an important advantage of using a physics-based model; other parametric models could be used to represent the measured data, but unless the model parameters relate to scattering physics, it is very difficult to model the prediction uncertainty $f(X | H_{ij})$ in Fig. 1. In addition, a parameter-estimation formulation of feature extraction provides means for describing feature uncertainty $f(Y | U)$ and for bounding it with the Cramér–Rao bound.

The model-based interpretation of images permits an information-theoretic view of SAR imaging. We consider two implications of this viewpoint. First, we define SAR image resolution in terms of uncertainty in estimated parameters. Second, we consider performance degradation when incomplete data are available. Incomplete data availability results in higher feature uncertainty as measured by relative information; as an example, we consider the increase in uncertainty that results from the common practice of discarding the phase of the SAR image.

A. A Parametric Model for Object Scattering

Most feature extraction models used with SAR rely on processing of the magnitude image. For example, features used in the MSTAR program are peaks (local maxima of the SAR magnitude image) and ridges obtained from directional derivatives of the SAR magnitude image [3]. When the complex-valued SAR image is used, the point-scattering model is most commonly employed; in this model, the backscattered amplitude is assumed to be independent of frequency and aspect. The point-scattering assumption leads to a two-dimensional harmonic scattering model, and parameter estimation becomes a two-dimensional harmonic retrieval problem [4], [5]. One drawback of peak and point scattering models is that a single scattering object, such as a dihedral, is modeled as several peaks or point scatterers; in this case, the correlated uncertainty in the estimated parameters is difficult to model. Similarly, the relationship of ridge features to scattering geometry is not well understood, and feature uncertainty is hard to predict.

In this paper we adopt the physical radar scattering model from Gerry *et al.* [9], which assumes a data collection scenario consistent with SAR imaging. A reference point is defined, and the radar trajectory is required to be coplanar with the reference point. This plane, the imaging plane, is labeled using an $x - y$ Cartesian coordinate system with origin at the reference point. The radar position is then described by an angle ϕ defined counterclockwise from the x direction. Far-zone backscatter is assumed, and therefore planewave incidence is obtained on illuminated objects.

From the geometric theory of diffraction (GTD) [12], [13], if the wavelength of the incident excitation is small relative to the object extent, then the backscattered field from an object consists of contributions from electrically isolated scattering centers. The backscattered field of an individual scattering center is described as a function of frequency ω and aspect angle ϕ , and the total scattered field from a target is then modeled as the sum of these individual scatterers [9]

$$s(\omega, \phi; \theta) = \sum_{k=1}^N A_k \left(j \frac{\omega}{\omega_c} \right)^{\alpha_k} \text{sinc} \left(\frac{\omega}{c} L_k \sin(\phi - \bar{\phi}_k) \right) \cdot \exp(-j\omega\gamma_k \sin \phi) \cdot \exp \left(j \frac{\omega}{c/2} (x_k \cos \phi + y_k \sin \phi) \right). \quad (1)$$

In (1), ω_c is the center frequency of the radar bandwidth, and c is the speed of propagation. Each of N scattering centers is characterized by seven attributes: (x_k, y_k) denotes the scattering center location projected to the (x, y) -plane, A_k is a relative amplitude, L_k is the scattering center length, $\bar{\phi}_k$ its orientation angle, α_k characterizes frequency dependence of the scattering center, and γ_k models the mild aspect dependence of scattering center cross-section (for example, the projected cross-sectional area of a trihedral changes slightly with aspect angle). The scattering model is described by the parameter set $\theta = [\theta_1, \dots, \theta_N]$, where each vector $\theta_k = [x_k, y_k, A_k, \alpha_k, L_k, \bar{\phi}_k, \gamma_k]$ is the collection of the seven parameters, or attributes, defining each scattering center.

The frequency and aspect dependence of the scattering centers is an important distinction of this model and permits description of a rich variety of scattering primitives. The frequency dependence relates directly to the curvature of the scattering object and is parameterized by α_k , which takes on integer or half-integer values. For example, $\alpha_k = 1$ describes flat surface scattering, $\alpha_k = 1/2$ describes scattering from singly curved surfaces, and $\alpha_k = 0$ indicates scattering from doubly curved surfaces or edges. Values of α less than zero describe diffraction mechanisms, such as edges and tips. In addition, the sinc aspect dependence in (1) reveals the effective length L_k of the scattering primitive. Many scattering geometries, such as dihedrals, corner reflectors, and cylinders, are distinguishable by their (α, L) parameters [9], as shown in Fig. 2. Point scattering is a special case of the model in (1) for $\alpha_k = L_k = \gamma_k = 0$.

The model in (1) is based on GTD and physical optics approximations for scattering behavior and, while parsimonious, is able to describe a large class of scatterers. Scattering objects separated by approximately two or more wavelengths are distinguishable [10]. Physical behaviors not well modeled by (1) for small N include creeping waves and cavity scattering [9].

B. Parameter Estimation

Next, we describe an approximate maximum-likelihood technique for extracting the model parameters in (1) from measured sensor data. The measured data is modeled as

$$r(\omega, \phi) = s(\omega, \phi; \theta) + n(\omega, \phi) \quad (2)$$







	$\alpha=1$	$\alpha=1/2$	$\alpha=0$
$L=0$	 Trihedral	 Top Hat	 Sphere
$L>0$	 Dihedral	 Cylinder	 Edge

Fig. 2. Canonical scattering geometries that are distinguishable from (α, L) pairs in the scattering model.

where $n(\omega, \phi)$ is a noise term that represents the modeling error (background clutter, sensor noise, model mismatch, incomplete motion compensation, antenna calibration errors, etc.) and can be modeled as a zero-mean, Gaussian noise process with known covariance.

The measured data is often transformed into the image domain as an array of complex-valued pixels. The transformation comprises equalization (to compensate for nonideal sensor characteristics), windowing, zero padding, and discrete Fourier transformation. The transformation can be represented by the linear operator \mathcal{L} ; thus

$$\begin{aligned}\tilde{r}(x, y) &= \mathcal{L}[s(\omega, \phi; \theta) + n(\omega, \phi)] \\ &= \tilde{s}(x, y; \theta) + \tilde{n}(x, y)\end{aligned}\quad (3)$$

for a finite array of sample points (x, y) . We see that $\tilde{n}(x, y)$ is a zero-mean Gaussian noise process with known covariance. The feature-extraction problem is thus one of estimating the parameter vector θ from the measurement $\tilde{r}(x, y)$.

R. A. Fisher's pioneering work laid a foundation for parametric modeling as a method of data compression, and established maximum-likelihood procedures for estimation of the unknown parameters [8]. Since $\tilde{r}(x, y)$ are Gaussian measurements, the parameter vector θ which maximizes the likelihood function is found as

$$\hat{\theta}_{\text{ML}} = \arg \min_{\theta} J(\theta) \quad (4)$$

$$J(\theta) = [\tilde{r} - \tilde{s}(\theta)]^H \hat{\Sigma}^{-1} [\tilde{r} - \tilde{s}(\theta)] \quad (5)$$

where \tilde{r} , \tilde{n} , and $\tilde{s}(\theta)$ are vectors obtained by stacking the columns of $\tilde{r}(x, y)$, $\tilde{s}(x, y; \theta)$, and $\tilde{n}(x, y)$, respectively; $\hat{\Sigma} = \text{cov}(\tilde{n})$, and $(\cdot)^{\dagger}$ denotes Moore–Penrose pseudoinverse. Furthermore, this estimator is robust to model mismatch [15]. Equation (4) is a nonlinear least squares minimization problem.

We make use of the fact that scattering center responses are localized in the image domain to develop a computationally simpler approximate maximum-likelihood estimator for θ [14]. The minimization in (4) is decomposed into smaller estimation problems. We partition the image into M disjoint regions R_i of high energy and a remainder region R_0 . Defining Π_i as the projection onto region R_i , we have

$$J(\theta) \approx \sum_{i=0}^M [\tilde{r} - \tilde{s}(\theta^i)]^H \Pi_i \hat{\Sigma}^{-1} \Pi_i [\tilde{r} - \tilde{s}(\theta^i)] + C \quad (6)$$

where θ^i is a vector containing the parameters for scattering centers in region R_i and C is a constant independent of θ . Since

the number of pixels in R_i is much less than the total number of image pixels in \tilde{r} and the θ^i form a partition of θ , the individual minimization problems in (6) are decoupled and have many fewer unknowns than the minimization problem in (5). The weighted least squares estimator is tractable and provides nearly efficient parameter estimates for data satisfying the scattering model in (1) with colored Gaussian noise on image pixels [14].

An additional advantage of the approximate maximum-likelihood (ML) algorithm is its robustness to the assumed noise model. The assumption of correlated Gaussian noise across the entire image is not very accurate for scenes where clutter is present in the form of trees, power lines, etc. However, this assumption is much better over small image regions that primarily contain target-scattering centers. Image segmentation also facilitates model order selection, which is implemented using the minimum description length principle.

As an illustration of the approximate ML estimation, Fig. 3 shows the results of feature extraction on a measured SAR image from the MSTAR Public Targets dataset [26]. For $N = 30$, the algorithm models 96.5% of the energy in the image chip shown. In addition, the T-72 tank barrel segment is modeled as a single scattering center whose length is modeled within 10 cm of the actual 1.37-m length. In comparison, peak-based scattering center extraction methods model this segment as three peaks spaced along the barrel. Execution time for extraction of 30 scattering features using unoptimized Matlab code on a 450-MHz Pentium processor is approximately 140 s using (5) and approximately 50 s if a suboptimal but computationally efficient estimator is employed.

C. Parameter Uncertainty

Use of estimated model parameters for Bayesian hypothesis testing requires that an uncertainty be associated with each estimate. The inverse of Fisher information is used to predict the error covariance of the approximate maximum-likelihood estimation algorithm in (6).

The Cramér–Rao lower bound is derived in [9] and provides an algorithm-independent lower bound on the error variance for unbiased estimates of the model parameters. The derivation assumes the data model in (3). For any choice of model parameters, the covariance bound is computed by inversion of the Fisher information matrix [16]

$$I(\theta) = -E \left\{ \frac{\partial^2 \ln f(s|\theta)}{\partial \theta^2} \right\} \quad (7)$$

where $f(s|\theta)$ is density on the sensor data s conditioned on the parameter θ .

D. Image Resolution

As noted in [17], “a universally acceptable definition of resolution as a performance measure is elusive.” In SAR, image resolution is typically reported as the width of a point-spread function. This definition is a Rayleigh resolution and is determined by sensor bandwidth, range of viewing angles, and degree of sidelobe suppression in image formation. In contrast, for model-based interpretation of SAR imagery we define reso-

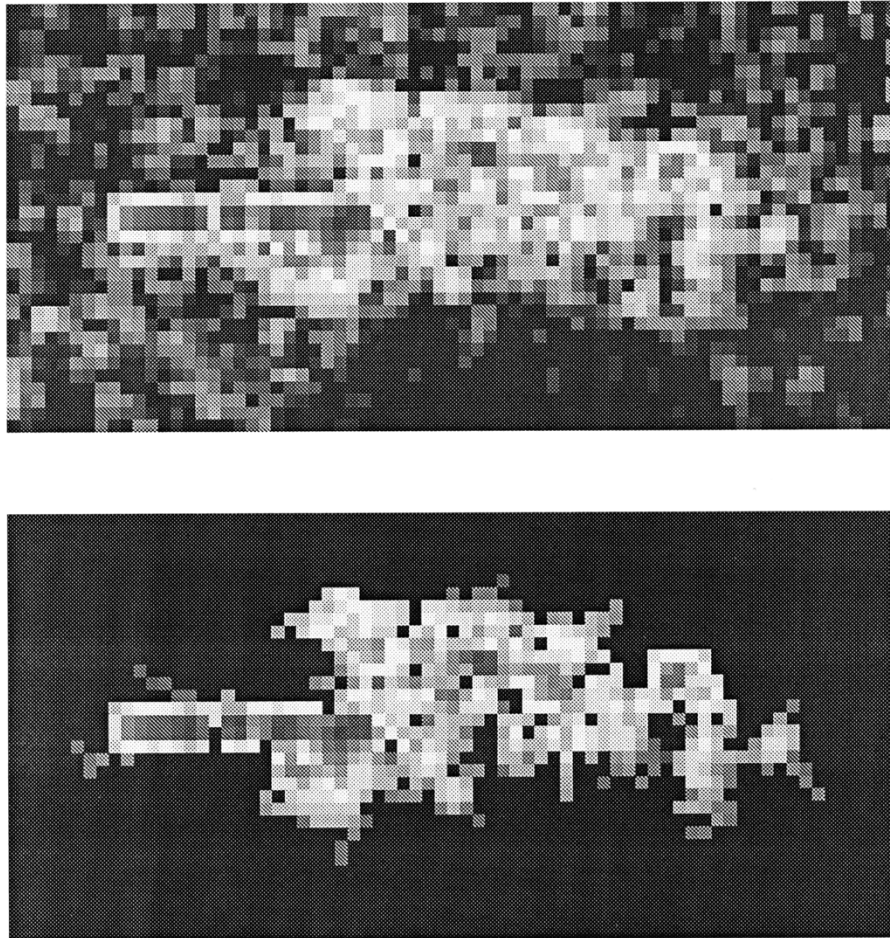


Fig. 3. Measured SAR Image of T-72 Tank (top) and reconstruction from estimated parameters (bottom). Images are in decibel magnitude with a total range of 40 dB.

lution in terms of a bound on the uncertainty in estimated parameters. Prior knowledge of the scattering behavior, as encoded in (1), results in an uncertainty-based resolution that is often much finer than the Rayleigh resolution. For example, consider application of feature uncertainty to the classical notion of separating closely spaced point sources, i.e., $\alpha = L = \gamma = 0$ in (1). For a given signal-to-noise ratio (SNR) of a single-point scatterer (SNR per mode), let the resolution be defined as the minimum distance between two equal amplitude scattering centers resulting in nonoverlapping 95% confidence regions for the estimated locations [9], [18].

Adopting this definition, resolution versus SNR per mode is shown in Fig. 4 for a SAR with Rayleigh resolution of 30 cm. The resolution depends on the orientation of the two point scatterers. The dashed line shows resolution for point scatterers separated an equal distance in both down range and cross range (i.e., aligned 45° to the aperture). The solid line and the dash-dot line show resolution for two point scatterers aligned parallel and orthogonal to the aperture, respectively. For an SNR per mode of -5 dB and 500-MHz bandwidth, the limit of resolution achievable by model-based scattering analysis is approximately one-half the Rayleigh resolution; model-based resolution is limited by sensor bandwidth and SNR, which includes mismatch from the model in (3).

In the figure, we report signal-to-noise (SNR) values using the ratio of signal energy to noise energy computed for the frequency-aspect domain samples. Alternatively, SNR may be interpreted in the image space as a difference between peak signal level and clutter floor. However, this image space definition of SNR varies depending on the specific values of the parameter vector θ describing the scattering center.

E. Magnitude-Only Fourier Data

The parameter uncertainty definition of resolution can be directly applied to image reconstruction from incomplete data; for example, in SAR image formation a common practice is to discard image phase. In this case, the estimation of $s(\omega, \phi)$ becomes reconstruction from magnitude-only Fourier data. The Fisher information $I_{\text{mag}}(\theta)$ can be computed for the sampled magnitude of the image data, using (1) and knowledge of both the sensor transfer function and the image formation operator. The relative information [19] is the ratio of Fisher matrices $I_{\text{mag}}(\theta)/I(\theta)$ and quantifies the loss of information incurred by discarding the image phase. Likewise, the increase in variance in any parameter estimate can be predicted, for efficient estimators, using the Cramér–Rao bounds.

For example, for $\omega_c/(2\pi) = 10$ GHz, 3-GHz bandwidth, and 10-dB SNR, the Cramér–Rao bound on standard deviation in

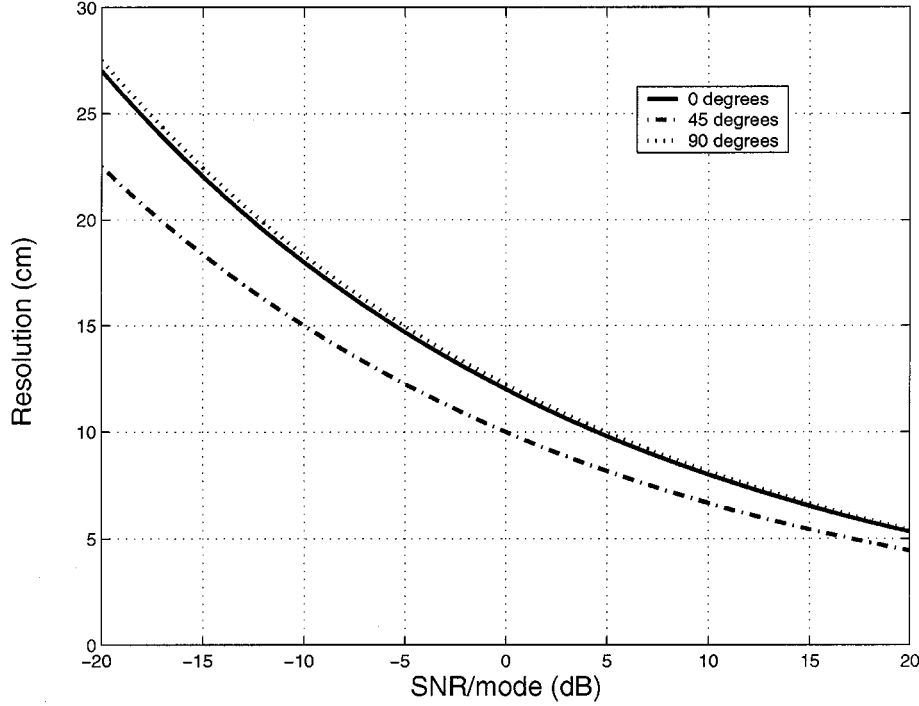


Fig. 4. Resolution versus SNR for three different orientations of two point scatterers; $\omega_c/(2\pi) = 10$ GHz, and bandwidth is 500 MHz.

estimation error for α is $\sigma = 1.0$ using a magnitude image. In contrast, estimation of α from the complex-valued image results in $\sigma = 0.02$. Thus use of complex-valued imagery allows inference of the frequency-dependent scattering behavior, whereas use of magnitude-only imagery does not.

III. HYPOTHESIS TESTING

A. Problem Statement

In this section we derive the Bayes match function used for classification from feature vectors. At the input to the classifier stage, we have a given region of interest (a SAR image chip), along with a set $\mathcal{H} = \{H_{ij}, i, j \in L\}$ of K candidate target hypotheses and their prior probabilities $P(H_{ij})$. Each hypothesis contains both target class and subclass information; the set \mathcal{H} may contain all possible hypotheses but typically contains a reduced set as generated from an earlier Index stage as depicted in Fig. 1. From the image chip we extract a feature vector Y , and from each candidate hypothesis $H \in \mathcal{H}$ we generate a predicted feature vector X , where

$$X = [X_1, X_2, \dots, X_m]^T \quad Y = [Y_1, Y_2, \dots, Y_n]^T \quad (8)$$

and where m and n are the number of predicted and extracted features, respectively. Each feature X_i and Y_j is an ordered vector of feature attributes; for example, these attributes can be $(x, y, A, \alpha, L, \bar{\phi}, \gamma)$ parameters from the model in (1). However, the features themselves are unordered. In addition, there is uncertainty in both the predicted and extracted features.

There are two hypothesis testing goals that may be of interest. First, we may wish to classify the extracted feature vector Y as a measurement of one of the M class hypotheses. Second, we may wish to classify Y as one of the K class-subclass hypotheses in

the set \mathcal{H} . For both cases, we adopt a maximum *a posteriori* probability (MAP) rule; thus we must find the posterior likelihoods

$$\Lambda_{ij} = P(H_{ij} | Y), \quad H_{ij} \in \mathcal{H}. \quad (9)$$

If our goal is to classify Y as one of the K Index hypotheses (which include both class and subclass information), we choose the hypothesis that corresponds to the maximum Λ_{ij} . If our goal is to classify Y as one of the M class hypotheses, we form

$$\tilde{\Lambda}_i = \sum_j \Lambda_{ij}, \quad 1 \leq i \leq M \quad (10)$$

and choose the class i corresponding to the maximum $\tilde{\Lambda}_i$.

The above formulation gives an interpretation of the Index block in Fig. 1 as modifying the prior probabilities of the class and subclass hypotheses. The optimal MAP classifier maximizes or sums over all possible H_{ij} classes, and not just those provided by the Index stage. The Index stage computes a statistic $Z = g(U)$ from the image U , and essentially updates probabilities of hypotheses H_{ij} by finding posterior probabilities $P(H_{ij} | Z)$. A subset of hypotheses with sufficiently high posterior probabilities are retained for further processing. The final hypothesis test involves computing $P(H_{ij} | U)$; thus the feature-based match processing seeks to extract information in U not contained in $Z = g(U)$ to obtain a final classification decision. We see that the Index stage does not impact optimality in (9) provided the correct hypothesis is one of the K hypotheses passed. On the other hand, from (10) we see that the optimal MAP rule involves summation over all subclasses in class i , not just those passed by the Index stage. Thus (10) is optimal only under the stronger condition that the likelihoods $P(H_{ij} | Z)$ of all subclasses not passed by the Index stage are

equal to zero. In either case, the computational reduction of maximizing or summing on a reduced set of subclasses often justifies the deviation from optimality of the resulting classifier.

To compute the posterior likelihood in (9), we apply Bayes rule for any $H \in \mathcal{H}$ to obtain¹

$$\begin{aligned} P(H|Y) &= P(H|Y, n) = \frac{f(Y|H, n)P(H|n)}{f(Y|n)} \\ &= \frac{f(Y|H, n)P(n|H)P(H)}{f(Y|n)P(n)}. \end{aligned} \quad (11)$$

The conditioning on n is used because the number of features in Y is itself a random variable, but it suffices to consider only vectors Y of length n in the right-hand side of (11). Since the denominator of (11) does not depend on hypothesis H , the MAP decision is found by maximizing $f(Y|H, n)P(H)P(n|H)$ over $H \in \mathcal{H}$. The priors $P(H)$ and $P(n|H)$ are assumed to be known, or are provided by the Index stage.

The determination of $f(Y|H, n)$ includes both prediction and extraction uncertainties, which are related in the following way. Assume we have an object in the field with feature vector \hat{X} . We measure that object with a sensor, and obtain a feature vector Y . The measured feature vector differs from \hat{X} due to noise, sensor limitations, etc. We write this difference notionally as $Y = \hat{X} + N_e$ where N_e is some extraction error described by an uncertainty probability density function (pdf) $f(Y|\hat{X})$. In addition, if we suppose a hypothesis H , we can predict a feature vector X that differs from \hat{X} because of electromagnetic modeling errors, quantization errors of the assumed object subclass states (e.g., pose angle quantization errors), and differences between the actual object in the field and the nominal object that is modeled. We express this difference as $\hat{X} = X + N_p$ where N_p is a prediction error which we describe with an uncertainty $f(\hat{X}|H)$. Note that X is completely determined from H . To find the conditional uncertainty of Y given hypothesis H , we have

$$f(Y|H, n) = \int f(Y|\hat{X}, H, n)f(\hat{X}|H, n)d\hat{X} \quad (12)$$

where $f(\hat{X}|H, n)$ models the predict uncertainty, and $f(Y|\hat{X}, H, n)$ models extract uncertainty. The computation of $f(Y|\hat{X}, H, n)$ is complicated by the fact that the features in the Y and \hat{X} vectors are unordered, so a correspondence between the elements of Y and \hat{X} , or equivalently between Y and X , is needed.

B. Feature Correspondence

Computing the likelihood $f(Y|\hat{X}, H, n)$ requires that we form a correspondence map Γ between extracted and predicted features. The correspondence map is a nuisance parameter that arises because an extracted feature vector is not ordered with respect to the predicted feature vector. The correspondence also accounts for extracted features that are not in the predicted vector (false alarms) as well as predicted features that are not extracted (missed features).

For general pattern matching applications, a predicted feature X_i may correspond to none, one, or several extracted features

¹For notational simplicity, we drop the subscripts on the hypotheses in the sequel, and consider a general $H \in \mathcal{H}$.

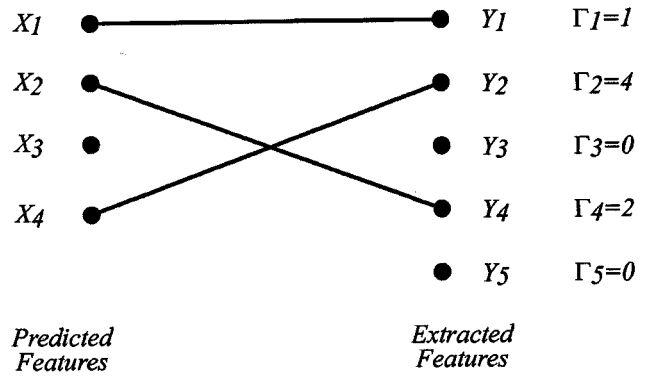


Fig. 5. An example one-to-one correspondence mapping for $m = 4$ and $n = 5$. Extracted features Y_3 and Y_5 are false alarms, and predicted feature X_3 is missed.

Y_j . Further, an extracted feature may correspond to one or more predicted features, or be a false alarm. We denote by \mathcal{G} the set of all such correspondence maps. For specific applications, a smaller set $\mathcal{G}_A \subset \mathcal{G}$ of admissible correspondence maps need only be considered. For example, in the SAR classification application presented in this paper, \mathcal{G}_A is the set of all one-to-one maps, where a one-to-one map corresponds to at most one predicted feature with each extracted feature and conversely. An example one-to-one correspondence map is shown in Fig. 5. Following [20], we let $\Gamma_j = i$ denote a correspondence between the X_i and Y_j . For notational conciseness, we write $\Gamma_j = 0$ to denote that Y_j does not correspond to any X_i , and therefore is a false-alarm feature.

We consider two correspondence mappings, random and deterministic. These two correspondence mappings lead to two different expressions of the posterior likelihoods Λ_{ij} .

Probabilistic Correspondence: If we assume a probabilistic correspondence model, then we have the Bayes likelihood

$$f(Y|H, n) = \sum_{\Gamma \in \mathcal{G}_A} f(Y|\Gamma, H, n)P(\Gamma|H, n) \quad (13)$$

where, similarly to (12),

$$f(Y|\Gamma, H, n) = \int f(Y|\hat{X}, \Gamma, H, n)f(\hat{X}|\Gamma, H, n)d\hat{X}. \quad (14)$$

The conditioning on n , the number of extracted features, is needed above because Γ is a correspondence between m predicted features and n extracted features; without the conditioning on n , $P(\Gamma|H, n)$ cannot be computed independently of Y .

The main difficulties in implementing (13) are: 1) knowledge of the priors $P(\Gamma|H, n)$ and 2) the high computational cost of summing over all possible correspondences. The correspondence prior probabilities can, in principle, be determined from knowledge of the predict and extract uncertainties for each hypothesis, but the derivation is quite difficult for many applications. More importantly, the summation contains a (very) large number of components; for example, there are more than $\min(m!, n!)$ possible one-to-one maps from a set of m predicted to n extracted features.

One can simplify computation of the large sum in (13) with assumptions of equal priors on Γ and independence of features

[20], [21]. If the priors are not equal or the features are not independent, then the resulting classifier will be suboptimal. It is difficult to predict the performance loss due to mismatch between the assumed and actual priors.

Deterministic Unknown Correspondence: If we assume the correspondence is deterministic but unknown, then it becomes a nuisance parameter in the classification. In this case, no uniformly most powerful test exists. We thus resort to the Generalized Likelihood Ratio Test (GLRT) classifier, in which we estimate Γ , then use the estimated Γ to estimate the likelihoods $f(Y | \hat{X}, H, n)$

$$f(Y | H, n) \approx \max_{\Gamma \in \mathcal{G}_A} f(Y | \Gamma, H, n) \quad (15)$$

where $f(Y | \Gamma, H, n)$ is computed using (14). The GLRT approach in (15) avoids the summation in (13), but requires a search for the best correspondence. Graph matching algorithms [22] can be used to simplify this search.

Discussion: The GLRT estimate of the conditional likelihood is a good estimate of the Bayes likelihood if the “best” correspondence term (15) dominates the sum in (13). This happens in the SAR classification problem, for example, when the feature uncertainties are small compared to their feature distances; for example, the match likelihood when corresponding two features with widely differing (x, y) locations is negligibly small compared to the likelihood found from associating all pairs of features with similar (x, y) locations. The presence of additional scattering attributes helps increase the feature distances even for scattering features that have similar locations; for example, two physically close scattering centers with different α and L parameters have lower likelihood of an incorrect match pairing than they would if match scores were based only on scatterer location and amplitude.

For the SAR classification application, we adopt both a one-to-one map and a GLRT classifier. The one-to-one map makes physical sense: an extracted scattering center corresponds to at most one predicted scattering center, and conversely. The GLRT classifier assumes a deterministic but unknown correspondence map, and avoids summation over a large set of possible correspondence maps. The probabilistic map assumption for this model is considered in [21]. In addition, [20] considers other classifiers derived for SAR features using only location attributes.

C. Conditional Feature Likelihood

To implement either (13) or (15), one must have available a model for $f(Y | \hat{X}, \Gamma, H, n)$. In this section we develop a model based on [20] that applies to SAR scattering center features.

We assume the X_i are conditionally independent given H , and that Y_j are conditionally independent given H , \hat{X} , and n . The independence of the X_i is reasonable because the prediction errors of separate scattering centers are due to variations in components on the target that make up that scattering center, and these variations can be assumed to be unconnected. The independence of the Y_i is supported by the near block diagonality of the CRB matrix for well-separated scattering cen-

ters [9]. In addition, the independence assumptions simplify the Bayes matcher significantly. Thus we have

$$f(Y | \Gamma, H, n) = \prod_{j=1}^n f(Y_j | \Gamma, H, n). \quad (16)$$

For a one-to-one correspondence, the j th extracted feature corresponds either to a particular predicted feature (say, the i th one), or to a false alarm. We denote these two cases as $\Gamma_j = i$ or $\Gamma_j = 0$, respectively. Thus for a given correspondence, there may be some predict-extract feature correspondences, some missed predicted features (which correspond to no extracted feature), and some false-alarm extracted features which have no corresponding predict feature; see Fig. 5. For a given correspondence, let n_F denote the number of false-alarm features.

We model the conditional feature likelihood as

$$\begin{aligned} f(Y | \Gamma, H, n) &= P(n_F \text{ false alarms}) \prod_{\{j: \Gamma_j=0\}} f_{FA}(Y_j) \\ &\cdot \prod_{\{j: \Gamma_j=i>0\}} P_i(H) f(Y_j | \Gamma_j = i, H, n) \\ &\cdot \prod_{\{i: \Gamma_j \neq i, \forall j\}} (1 - P_i(H)) \end{aligned} \quad (17)$$

where $P_i(H)$ is the detection probability of the i th predicted feature under hypothesis H . The first term on the right-hand side models the likelihood of false-alarm features, and $f_{FA}(Y_j)$ is the pdf of feature Y_j if it corresponds to a false alarm. The second line is the likelihood of extracted features that correspond to predicted features, and the third line represents the miss probabilities for predicted features that have no corresponding extract feature.

D. Implementation of the Correspondence Search

The GLRT hypothesis selection rule in (15) involves finding the correspondence Γ that maximizes $f(Y | \Gamma, H, n)$ in (17) for each candidate hypothesis $H \in \mathcal{H}$. In general, the search is computationally intensive, but for some cases can be implemented efficiently. Specifically, in the case that

$$P(n_F \text{ false alarms}) = c\beta^{n_F} \quad (18)$$

for some constants c and β , the search can be efficiently implemented in $O((m+n)^3)$ operations using the Hungarian algorithm [22].

We briefly summarize the implementation of the Hungarian algorithm for this problem.² From (17) and (18) we have

$$\begin{aligned} -\log f(Y | X, \Gamma, H, n) &= -\log c - \sum_{\{j: \Gamma_j=0\}} \log[\beta f_{FA}(Y_j)] \\ &- \sum_{\{j: \Gamma_j=i>0\}} \log[P_i(H) f(Y_j | \Gamma_j = i, H, n)] \\ &- \sum_{\{i: \Gamma_j \neq i, \forall j\}} \log[1 - P_i(H)]. \end{aligned} \quad (19)$$

²The authors thank William Irving for noting the application of the Hungarian algorithm to this search problem.

	Y_1	\cdots	Y_n	misses	
X_1	c_{11}	\cdots	c_{1n}	M_1	∞
\vdots	\vdots	\ddots	\vdots	\ddots	
X_m	c_{m1}	\cdots	c_{mn}	∞	M_m
false	F_1		∞	0	\cdots 0
alarms		\ddots		\vdots	\ddots
	∞		F_n	0	\cdots 0

Fig. 6. The cost matrix for the one-to-one matcher in (17). Here, $c_{ij} = -\log[P_i(H)f(Y_j|\Gamma_j = i, H)]$, $F_j = -\log[\beta_{FA}(Y_j)]$, and $M_k = -\log[1 - P_k(H)]$.

We insert the elements of the above equation for all possible i and j into an $(m+n) \times (m+n)$ array, as shown in Fig. 6. Then the minimum of (19) over all one-to-one maps reduces to the problem of selecting exactly one element from each row and column of the array such that the sum of the selected entries is minimized. The resulting solution also gives the optimal correspondence. Specifically, if c_{ij} is selected, then predicted feature X_i corresponds to extracted feature Y_j ; if F_j is selected, then Y_j is a false-alarm feature; if M_i is selected, then X_i is not present (missed) in the extracted features Y .

The search is equivalent to finding a permutation of the cost matrix that minimizes its trace. Such a permutation is found efficiently using the Hungarian algorithm [22]. A related algorithm [23] can also find the k permutations that give the k smallest trace values, which is useful if the “best” k correspondences are of interest.

As an alternative, geometric hashing [24], [25] can be used to efficiently search a set of candidate hypotheses for the highest likelihood match. Hashing methods precompute information about patterns of features in a hash table that later can be efficiently searched to vote for hypotheses that are close matches. On the other hand, hashing requires the formation of a large table, containing entries for every hypothesis H_{ij} ; this table can be prohibitively large for high-dimensional classification applications.

IV. BAYES CLASSIFICATION EXAMPLE

In this section we present an example of feature-based classification using SAR scattering center attributes. We use synthetic feature vector means based on measured SAR imagery and an assumed feature perturbation model. We select target classes, feature sets, feature attribute uncertainties, and priors to be representative of a realistic X -band SAR target-recognition problem. The synthetic data results serve to emphasize, by example, that the Bayes classifier is tractable for problem sizes encountered in SAR target recognition given the assumption of conditionally independent features in Section III-C. The proposed technique permits estimation of the optimal error rate given a set of assumed priors and feature uncertainties. In addition, we demonstrate by example that the Bayes classifier can be used to explore the performance effects of sensor parameters. Finally, the Bayes classifier can be used as a simulation tool

to investigate sensitivity of the estimated error rate to errors in assumed priors and feature attribute uncertainty. Accurate prediction of absolute classification performance would require an electromagnetic prediction module as in Fig. 1 and extraction uncertainties empirically verified from ground truth; neither is presently available.

We generate class means using a combination of synthetic generation and feature extraction from SAR imagery. We synthesize 2747 class mean features for ten composite target classes in the MSTAR Public Targets dataset [26]. The data set contains X -band image chips with 128×128 pixels and $1 \text{ ft} \times 1 \text{ ft}$ resolution SAR data chips of 10 targets at 17° depression angle. For each target, approximately 270 images are available covering the full 360° aspect angles, for a total of 2747 images. Down-range and cross-range locations and amplitudes of scattering centers are synthesized from local maxima on the image chips. The targets are the 2S1, BMP-2, BRDM-2, BTR-70, BTR-60, D-7, T-62, T-72, ZIL-131, and ZSU-23-4. Examples of the SAR image chips are shown in Fig. 7. From each image we extract locations and amplitudes of scattering centers from local maxima in the SAR image. The α and L parameters are not provided by current prediction modules, so are generated synthetically. The α attribute for each feature is generated as $\alpha \sim \mathcal{N}(0.5, 0.25)$. The Gaussian approximation to a discrete variable is used to avoid the combinatorial number of likelihood evaluations for all possible α choices from prediction and extraction; experiments verify that the Gaussian approximation gives very similar results at lower computational cost [21]. The length parameter is quantized to 1 bit ($L = 0$ or $L > 0$), and the nominal values of the length attribute are generated using a Bernoulli random variable with $P(L > 0) = 0.3$. We quantize L because existing electromagnetic prediction codes do not provide L ; further, prediction uncertainty is unknown, so we choose to adopt only coarse uncertainty assumptions in the simulation. The γ and ϕ parameters in (1) are not used in the experiments because no strong evidence exists that these parameters can be predicted and extracted at 1-ft SAR resolution with sufficiently low uncertainty to substantially improve classification performance; nonetheless, these two parameters are retained in (1) both for reduced bias and for application to higher resolution SAR imagery. For example, using 2-in resolution SAR imagery at X -band, pose angle can be estimated to an accuracy that is a small fraction of the 20° phase history angle used in the image formation.

Prediction and extraction feature uncertainties are needed in the Bayes classifier. We evaluate $f(Y_j|\Gamma_j = i, H, n)$ in (17) as follows. Recall each Y_j is a feature vector $[Y_{j1}, \dots, Y_{js}]$, where the Y_{jk} are the individual feature attributes $x_j, y_j, A_j, \alpha_j, L_j, \phi_j$, and γ_j from the model in (1) (or a subset of these). For simplicity, we assume the uncertainties of the feature attributes are independent; experiments using dependent attribute uncertainties are presented in [21]. For independent attributes

$$f(Y_j|\Gamma_j = i, H, n) = \prod_k f(Y_{jk}|\Gamma_j = i, H, n) \quad (20)$$

where k denotes an index on the feature attributes. We further model each attribute k in an extracted feature Y_j that corresponds to predicted feature X_i with a conditional likelihood

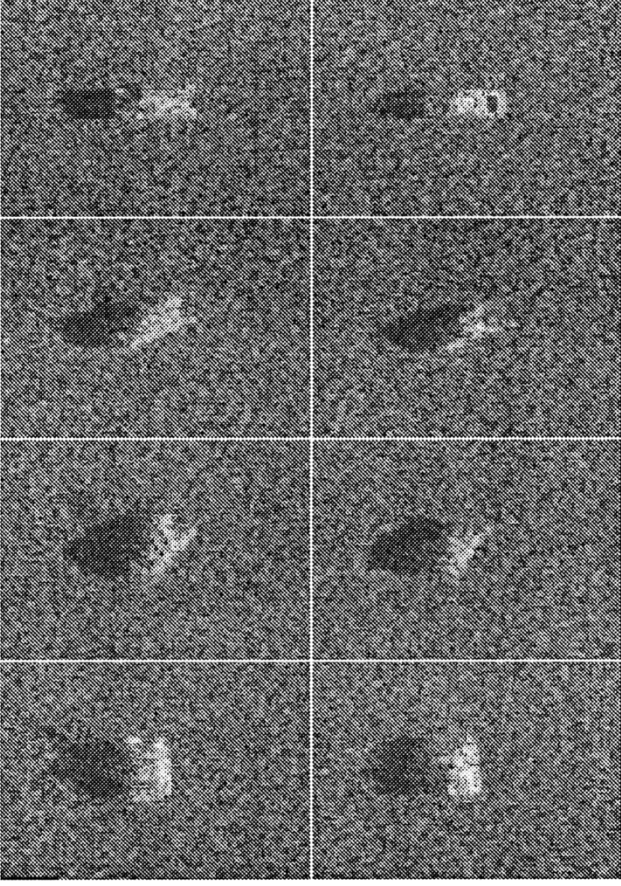


Fig. 7. Examples of the MSTAR SAR image chips used in the Bayes classification example. Four T-72 (left) and BMP-2 (right) images are shown.

$f(Y_{jk} | \hat{X}_{ik}, H, n)$ as follows. If k corresponds to $x, y, \log|A|$, or α , the conditional likelihood is assumed to be Gaussian

$$f(Y_{jk} | \hat{X}_{ik}, H, n) \sim \mathcal{N}(\hat{X}_{ik}, \sigma_{e,k}^2). \quad (21)$$

We assume a similar predict uncertainty for each attribute

$$f(\hat{X}_{ik} | X_{ik}, H, n) \sim \mathcal{N}(X_{ik}, \sigma_{p,k}^2). \quad (22)$$

Thus from (14), (21), and (22) we have

$$f(Y_{jk} | X_{ik}, H, n) \sim \mathcal{N}(X_{ik}, \sigma_{p,k}^2 + \sigma_{e,k}^2) \quad (23)$$

which gives the needed terms in (20). Similarly, for a discrete attribute (the quantized length L) the likelihood is a weighted sum

$$P(Y_{jk} | X_{ik}, H, n) = \sum_{\hat{X}_{ik}} P(Y_{jk} | \hat{X}_{ik}, H, n) P(\hat{X}_{ik} | X_{ik}, H, n) \quad (24)$$

and is thus described by probability mass functions on the predicted features along with predict and extract confusion matrices.

From (23) and (24) we see that only the sum of prediction and extraction uncertainties is needed. Table I lists the uncertainty values used in the simulations. We assume no prediction uncertainty in α or L , and log-normal uncertainty in $|A|$. The total location uncertainty is assumed to have a standard deviation of one Rayleigh resolution for both x and y . In addition, Table II specifies the false-alarm pdf $f_{FA}(Y_j)$.

We emulate the Index stage as follows. For each of the 2747 target image chips, we find the five image chips in each of the ten target classes that have the highest correlation. The target classes and poses (pose is in this case azimuth angle) corresponding to these 50 image chips form the initial hypothesis list generated by the Index stage. For each class mean vector, we generate a predict feature vector for each of the 50 hypotheses from the Index stage by randomly perturbing the mean vector using the predict uncertainty model above. We similarly generate an extracted feature vector from the mean vector. The extracted feature vector assumes each scattering center has a probability of detection of $P_d = 0.5$, so not all scattering centers are present in the extracted feature vector. We also add clutter scattering centers to the extract feature vector. We then compute the GLRT hypothesis test using (11), (15), and (17), assuming equally likely priors ($P(H)P(H|n) = \text{constant}$) on the 50 Index hypotheses. We use the Hungarian algorithm to search for the best correspondence map, using $\beta = \lambda$ in (18). We record the target class corresponding to the one of the 50 hypotheses with the highest likelihood score. We repeat this experiment ten times for each class mean vector; this gives a total of 27 470 classifications from $27\,470 \times 50$ matches. For each candidate hypothesis, computation of the correspondence is $O(m^3)$, where m is the number of predicted features. Execution times for the 50 likelihood computations average 4.6 s using unoptimized Matlab code on a 333-MHz Pentium processor.

Table III presents the results of the above experiment for a SAR Rayleigh resolution of 1 ft, using the uncertainty values in Table I. We summarize the overall performance as an average probability of correct classification P_c , which is 86.8% for this case.

Fig. 8 presents probability of correct classification results as a function of both the number of feature attributes and the system bandwidth. First we compare the use of location features with location features coupled with other attributes. The amplitude attribute provides only modest improvement (1–2 dB) in the probability of error, due to its relatively high uncertainty. The addition of the frequency dependence and length attributes provides more significant improvement in classification performance, especially for the higher resolutions considered. The amount of improvement depends critically on the assumed attribute uncertainty and its correlation with other attributes.

Fig. 8 also presents results of an experiment in which we predict classification performance as a radar system parameter, namely, Rayleigh resolution, changes. The bandwidths and integration angles correspond approximately to SAR image Rayleigh resolutions of 2, 1, 1/2, and 1/4 ft. We assume decreasing uncertainty in the location, frequency dependence, and length attributes as Rayleigh resolution becomes finer, as shown in Table I. From Fig. 8 we see that classification performance improves significantly as radar bandwidth and integration angle increase. Specifically, the error probability $1 - P_c$ decreases by about 15 dB as the SAR Rayleigh resolution improves from 2 ft to 1/4 ft. Here we see a clear benefit of increased bandwidth because it results in decreased feature uncertainty.

Fig. 9 shows the effect on classification performance when the assumed uncertainty model in the Bayes classifier is in error. In this experiment we set the location uncertainty standard de-

TABLE I
SUM OF PREDICTION AND EXTRACTION FEATURE ATTRIBUTE UNCERTAINTIES USED IN THE BAYES CLASSIFIER EXAMPLE

Feature Attribute	SAR Rayleigh Resolution			
	2 ft	1 ft	1/2 ft	1/4 ft
locations x, y $\mathcal{N}(0, \sigma^2)$	$\sigma = 2$ ft	$\sigma = 1$ ft	$\sigma^2 = 1/2$ ft	$\sigma = 1/4$ ft
amplitude $\log_{10}(A) \sim \mathcal{N}(0, \sigma^2)$	$\sigma^2 = 0.5$	$\sigma^2 = 0.5$	$\sigma^2 = 0.5$	$\sigma^2 = 0.5$
frequency dependence $\alpha \sim \mathcal{N}(0, \sigma_\alpha^2)$	$\sigma_\alpha = 1$	$\sigma_\alpha = 1/2$	$\sigma_\alpha = 1/4$	$\sigma_\alpha = 1/8$
length $\begin{bmatrix} P(L=0 L=0) & P(L=0 L>0) \\ P(L>0 L=0) & P(L>0 L>0) \end{bmatrix}$	$\begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$	$\begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$	$\begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$	$\begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$

TABLE II
FALSE-ALARM pdf $f_{FA}(Y_j)$ USED IN THE BAYES CLASSIFIER EXAMPLE

Feature Attribute	Feature pdf used
number	Poisson with rate $\lambda = 3$ per image chip $P(n_F \text{ false alarms}) = e^{-\lambda} \lambda^{n_F} / (n_F!)$
locations x, y	uniformly distributed over the image
amplitude	$\log_{10}(A) \sim \mathcal{N}(\mu, 0.25)$ $\mu = \log_{10}(\text{median amplitude of predicted scattering centers})$
curvature	$\alpha \sim \mathcal{N}(0.5, 1)$
length	L is Bernoulli with $P(L > 0) = 0.3$

TABLE III
ONE-TO-ONE CLASSIFICATION RESULTS USING FIVE FEATURE ATTRIBUTES. THE ij TH ENTRY GIVES THE NUMBER OF TIMES THE OBJECT WAS CLASSIFIED AS OBJECT j GIVEN THAT OBJECT i IS THE TRUE OBJECT. OVERALL $P_c = 86.8\%$
Optimal One-to-One Map, five feature attributes, $P_d = 0.5$

	2S1	BMP	BRDM	BTR	BTR	D-7	T-62	T-72	ZIL	ZSU	Total
		2	2	70	60				131	23-4	
2S1	2574	42	33	48	43	38	72	52	56	32	2990
BMP-2	42	2023	55	37	29	29	30	31	29	25	2330
BRDM-2	60	41	2554	71	39	39	50	40	29	57	2980
BTR-70	29	32	46	2046	33	20	34	34	37	19	2330
BTR-60	45	28	38	36	2280	23	26	27	32	25	2560
D-7	49	35	39	13	25	2639	68	37	36	49	2990
T-62	51	44	58	31	36	37	2584	53	46	50	2990
T-72	36	43	38	36	32	20	44	1981	47	43	2320
ZIL-131	55	53	34	45	44	30	44	38	2614	33	2990
ZSU-23-4	37	61	52	39	33	63	60	57	45	2543	2990

viation to 0.5, 1, and 2 times the correct location uncertainty; the other attributes use the correct uncertainty models. Here, $P_i(H) = 0.9$. We see the correct classification rate drops by 10%–20% as a result of the mismatch, and that a greater performance loss occurs when the model-based classifier assumes too low an uncertainty.

V. CONCLUSION

We have presented a model-based framework for image processing when the processing goal is object classification. The Bayesian formalism allows clear and explicit disclosure of all assumptions, in contrast to *ad hoc* classification procedures.

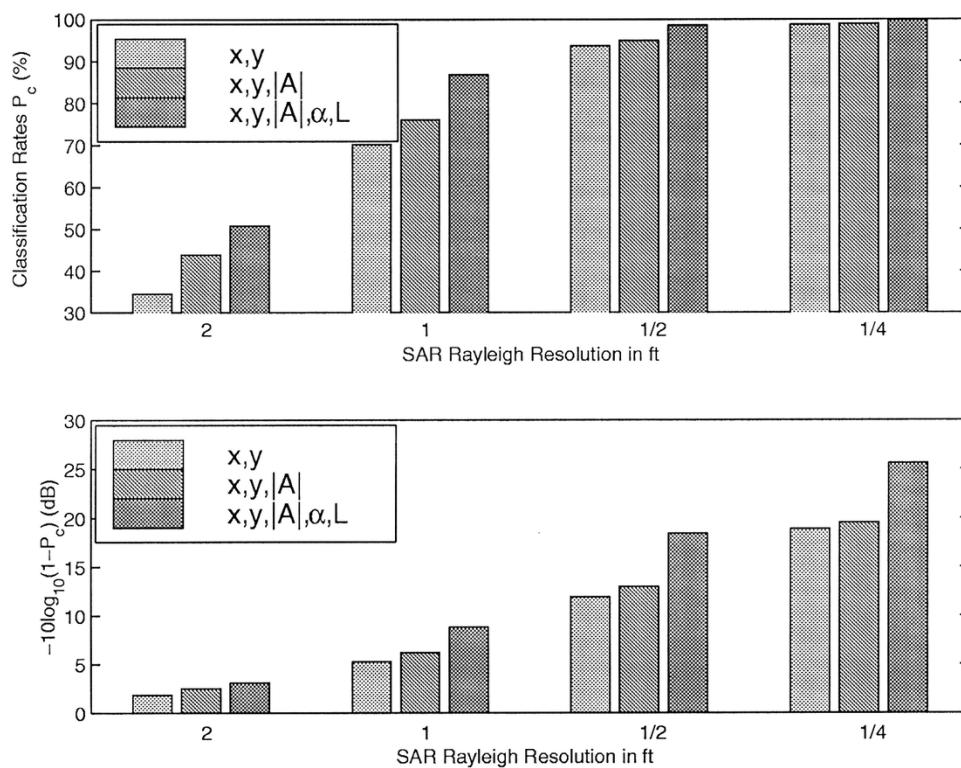


Fig. 8. Classification performance as a function of number of feature attributes and radar bandwidth. The top graph shows average probability of correct classification (P_c); the bottom graph shows the same data plotted as average probability of error ($1 - P_c$) in decibels.

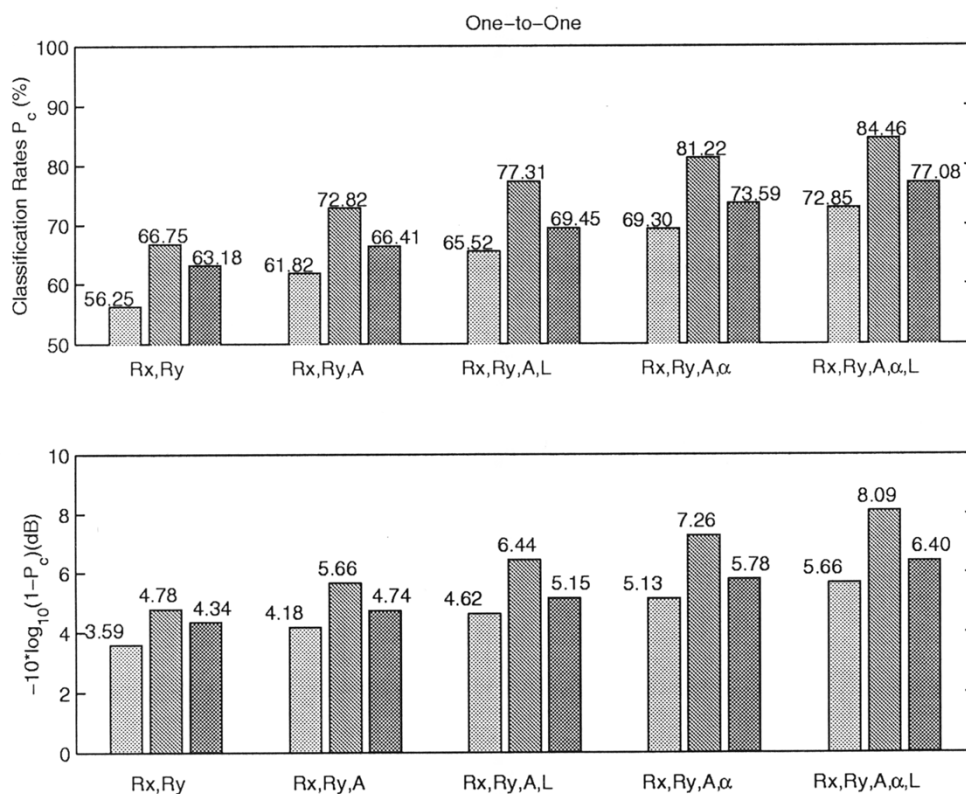


Fig. 9. Classification performance using correct (center) and erroneous location uncertainties in the Bayes classifier. The left (right) bars assume 0.5 (2) times the true location uncertainty. The top graph shows average probability of correct classification (P_c); the bottom graph shows the same data plotted as average probability of error ($1 - P_c$) in decibels.

Moreover, we demonstrated that the Bayes approach, including the associated correspondence problem, is tractable and leads to implementable algorithms.

We have presented the Bayes approach to model-based classification in the application context of synthetic aperture radar (SAR) imagery. By modeling electromagnetic scattering behavior and estimating physically meaningful parameters from complex-valued imagery, we computed features as statistics for use in hypothesis testing. For radar systems with a significant fractional bandwidth, the features provide richer information than local peaks in magnitude imagery.

A complete empirical evaluation of the proposed classifier requires an electromagnetic scattering code to provide predicted features conditioned on a target hypothesis; at the time of this publication, such a code is being developed by Veridian-ERIM International, Inc., as a hybrid combination of the ray-tracing and scattering primitive codes. Further, the efficacy of the proposed likelihood estimation technique requires additional empirical verification of the feature uncertainties adopted in Section IV; to do so requires ground truth that is not currently available, but would be provided by the scattering prediction code under development.

The implementable Bayes classifier allows estimation of optimal error rates, given assumed priors and feature uncertainties, and the simulation of performance sensitivity to assumed priors, to assumed feature uncertainties, and to sensor characteristics.

ACKNOWLEDGMENT

The authors wish to thank William Irving and Edmund Zelnio for informative discussions on correspondences of random feature vectors, staged hypothesis evaluation, and application to microwave radar images.

REFERENCES

- [1] J. C. Curlander and R. N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*. New York: Wiley, 1991.
- [2] E. R. Keydel, S. W. Lee, and J. T. Moore, "MSTAR extended operating conditions: A tutorial," in *Algorithms for Synthetic Aperture Radar Imagery III (Proc. SPIE)*, E. Zelnio and R. Douglass, Eds., no. 2757, 1996, pp. 228–242.
- [3] T. W. Ryan and B. Egaas, "SAR target indexing with hierarchical distance transform," in *Algorithms for Synthetic Aperture Radar Imagery III (Proc. SPIE)*, E. Zelnio and R. Douglass, Eds., no. 2757, 1996, pp. 243–252.
- [4] D. F. DeLong and G. R. Benitz, "Extensions of high definition imaging," in *Algorithms for Synthetic Aperture Radar Imagery II (Proc. SPIE)*, D. A. Giglio, Ed., no. 2487, 1995, pp. 165–180.
- [5] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Processing*, vol. 44, pp. 1469–1484, June 1996.
- [6] J. Wissinger, R. Washburn, D. Morgan, C. Chong, N. Friedland, A. Nowicki, and R. Fung, "Search algorithms for model-based SAR ATR," in *Algorithms for Synthetic Aperture Radar Imagery III (Proc. SPIE)*, E. Zelnio and R. Douglass, Eds., no. 2757, 1996, pp. 279–293.
- [7] E. R. Keydel and S. W. Lee, "Signature prediction for model-based automatic target recognition," in *Algorithms for Synthetic Aperture Radar Imagery III (Proc. SPIE)*, E. Zelnio and R. Douglass, Eds., no. 2757, 1996, pp. 306–317.
- [8] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Phil. Trans. Roy. Soc. London*, vol. A222, pp. 309–368, 1922.
- [9] M. J. Gerry, L. C. Potter, I. J. Gupta, and A. van der Merwe, "A parametric model for synthetic aperture radar measurements," *IEEE Trans. Antennas Propagat.*, vol. 47, pp. 1179–1188, July 1999.
- [10] L. C. Potter, D.-M. Chiang, R. Carrière, and M. J. Gerry, "A GTD-based parametric model for radar scattering," *IEEE Trans. Antennas Propagat.*, vol. 43, pp. 1058–1067, Oct. 1995.
- [11] L. C. Potter and R. L. Moses, "Attributed scattering centers for SAR ATR," *IEEE Trans. Image Processing*, vol. 6, pp. 79–91, Jan. 1997.
- [12] J. B. Keller, "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, vol. 52, pp. 116–130, 1962.
- [13] R. G. Kouyoumjian and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," *Proc. IEEE*, vol. 62, pp. 1448–1461, Nov. 1974.
- [14] M. Koets and R. Moses, "Image domain feature extraction from synthetic aperture imagery," in *Proc. ICASSP*, paper 2438, Phoenix, AZ, Mar. 1999.
- [15] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica*, vol. 50, pp. 1–25, 1982.
- [16] H. L. Van Trees, *Detection, Estimation, and Modulation Theory Part I*. New York: Wiley, 1968.
- [17] J. A. O'Sullivan, R. E. Blahut, and D. L. Snyder, "Information-theoretic image formation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2094–2123, Oct. 1998.
- [18] M. P. Clark, "On the resolvability of normally distributed vector parameter estimates," *IEEE Trans. Signal Processing*, vol. 43, pp. 2975–2981, Dec. 1995.
- [19] S. P. Bruzzone and M. Kaveh, "A criterion for selecting information-preserving data reductions for use in the design of multiple parameter estimators," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 466–470, May 1983.
- [20] G. Ettinger, G. A. Klanderman, W. M. Wells, and W. Grimson, "A probabilistic optimization approach to SAR feature matching," in *Algorithms for Synthetic Aperture Radar Imagery III (Proc. SPIE)*, E. Zelnio and R. Douglass, Eds., no. 2757, Apr. 1996, pp. 318–329.
- [21] H.-C. Chiang, "Feature-based classification with application to synthetic aperture radar," Ph.D. dissertation, Ohio State Univ., Columbus, OH, 1999.
- [22] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization Algorithms and Complexity*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [23] M. L. Miller, H. S. Stone, and I. J. Cox, "Optimizing Murty's ranked assignment method," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 851–861, July 1997.
- [24] Y. Lamdan, J. T. Schwartz, and H. Wolfson, "Affine invariant model-based object recognition," *IEEE Trans. Robotics Automat.*, vol. 6, pp. 578–589, Oct. 1990.
- [25] I. Rigoutsos and R. Hummel, "A Bayesian approach to model matching with geometric hashing," *Comput. Vision Image Understanding*, vol. 62, pp. 11–26, July 1995.
- [26] MSTAR (Public) Targets: T-72, BMP-2, BTR-70, SLICY. [Online]. Available: <http://www.mbvlab.wpafb.af.mil/public/MBVDATA>