Classification Performance Prediction Using Parametric Scattering Feature Models

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

Department of Electrical Engineering
The Ohio State University
Columbus, OH 43210-1272, USA

ABSTRACT

We consider a method for estimating classification performance of a model-based synthetic aperture radar (SAR) automatic target recognition system. Target classification is performed by comparing an unordered feature set extracted from a measured SAR image chip with an unordered feature set predicted from a hypothesized target class and pose. A Bayes likelihood metric that incorporates uncertainty in both the predicted and extracted feature vectors is used to compute the match score. Evaluation of the match likelihoods requires a correspondence between the unordered predicted and extracted feature sets. This is a bipartite graph matching problem with insertions and deletions; we show that the optimal match can be found in polynomial time.

We extend the results in[1] to estimate classification performance for a ten-class SAR ATR problem. We consider a synthetic classification problem to validate the classifier and to address resolution and robustness questions in the likelihood scoring method. Specifically, we consider performance versus SAR resolution, performance degradation due to mismatch between the assumed and actual feature statistics, and performance impact of correlated feature attributes.

Keywords: synthetic aperture radar; model-based target recognition; scattering centers; structural matching; Bayes classification; performance estimation

1. INTRODUCTION

Synthetic Apterture Radar (SAR) classification problems are often characterized by the combined complexity of a high-dimensional observation space and a large set of multi-modal candidate hypotheses. Pattern recognition from measured imagery is characterized by a high dimensional observation space. A typical image may comprise a 256×256 array of pixels, yielding an observation vector in R^N, where $N = 2^{16}$. For both computational simplicity and performance robustness, feature extraction is used to reduced the data to lower dimension. In addition, pattern recognition problems in general, and SAR ATR problems in particular, are characterized by a complex hypothesis space. The hypothesis set consists of M classes, or objects. The complexity arises in that each object may be observed in a variety of poses, configurations and environments, thereby resulting in an intractable density function for the measurement conditioned on the object. The number of enumerated subclasses explodes exponentially; a typical application for $10 \leq M \leq 50$ object classes might dictate 10^{12} subclasses.[2]

To address hypothesis complexity, a staged, coarse-to-fine classification strategy is used to efficiently search the hypothesis space. In addition, a 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 measurement and storage of a prohibitively large catalog of image templates. The classification approach is summarized in Figure 1. A state of nature is characterized by the hypothesis of an object class or subclass, H_k, from which a measurement U is drawn. A Feature Extraction stage serves to reduce the dimensionality of the measured data; parameters are estimated from imagery and used as low-dimensional surrogates for sufficient statistics. The uncertainty in these parameters is given as a density function, $f(Y|U)$, and reflects the sensitivity of parameter estimates, Y, to noisy sensor data, given the measured data, U.

The complexity of the hypothesis space is addressed in a coarse-to-fine approach. An Index stage provides a list of candidate hypotheses, $H_k,\ k = 1, \ldots, K$, based on a coarse partitioning of the hypothesis space. Evaluation of

*This research was supported in part by DARPA and the US Air Force Research Laboratory under Grant F33615-979191020.

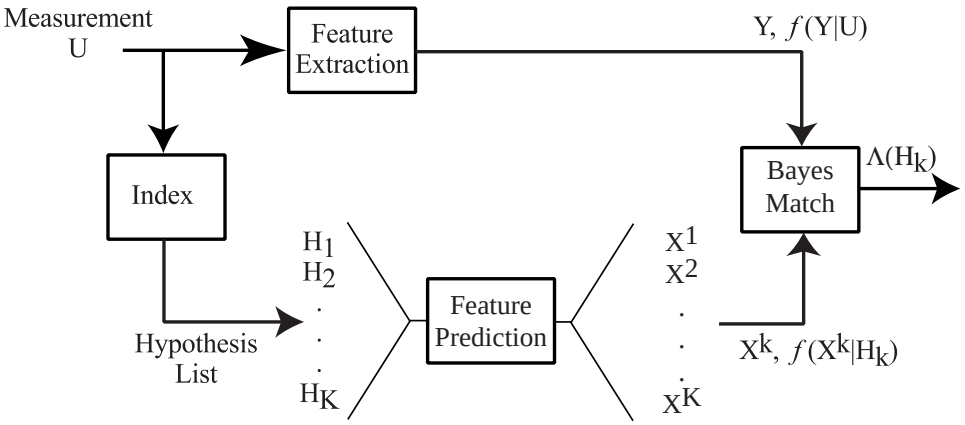

Figure 1. A staged approach to manage high-dimensional observations and complex hypothesis sets encountered in classification of radar images.

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_k. The predicted feature set, X^k, has an associated uncertainty $f(X|H_k)$ acknowledging error in the modeling and variation among objects in the subclass. The predicted and extracted feature sets are combined in a fine classification, or Bayes Match, stage to compute the posterior probability of a candidate hypothesis, $\Lambda(H_k)$. The top hypotheses, and their likelihoods, are reported as the output of the classification system. Computation of the likelihood scores requires a correspondence between extracted and predicted features and an integration over feature uncertainty.

Inexact feature matching, and the related inexact graph matching problem, have been the subject of much work in the pattern recognition community.[3–6] Recent work on Bayesian structural matching considers subgraph matches and edit distances.[5,6] These structural match problems typically have a large number of nodes that are not fully connected by edges, and the research emphasis has been on computationally efficient search solutions that exploit the edge structure. In contrast, we consider a problem in which the graphs are fully connected, and seek Bayes optimal matches that incorporate prior probabilities of node deletions and extraneous nodes.

This paper focuses on the application of model-based Bayesian classification approach to recognition of vehicles using synthetic aperture radar (SAR) images. We consider a model-based Bayesian matcher derived from the MSTAR matcher.[7–9] We adopt an electromagnetic scattering model for use in both the Feature Extraction and Feature Prediction stages. The feature set used for classification is based on an *attributed scattering center* model for target scattering.[10,11] We present synthetic classification performance predictions using class means estimated from measured X-band SAR imagery of ten vehicles. We also consider sensitivity of classification performance to the assumed feature uncertainties and priors. The results illustrate that the Bayes approach to model-based classification is tractable given priors and feature uncertainties. The examples also illustrate how one can use the Bayes classifier as a simulation tool to explore performance as a function of sensor parameters or to explore sensitivity of classification performance to assumed priors and feature uncertainties.

This paper extends earlier work[1] and uses the same feature set and classifiers. We briefly present the attributed scattering center features in Section 2 and the Bayes classifiers in Section 3. In Section 4 we present recent classification performance results using this scattering center model and classifier. Section 5 concludes the paper.

2. ATTRIBUTED SCATTERING CENTER MODEL

We adopt a parametric data model which is based on approximate radar scattering physics.[11] From the geometric theory of diffraction, 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 total

scattered field from a target is modeled as the sum of p individual scattering centers[11]

$$E(f,\phi) = \sum_{k=1}^{p} E_k(f,\phi) \tag{1}$$

where each scattering center is modeled as

$$E_k^s(f,\phi) = A_k \left(j\frac{f}{f_c}\right)^{\alpha_k} \exp\left(j\frac{4\pi f}{c}(Rx_k\cos\phi + Ry_k\sin\phi)\right) \operatorname{sinc}\left(\frac{2\pi f}{c}L_k\sin(\phi-\bar{\phi}_k)\right) \tag{2}$$

where f_c is the center frequency of the radar and c is the speed of propagation. Each of the p scattering centers is characterized by six attributes: (Rx_k, Ry_k) denote the scattering center location, A_k is the amplitude, L_k is a length, $\bar{\phi}_k$ is a pose angle, and the discrete parameter α_k characterizes curvature of the scattering center. The feature set $\{\theta_1,\ldots,\theta_p\}$, where

$$\theta_k = [Rx_k, Ry_k, A_k, \alpha_k, L_k, \bar{\phi}_k], \quad k = 1,...,p \tag{3}$$

forms the set of features used for classification. Note that the features θ_k are unordered, but the feature attributes (the elements of θ_k) are well-ordered.

The model in (1)–(2) is based on geometric theory of diffraction and physical optics approximations for scattering behavior and allows compression of the high-dimensional measured image into a low-dimensional feature space. Primitive scattering geometries, such as dihedrals, corner reflectors and cylinders, are distinguishable by their (α, L) parameters.[11]

3. BAYES CLASSIFICATION

3.1. Problem Statement

The Bayes matching problem is given as follows (see Figure 1). At the input to the classifier stage, we are given a vector of n features

$$Y = [Y_1, Y_2, \ldots, Y_n]^T \tag{4}$$

extracted from a measurement, U. Each feature Y_i is an $\ell \times 1$ vector of ordered attributes, either θ_i given by (3) or an ordered subset of the attributes. We are also given a set $\mathcal{H} = \{H_k,\ k \in [1\ldots K]\}$ of K candidate hypotheses, along with their prior probabilities $P(H_k)$. The set $\mathcal{H}$ typically is provided by the preceding Index stage, as depicted in Figure 1.

We assume available a feature prediction function which maps a hypothesis H_k to a set of m_k predicted features

$$X^k = [X_1^k, X_2^k, \ldots, X_{m_k}^k]^T \tag{5}$$

Each feature X_i^k is an ordered attribute vector whose attributes match those of the Y_i vectors. Finally, we assume a known probability model for the uncertainty of the attributes for any predicted or extracted feature, as well as models for appearance of a given feature in Y and for the feature attributes if Y_i is a false-alarm feature (*i.e.*, Y_i does not correspond to a predicted feature). We seek to compute the posterior likelihood of the observed features, Y, under each of the K hypotheses in the set $\mathcal{H}$.

$$\Lambda_k = P(H_k|Y), \quad H_k \in \mathcal{H} \tag{6}$$

The most likely hypotheses and their corresponding likelihoods are reported at the output of the classifier.

To compute the posterior likelihood in (6), we apply Bayes rule for any $H \in \mathcal{H}$ to obtain[†]

$$P(H|Y) = P(H|Y,n) = \frac{f(Y|H,n)P(H|n)}{f(Y|n)} = \frac{f(Y|H,n)P(n|H)P(H)}{f(Y|n)P(n)} \tag{7}$$

[†]We drop the subscript on the hypothesis H_k in the sequel, and consider a general $H \in \mathcal{H}$ Correspondingly, we drop the k on X^k and m_k in (5).

Since the denominator of (7) 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.

To compute $f(Y|H,n)$ we incorporate uncertainty in both the predicted and extracted feature sets. Assume the object being measured has a true 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 a feature extraction error described by a probability density function $f(Y|\hat{X})$. In addition, if we suppose a hypothesis H, we can predict a feature vector X that differs from $\hat{X}$ because of modeling errors. We express this difference as $\hat{X} = X + N_p$ where N_p is a prediction error with probability density function $f(\hat{X}|X) = f(\hat{X}|H)$. Note that X is completely determined from the hypothesis H. Therefore, to determine the conditional uncertainty of Y given hypothesis H, we have

$$f(Y|H,n) = \int f(Y|\hat{X},H,n) f(\hat{X}|H,n)\, d\hat{X} \tag{8}$$

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. The correspondence must also account for unmatched features in both X and Y.

3.2. Feature Correspondence

Computation of the likelihood $f(Y|\hat{X},H,n)$ requires 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).

3.2.1. Random versus Deterministic Correspondences

We consider two correspondence mappings, probabilistic and deterministic. These two correspondence mappings lead to two different expressions for the posterior likelihoods Λ_k.

We denote by $\mathcal{G}$ the set of all admissible correspondence maps. For a probabilistic correspondence model the Bayes likelihood is

$$f(Y|H,n) = \sum_{\Gamma \in \mathcal{G}} f(Y|\Gamma,H,n) P(\Gamma|H,n) \tag{9}$$

where, similarly to (8),

$$f(Y|\Gamma,H,n) = \int f(Y|\hat{X},\Gamma,H,n) f(\hat{X}|\Gamma,H,n)\, d\hat{X} \tag{10}$$

The conditioning on n, the number of extracted features, is required in (9)–(10) because Γ is a correspondence between m predicted features and n extracted features.

If the correspondence Γ is assumed to be deterministic but unknown, then Γ is a nuisance parameter in the classification problem, and we seek uniformly most powerful (UMP) or UMP-invariant decision rules.[12] In this case, no uniformly most powerful classifier exists.[13] Therefore, we adopt the Generalized Likelihood Ratio Test (GLRT) classifier,

$$f(Y|H,n) \approx \max_{\Gamma \in \mathcal{G}} f(Y|\Gamma,H,n) \tag{11}$$

where $f(Y|\Gamma,H,n)$ is computed using (10). The GLRT approach in (11) avoids the summation in (9), but requires a search for the best correspondence Γ.

3.2.2. Conditional Feature Likelihood

To implement either (9) or (11), 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. Note that the features need not be independent; only the feature *uncertainties* are assumed conditionally independent. Independence is assumed because separate features are often physically unrelated. Further,

even if the physical features contain some common attribute, the physical processes that result in feature uncertainty may be unrelated. The independence assumption significantly simplifies computation of the likelihood to yield

$$f(Y|\Gamma, H, n) = \prod_{j=1}^{n} f(Y_j|\Gamma, H, n) \tag{12}$$

In (12), each term in the product is a feature likelihood conditioned on a correspondence Γ and hypothesis H.

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

$$f(Y_j|\Gamma_j = 0, H, n) = f_{FA}(Y_j) \tag{13}$$

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 (10). In particular, from (10) it follows that for $i > 0$,

$$f(Y_j|\Gamma_j = i, H, n) = \int f(Y_j|\hat{X}_i, H, n) f(\hat{X}_i|X_i, H)\, d\hat{X}_i \tag{14}$$

For the special case of Gaussian uncertainties, (14) admits a closed-form solution. If the extracted feature attributes has Gaussian uncertainty, $f(Y_j|\hat{X}_i, H, n) \sim \mathcal{N}(\hat{X}_i, \Sigma_e)$, and if the predicted feature has uncertainty $f(\hat{X}_i|X_i, H, n) \sim \mathcal{N}(X_i, \Sigma_p)$, then from (14) we have

$$f(Y_j|\Gamma_j = i, H, n) = f(Y_j|X_i, H, n) \sim \mathcal{N}(X_i, \Sigma_p + \Sigma_e) \tag{15}$$

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

$$P(Y_j|X_i, H, n) = \sum_{\hat{X}_i} P(Y_j|\hat{X}_i, H, n) P(\hat{X}_i|X_i, H, n) \tag{16}$$

3.2.3. Correspondence Match Scores

We consider two ways to compute a likelihood match score between two sets of features from equation (9) or (11): a probabilistic many-to-many score in which every predicted feature may correspond to every extracted feature with some probability, and a one-to-one score in which we adopt a deterministic unknown correspondence assumption and maximize (11) over all one-to-one maps Γ.

Many-to-Many Likelihood Score

Two main difficulties are faced when implementing a probabilistic correspondence in (9): (i) knowledge of the priors $P(\Gamma|H, n)$, and (ii) the high computational cost of summing over all possible correspondences. The derivation of correspondence prior probabilities appears intractable for this application.

The sum over all correspondences in (9) can be simplified if the correspondence priors $P(\Gamma_j = i)$ are independent of j. Following,[8,13,1] let λ be the average number of false alarm features present in Y, and let $P_i(H),\ 1 \le i \le m$ denote the probability of detecting the ith predicted feature under hypothesis H. It follows that[8] $B = \lambda/[\lambda + \sum_{k=1}^{m} P_k(H)]$ 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 ith predicted feature. With these assumptions, the probabilistic many-to-many likelihood score is

$$f(Y|X) = f(Y|H) = \prod_{j=1}^{n} f(Y_j|X, H) = \prod_{j=1}^{n} \left[B f_{FA}(Y_j) + \sum_{i=1}^{m} D_i(H) f(Y_j|X_i, H) \right] \tag{17}$$

where $f(Y_j|X_i, H)$ is the likelihood that extracted feature Y_j corresponds to predicted feature X_i under hypothesis H.

One-to-One Likelihood Score

For a given one-to-one correspondence map Γ (in such a map each extracted feature maps to at most one predicted feature and conversely), the likelihood score is given by

$$\begin{aligned} f(Y|\Gamma,H,n) &= \left\{ P(n_F \text{ false alarms}) \prod_{\{j:\Gamma_j=0\}} f_{FA}(Y_j) \right\} \\ &\quad \cdot \left\{ \prod_{\{j:\Gamma_j=i>0\}} P_i(H) \right\} \cdot \left\{ f(Y_j|\Gamma_j=i,H,n) \prod_{\{i:\Gamma_j\neq i,\forall j\}} (1-P_i(H)) \right\} \end{aligned} \tag{18}$$

where $P_i(H)$ is the detection probability of the ith predicted feature under hypothesis H. Note that Γ defines the feature correspondences, determines a set of n_F extracted features that correspond to no predicted features (false alarms), and identifies predicted features that correspond to no extracted features (missed features). The first braced term in (18) models the likelihood of false alarm features, the second term gives the likelihood of corresponding predict-extracted feature pairs, and the third term penalizes the missed predict features.

The GLRT hypothesis selection rule in (11) finds the correspondence Γ that maximizes $f(Y|\Gamma,H,n)$ in (18) for each candidate hypothesis $H \in \mathcal{H}$. For the case that $P(n_F \text{ false alarms}) = ce^{-\beta n_F}$ for some constants c and β, the search can be implemented in $O((m+n)^3)$ operations (where m and n are the lengths of the predicted and extracted feature vectors, respectively) using the Hungarian algorithm.[14] The Hungarian algorithm finds, in $O(k^3)$ computations, the one-to-one correspondence between the elements of the $k \times 1$ vectors $[x_1,\ldots,x_k]^T$ and $[y_1,\ldots,y_k]^T$ that minimizes the cost of the correspondence, where the cost of corresponding x_i with y_j is given by the ijth entry of the $k \times k$ matrix C. The correspondence is equivalent to selecting exactly one element from each row and column of the array such that the sum of the selected entries is minimized.

The Hungarian algorithm can be modified to find the optimal correspondence between $[X_1,\ldots,X_m]^T$ and $[Y_1,\ldots,Y_n]^T$ that includes both insertions and deletions in the correspondence. To do this we employ the $(m+n)\times(m+n)$ cost matrix C given in Table 1. From (18) we observe

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

The elements on the right hand side of (19) appear in the cost matrix in Table 1. From this table we see that the Hungarian algorithm thus finds the correspondence that minimizes the log-likelihood score (19) in $O((m+n)^3)$ computations. By exploiting the structure of the cost matrix in Table 1, the computational cost can be further reduced to $O(\max(m,n)^3)$ operations.

	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		$\ddots$		$\vdots$	$\ddots$	$\vdots$
	∞		F_n	0	$\cdots$	0

Table 1. The cost matrix for the one-to-one matcher in equation (18). Here, $c_{ij} = -\log[P_i(H) f(Y_j|\Gamma_j=i,H)]$, $F_j = -\log[\beta f_{FA}(Y_j)]$, and $M_k = -\log[1-P_k(H)]$.

4. PERFORMANCE ESTIMATION

In this section we present synthetic classification performance predictions by applying the Bayes matcher to a ten-class vehicle recognition problem using X-band synthetic aperture radar imagery. We use synthetic feature vector means based on measured SAR imagery, coupled with an assumed feature perturbation model. We compare performance when using five feature attributes (Rx, Ry, $|A|$, α, and L) and when using subsets of these attributes. The experiments illustrate that the Bayes classifier is tractable for problem sizes encountered in SAR target recognition, and that it permits estimation of the Bayes error given a model for priors and feature uncertainties. The experiments also illustrate how one can explore classification performance as a function sensor parameters (*e.g.,* bandwidth, signal-to-noise ratio, extracted feature sets), and how one can explore the sensitivity of the performance to the assumed priors and feature uncertainties.

To synthesize the class means, we extract location and amplitude features for ten targets in the MSTAR Public Targets data set.[15] The data set contains 0.31 ft $\times$ 1 ft resolution SAR images of ten targets at 17° depression angle. Each image is 128×128 complex-valued pixels. For each target, approximately 270 images are available covering the full 360 degree aspect angles, for a total of 2747 images. 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.

From each image, downrange and crossrange locations and amplitudes of peaks are extracted from each image chip by finding local maxima in the SAR image. We keep the ten largest amplitude peaks. The remaining parameters are not provided by publicly available prediction modules, so are generated synthetically. The nominal values of the type attribute are generated as $\alpha \sim \mathcal{N}(0.5, 0.25)$. The length parameter is quantized to one bit for this study, and the nominal values of the length attribute are generated using a Bernoulli random variable with $P(L > 0) = 0.3$. We quantize L to avoid the computational expense of evaluating likelihoods for a non-Gaussian, mixed distribution. The γ and ϕ parameters in (2) are not used in the experiments because no strong evidence exists that these parameters can be predicted and extracted with low enough uncertainty to substantially improve classification performance. These form the 2747 class mean vectors for the 10 composite target classes. The uncertainty models for the feature attribute means, and the false alarm feature density function, are:

- **Predict Feature Uncertainties $f(\hat{X}_i|H) - X_i$:**
  $Rx,\ Ry \sim \mathcal{N}(0, 1\ \text{ft}^2)$; $\log_{10}(|A|) \sim \mathcal{N}(0, 0.25)$; no uncertainty on α and L.
- **Extract Feature Uncertainties $f(\hat{Y}_j|H) - \hat{X}_i$:**
  $Rx,\ Ry \sim \mathcal{N}(0, 1\ \text{ft}^2)$; $\log_{10}(|A|) \sim \mathcal{N}(0, 0.25)$; $\alpha \sim N(0, 1/4)$;
  Probability of incorrectly estimating $L = 0$ or $L > 0$ is 0.2.
- **False Alarm Features: $f_{FA}(Y_j)$:**
  Number: Poisson with rate $\lambda = 3$ per image chip: $P(n_F \text{ false alarms }) = e^{-\lambda}\lambda^{n_F}/(n_F!)$;
  $(Rx, Ry) \sim$ uniform over the image;
  $\log_{10}(|A|) \sim \mathcal{N}(\mu, 0.25)$, with $\mu = \log_{10}(\text{median amplitude of predicted scattering centers})$;
  $\alpha \sim \mathcal{N}(0.5, 1)$; L: Bernoulli with $P(L > 0) = 0.3$

Each feature vector X_i or Y_i is the vector $[Rx_i, Ry_i, \log_{10}|A_i|, \alpha_i, L_i]$ or a subset of these attributes, depending on the particular experiment. We assume a Gaussian uncertainty for the first four of these attributes, so the covariance matrix for the first four attributes is given by $\Sigma_p + \Sigma_e = \text{diag}[2, 2, 0.5, 0.25]$. Nondiagonal uncertainty covariance matrices are considered in the third example below. The discrete parameter L_i is characterized by the confusion matrix

$$\begin{bmatrix} P(\hat{L} = 0|L = 0) & P(\hat{L} > 0|L = 0) \\ P(\hat{L} = 0|L > 0) & P(\hat{L} > 0|L > 0) \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix} \tag{20}$$

We emulate the Index stage in Figure 1 as follows. For each of the 2747 target image chips, we find the five image chips in each of the 10 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 match scores and posterior likelihoods assuming equally likely priors ($P(n|H)P(H) = \text{constant}$) on the 50 Index hypotheses. We record the target class corresponding to the one of the 50 hypotheses with the highest likelihood score. We repeat this experiment 10 times for each class mean vector; this gives a total of 27,470 classifications from 27,470×50 matches.

The CPU time needed to compute likelihood scores are summarized in Table 2. Shown are the average computation times for 1000 matches using the many-to-many and one-to-one match scores, for the case $m = n = 10$. The matchers are implemented in unoptimized Matlab code on a 333 MHz Pentium processor. Note that $O(mn)$ computations are needed for the many-to-many match score, and $O((m+n)^3)$ computations for the one-to-one match score. For this case, the computation of the entries in The computation of the entries in equation (17) and (19) comprise a significant fraction of the total compute time for this example.

	One-to-one	Many-to-many
Using (R_x, R_y)	73.28	14.78
Using five attributes	92.00	32.02

Table 2. Average CPU time (msec) to compute correspondences and likelihood scores for $m = 10$ predict and $n = 10$ extract features. Shown are results using only two feature attributes (R_x, R_y), and using all five feature attributes.

Below we present three experiments that illustrate how one can explore classification performance as a function sensor parameters and how one can explore the sensitivity of the performance to the assumed priors and feature uncertainties. The first example considers classification performance as a function of sensor resolution, and the second and third examples consider performance sensitivity due to scattering center instability and correlated feature attributes.

Performance versus Resolution

Figure 2 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.

Figure 2 also presents results of an experiment in which we predict classification performance as SAR sensor resolution changes; we consider SAR image Rayleigh resolutions of 2, 1, $\frac{1}{2}$, and $\frac{1}{4}$ ft. We assume decreasing uncertainty in the location, frequency dependence, and length attributes as Rayleigh resolution becomes finer, as shown in Table 3. From Figure 2 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 $\frac{1}{4}$ ft. We see a clear benefit of increased bandwidth because it results in decreased feature uncertainty.

Performance Sensitivity due to Scattering Center Scintillation

Figure 3 examines the effect of mismatch on the detection probability and clutter rate of scattering centers. In this experiment the true detection probability is $P_i(H) = 0.5$ and on average $\lambda = 5$ false alarm scattering centers, all constrained to lie within a vehicle mask. All five attributes were used in the classification. The experiment simulates the effect of scattering center scintillation (which is hypothesized to result in a scattering center detection probability significantly lower than 0.9) and explores the sensitivity of classification performance to the assumed detection probability. We estimate classification performance using the true detection and false alarm rates, and also using $P_i(H) = 0.9$ and $\lambda = 1$. We see the classification performance degrades significantly compared to the 1 ft baseline case in Figure 2 (in Figure 2 there are on average 3 false alarm scattering centers that lie anywhere in the image, whereas in Figure 3 there are 5 false alarm scattering centers all in close proximity to the vehicle scattering centers). On the other hand, the classification performance appears to be robust to uncertainty mismatch in this case, as the performance in the mismatched experiment is less than 0.6 dB lower than the performance when the correct uncertainties are used in the match score.

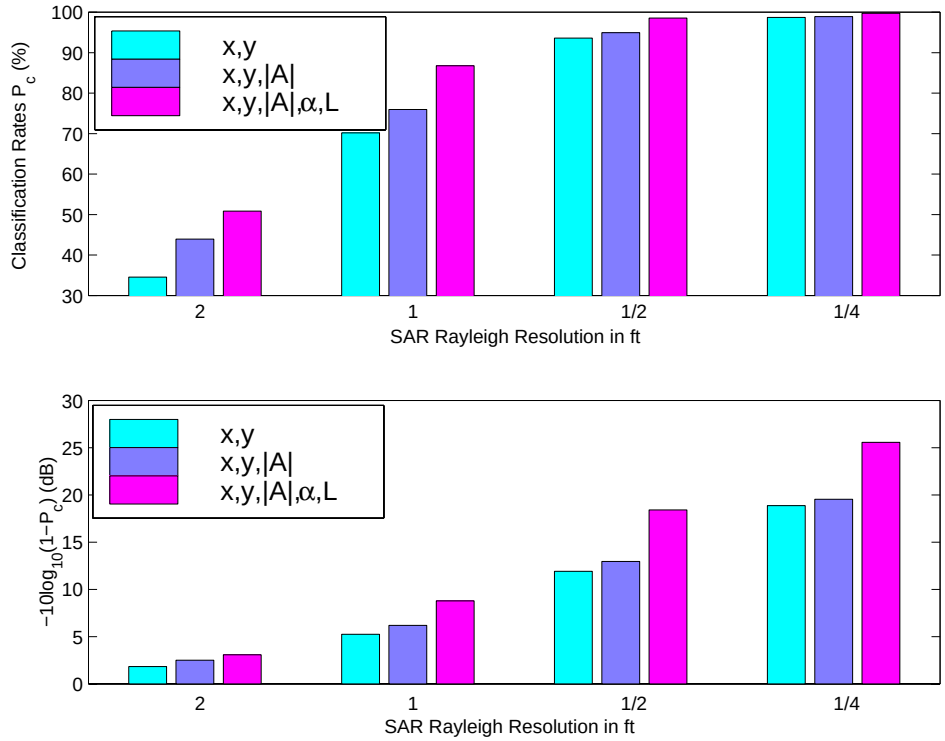

Figure 2. Classification performance as a function of number of feature attributes and radar bandwidth. The top figure shows average probability of correct classification (P_c); the bottom figure shows the same data plotted as average probability of error ($1 - P_c$) in dB.

Table 3. Sum of prediction and extraction feature attribute uncertainties and misdetection costs used in the Bayes classifier example.

	SAR Rayleigh Resolution			
Feature Attribute	**2 ft**	**1 ft**	**1/2 ft**	**1/4 ft**
locations x, y $\mathcal{N}(0, \sigma^2)$	$\sigma = 2\,\text{ft}$	$\sigma = 1\,\text{ft}$	$\sigma^2 = 1/2\,\text{ft}$	$\sigma = 1/4\,\text{ft}$
amplitude $\log_{10}(\lvert A\rvert) \sim \mathcal{N}(0, \sigma^2)$	$\sigma^2 = 0.5$	$\sigma^2 = 0.5$	$\sigma^2 = 0.5$	$\sigma^2 = 0.5$
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 $\begin{bmatrix} P(\hat{L}=0\|L=0) & P(\hat{L}=0\|L>0) \\ P(\hat{L}>0\|L=0) & P(\hat{L}>0\|L>0) \end{bmatrix}$	$\begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$	$\begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$	$\begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$	$\begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$

Performance Sensitivity to Correlated Feature Attributes

This example considers the effect on classification performance of attributes with nonzero cross-correlation. We assume a non-zero correlation coefficient of $\rho = 0.5$ or 0.9 between the downrange and crossrange location features; thus, the 2×2 top left submatrix of $\Sigma_p + \Sigma_e$ becomes

$$[\Sigma_p + \Sigma_e]_{1:2,1:2} = 2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{21}$$

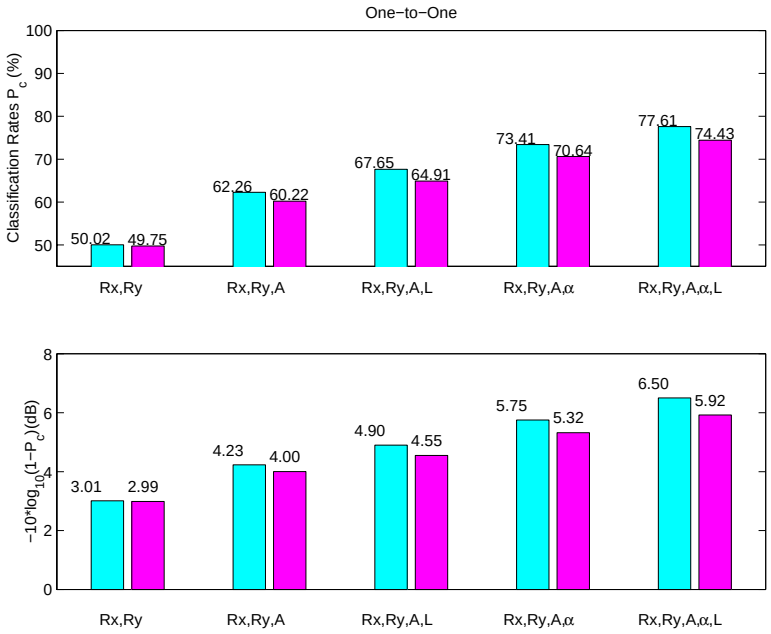

Figure 3. Classification performance using correct (left) and erroneous (right) detection probabilities and false alarm rates in the one-to-one match score.

We also perform a similar experiment in which the location features are uncorrelated features with downrange and crossrange feature variance equal to (1+ρ, 1-ρ); these are the eigenvalues of the matrix in (21). For both cases we apply two match metrics; one matcher uses the true uncertainty information and the other assumes $[\Sigma_p + \Sigma_e]_{1:2,1:2} = 2I_2$.

Figures 4 and 5 show the simulation results and for for $\rho = 0.5$ and $\rho = 0.9$, respectively, using the one-to-one matcher; the results for the many-to-many matcher are similar. The first two bars are for the correlated feature case, and show performance estimates for the matched (left) and mismatched (right) classifier; the second pair of bars give performance estimates for the uncorrelated but unequal uncertainty variance case, again with a matched and mismatched classifier. We see that the Bayesian classification performance for correlated features is nearly identicaly to classification performance for the uncorrelated attribute uncertainties if the uncertainty eigenvalues are equal in the two cases. We also see that the two mismatched classifiers give similar performance. Thus, the "rotation" of the covariance matrix has little effect on performance. In addition, the Bayesian classification performance for correlated attribute uncertainties using a mismatched matcher (that is, the matcher assumes uncorrelated attribute uncertainties with the same feature variances) is nearly identical to the Bayesian classification performance for uncorrelated attribute uncertainties with the same variance. We conclude that without knowing the mutual correlation on correlated attribute uncertainties and adopting a matcher designed for uncorrelated attribute uncertainties results in the same performance if those attribute uncertainties are actually uncorrelated, while knowing the correlation information on correlated attribute uncertainties helps to improve the ATR classification performance. This improvement can be significant for highly correlated attribute uncertainties. Furthermore, we can predict this improvement by evaluating the classification performance under uncorrelated attribute uncertainties with attribute uncertainty variances equal to the eigenvalues of the covariance matrix of the correlated attribute uncertainties.

5. CONCLUSIONS

We have considered a method for estimating classification performance for a model-based ATR system that employs attributed scattering center features. The classifier is based on a Bayes match between unordered sets of predicted and extracted scattering features. Feature uncertainty is included in the match metric. The match requires estimating the correspondence between extracted and predicted features, and the correspondence must take into account both false alarm and missed scattering centers in the extracted feature vector. We considered two match correspondence strategies, a one-to-one match and a computationally simpler many-to-many match.

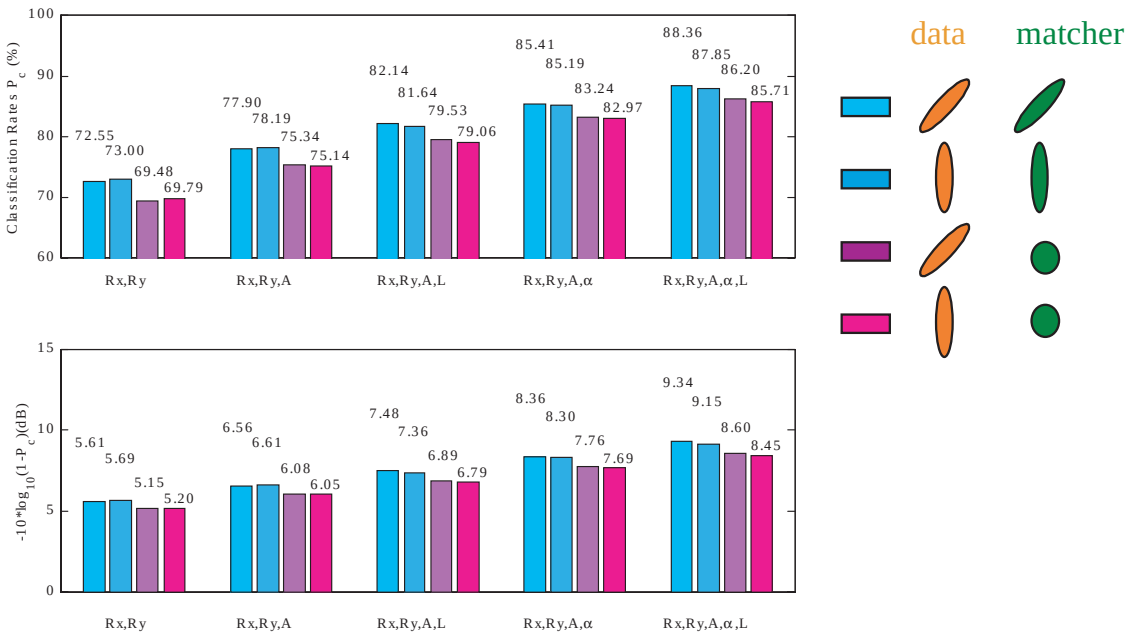

Figure 4. Classification performance results for correlated location attributes with $\rho = 0.5$ using the best one-to-one match score. The top figure shows average probability of correct classification (P_c); the bottom figure shows the same data plotted as average probability of error ($1 - P_c$) in dB.

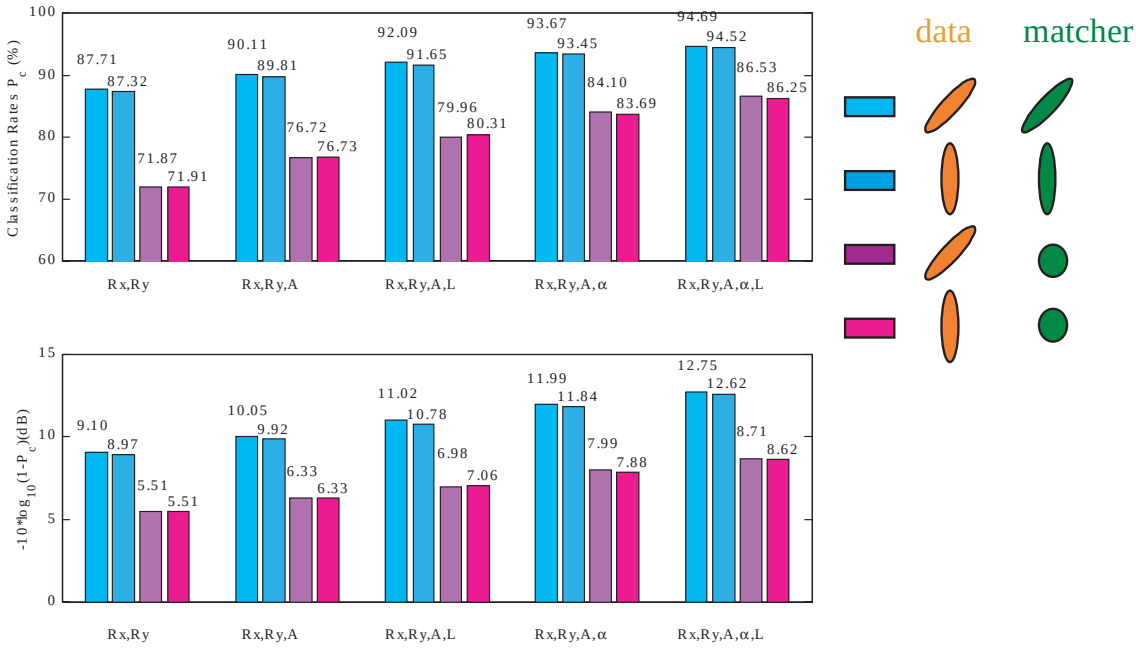

Figure 5. Classification performance results for correlated location attributes with $\rho = 0.9$ using the best one-to-one match score.

Examples presented show that a Bayesian formalism leads to tractable hypothesis testing in model-based object recognition using unordered feature sets. We presented classification performance predictions based on scattering centers extracted from measured SAR imagery of ten targets. Complexity of the Bayes matching is polynomial in the problem size. Tractability of the likelihood calculations in the Bayes approach requires only conditional independence assumptions. These are mild assumptions, much weaker than feature independence, and are reasonable for many physically-motivated feature sets.

In addition to providing a decision engine for M-ary hypothesis testing, the Bayes approach provides a structured method for predicting classification performance as a function of feature uncertainties and sensor physics. The probabilistic Bayes approach allows principled management of uncertainty in both measured data and class models. Significantly, the Bayes hypothesis testing reports normalized belief values — likelihood scores — which may be used for either maximum *a posteriori* probabilistic decisions or soft decisions.

REFERENCES

1. H.-C. Chiang and R. L. Moses, "ATR Performance Prediction Using attributed Scattering Features," in *SPIE*, (Orlando, Florida), April 1999.
2. H.-C. Chiang, R. L. Moses, and L. C. Potter, "Model-based bayesian feature matching with application to synthetic aperture radar target recognition," *Pattern Recognition*, 2000. (Special Issue on Data and Information Fusion in Image Processing and Computer Vision), (to appear).
3. L. P. Shapiro and R. M. Haralick, "Structural description and inexact matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 504–519, 1981.
4. I. Rigoutsos and R. Hummel, "A Bayesian approach to model matdching with geometric hashing," *Computer Vision and Image Uncerstanding*, vol. 62, pp. 11–26, July 1995.
5. R. C. Wilson and E. R. Hancock, "Structural matching by discrete relaxation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 634–648, June 1997.
6. R. Myers, R. C. Wilson, and E. R. Hancock, "Bayesian graph edit distance," in *Proceedings of the Tenth International Conference on Image Analysis and Processing*, (Venice, Italy), pp. 1166–1171, September 27–29 1999.
7. E. R. Keydel and S. W. Lee, "Signature Prediction For Model-Based Automatic Target Recognition," *SPIE*, vol. 2757, pp. 306–317, April 1996.
8. G. J. Ettinger, G. A. Klanderman, W. M. Wells, and W. E. L. Grimson, "A Probabilistic Optimization Approach To SAR Feature Matching," *SPIE*, vol. 2757, pp. 318–329, April 1996.
9. J. Wissinger, R. Washburn, D. Morgan, C. Chong, N. Friedland, A. Nowicki, and R. Fung, "Search Algorithms For Model-Based SAR ATR," *SPIE*, vol. 2757, pp. 279–293, April 1996.
10. L. C. Potter and R. L. Moses, "Attributed Scattering Centers for SAR ATR," *IEEE Trans. Image Processing*, vol. 6, pp. 79–91, January 1997.
11. M. J. Gerry, L. C. Potter, I. J. Gupta, and A. van der Merwe, "A parametric model for synthetic aperture radar measurements," *IEEE Transactions on Antennas and Propagation*, vol. 47, pp. 1179–1188, July 1999.
12. L. L. Scharf, *Statistical Signal Processing, Detection, Estimation, and Time Series Analysis*. Addison Wesley, 1991.
13. Hung-Chih Chiang, *Feature Based Classification with Application to Synthetic Aperture Radar*. PhD thesis, The Ohio State University, Columbus, OH, 1999.
14. C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization Algorithm and Complexity*. Englewood Cliffs, New Jersey: Prentice-Hall, 1982.
15. *MSTAR Public Target Database*. http://www.mbvlab.wpafb.af.mil/public/sdms/datasets/mstar.