PERGAMON

Pattern Recognition 34 (2001) 1539–1553

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

# Model-based Bayesian feature matching with application to synthetic aperture radar target recognition☆

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

*Department of Electrical Engineering, The Ohio State University, 2015 Neil Avenue, Columbus, OH 43210, USA*

Received 15 May 2000; accepted 15 May 2000

**Abstract**

We present a Bayesian approach for model-based classification from unordered, attributed feature sets. A set of features is estimated from measured data and is matched with a set predicted for each candidate hypothesis using a feature model. Both extracted and predicted feature sets have uncertainty, and some features may not be present in one set or the other. Computation of the match likelihoods requires a correspondence between estimated and predicted features, and two Bayesian correspondence methods are discussed. The proposed procedure is used to predict classification performance as a function of sensor parameters for a 10-vehicle target recognition problem using X-band synthetic aperture radar imagery. © 2001 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

*Keywords:* Hypothesis testing; Structural matching; Point correspondences; Performance Estimation; Synthetic aperture radar; Target recognition

## 1. Introduction

A statistical decision approach is presented for model-based M-ary classification using feature sets. The approach provides a structured, implementable method for managing complexity of the hypothesis set and measurement uncertainty. Model-based pattern matching combines uncertainty in both the object class models and the sensor data to compute posterior probabilities of hypotheses. Further, the approach permits tractable performance estimation.

### 1.1. Managing complexity

Classification tasks often must confront 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 $256 \times 256$ 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 reduce the data to lower dimension. Significantly, physically motivated features can allow a tractable alternative to a $2^{16} \times 2^{16}$ covariance matrix for description of measurement uncertainty. The features serve as statistics for the classification problem.

In addition, many pattern recognition problems 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

☆This paper is based, in part, on a paper presented at the Tenth International Conference on Image Analysis and Processing, Venice, Italy, Sep. 27–29, 1999. This work was sponsored by the US Air Force Materiel Command under contract F33615-97-1020. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the US Government.

* Corresponding author. Tel.: +1-614-292-1325; fax: +1-614-292-7596.

*E-mail address:* randy@ee.eng.ohio-state.edu (R.L. Moses).


PII: S0031-3203(00)00089-3

typical application for $10 \leqslant M \leqslant 50$ object classes might dictate $10^{12}$ subclasses [1]. Moreover, an application may require subclass decisions.

For such classification problems, a Bayes optimal decision rule is generally intractable. For example, consider the simple case in which each subclass is Gaussian with a common covariance matrix. Then, the Bayes optimal rule is to classify the image to the nearest subclass mean. The minimal sufficient statistic is the projection of measurement vector onto the smallest affine set containing the subclass means. Such a statistic would require measurement, storage and computation using $10^{12}$ templates.

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.

### 1.2. Model-based matching

In the first part of the paper we derive a Bayesian classifer that uses feature sets and model-based feature prediction to manage problem complexity. The classification approach is summarized in Fig. 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 dimensionali-ty 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. The candidate hypotheses may be class hypotheses or subclass hypotheses, depending on the application. 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 $H_k$. 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.

Finally, 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

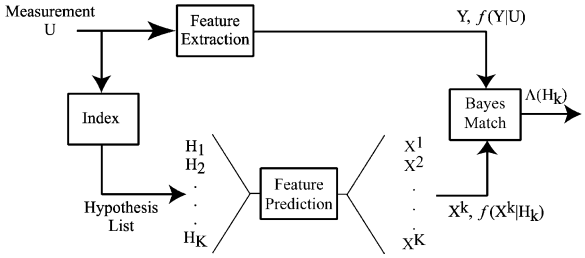


Fig. 1. A staged approach to manage high-dimensional observations and complex hypothesis sets encountered in classification of radar images.

a correspondence between extracted and predicted features and an integration over feature uncertainty. The task can be viewed as a probabilistic graph match of fully connected, attributed graphs with deletions and insertions of nodes.

Inexact feature matching, and the related inexact graph matching problem, have been the subject of much work in the pattern recognition community [2–7]. Early inexact matching approaches (e.g., Ref. [2]) use a relational distance metric that includes the ability to insert nodes into graphs to determine structural isomorphisms in the presence of structural error. Boyer and Kak [3] develop a structural matching technique that includes a conditional information measure to penalize attribute deviations in the match. Recent work on Bayesian structural matching considers subgraph matches and edit distances [6,7]. 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. Here, we consider a problem in which there is no structural relationship between the nodes; this corresponds to a graph matching problem in which the nodes are fully connected. We consider fully connected graphs and seek Bayes optimal matches that incorporate prior probabilities of node deletions and extraneous nodes; a related geometric hashing approach is considered in Ref. [5].

### 1.3. Example application

The second part of the paper focuses on application of the proposed model-based Bayesian classification approach to recognition of vehicles using synthetic aperture radar (SAR) images. We adopt an electromagnetic scattering model for use in both the feature extraction and feature prediction stages. The feature uncertainty can be predicted from knowledge of sensor noise and uncertainty in the model-based prediction process. We emulate a coarse classification stage using cross correlation between measured and catalog images and develop a

feature-based classifier using the scattering model features. The extracted feature set is perturbed by random deletions of predicted features or insertions of additional features not present in the predicted set.

We present synthetic classification performance predictions using class means estimated from measured X-band SAR imagery of 10 vehicles. In particular, we assess the value of specific feature attributes in the model. 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 and that the optimal Bayes error rate is estimable given priors and feature uncertainties. The example also illustrates how one can use the Bayes classifier as a simulation tool to explore performance as a function of sensor parameters (bandwidth, signal-to-noise ratio, number of extracted features, etc.) or to explore sensitivity of classification performance to assumed priors and feature uncertainties.

The paper is organized as follows. In Section 2 we outline the classification problem and discuss the Bayesian likelihood framework. Evaluation of the feature match likelihoods requires feature correspondence and feature likelihood scoring from uncertainty models; Section 3 discusses these issues. In Section 4 we propose two match scoring functions based on probabilistic and deterministic correspondence assumptions; we relate the two match scores and present two fast, approximate computational techniques for obtaining the deterministic correspondence. Section 5 outlines the SAR application; we present the feature and clutter models and parameter uncertainty. In Section 6 we present classification performance results obtained by applying the Bayes classifier to a 10-class vehicle recognition problem. Section 7 presents conclusions.

## 2. Classification problem statement

The Bayes matching problem we consider is given as follows. At the input to the classifier stage, we are given a vector of $n$ features

$$Y = [Y_1, Y_2, \ldots, Y_n]^{\mathrm{T}}, \tag{1}$$

extracted from a measurement, $U$. Each feature $Y_i$ is an $\ell \times 1$ vector of ordered attributes. The attributes may characterize location, amplitude, pose, or other properties of the features. We are also given a set $\mathscr{H} = \{H_k,\ k \in [1, K]\}$ of $K$ candidate hypotheses, along with their prior probabilities $P(H_k)$. The set $\mathscr{H}$ typically is provided by the preceding index stage, as depicted in Fig. 1.

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

$$X^k = [X_1^k, X_2^k, \ldots, X_{m_k}^k]^{\mathrm{T}}. \tag{2}$$

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). Uncertainty models are presented in Section 3B.

We seek to compute the posterior likelihood of the observed features, $Y$, under each of the $K$ hypotheses in the set $\mathscr{H}$.

$$\Lambda_k = P(H_k|Y), \quad H_k \in \mathscr{H}. \tag{3}$$

The most likely hypotheses and their corresponding likelihoods are reported at the output of the classifier. If the desired output is a final classification, we adopt the maximum a posteriori probability (MAP) decision rule and choose the class corresponding to the highest posterior likelihood score.

To compute the posterior likelihood in Eq. (3), we apply Bayes rule for any $H \in \mathscr{H}$ to obtain[1]

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

$$= \frac{f(Y|H,n)P(n|H)P(H)}{f(Y|n)P(n)}. \tag{4}$$

The conditioning on $n$ is required because the number of features in $Y$ is itself a random variable. Since the denominator of Eq. (4) does not depend on hypothesis $H$, the MAP decision is found by maximizing $f(Y|H,n)P(H)P(n|H)$ over $H \in \mathscr{H}$. 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. The uncertainty is modeled as shown in Fig. 2. 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 $H$.

[1] For notational simplicity, we drop the subscript on hypothesis $H_k$ in the sequel, and consider a general $H \in \mathscr{H}$ Correspondingly, we drop the $k$ on $X^k$ and $m_k$ in Eq. (2).

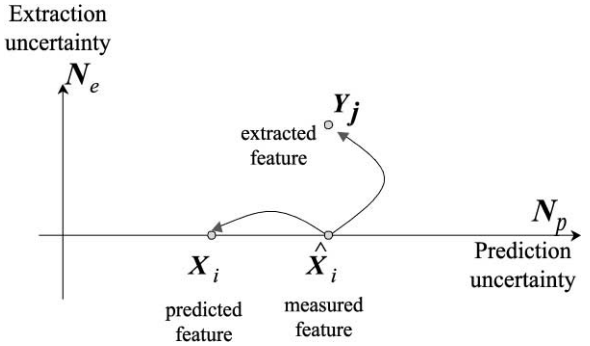


Fig. 2. Deviation of a "true" feature on an object being measured due to both measurement uncertainty $N_e$ and model uncertainty $N_p$.

Therefore, to determine the conditional uncertainty of $Y$ given hypothesis $H$, we have

$$f(Y|H,n) = \int f(Y|\hat{X},H,n) f(\hat{X}|H,n)\,\mathrm{d}\hat{X}, \tag{5}$$

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. Feature correspondence

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

We denote by $\mathscr{G}$ the set of all admissible correspondence maps. For some applications, only one-to-one maps are admissible; examples include Refs. [3,6]. In a one-to-one map, each predicted feature can correspond to at most one extracted feature, and conversely. For other applications, $\mathscr{G}$ might include many-to-many mappings [8].

### 3.1. 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 $\Lambda_k$.

For a probabilistic correspondence model the Bayes likelihood is

$$f(Y|H,n) = \sum_{\Gamma\in\mathscr{G}} f(Y|\Gamma,H,n) P(\Gamma|H,n), \tag{6}$$

where, similarly to Eq. (5),

$$f(Y|\Gamma,H,n) = \int f(Y|\hat{X},\Gamma,H,n) f(\hat{X}|\Gamma,H,n)\,\mathrm{d}\hat{X}. \tag{7}$$

The conditioning on $n$, the number of extracted features, is required in Eqs. (6)–(7) because $\Gamma$ 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 $Y$.

In contrast, if the correspondence $\Gamma$ is assumed to be deterministic but unknown, then $\Gamma$ is a nuisance parameter in the classification problem, and we seek uniformly most powerful (UMP) or UMP-invariant decision rules [9]. In this case, no uniformly most powerful classifier exists [10]. Therefore, we adopt the generalized likelihood ratio test (GLRT) classifier,

$$f(Y|H,n) \approx \max_{\Gamma\in\mathscr{G}} f(Y|\Gamma,H,n), \tag{8}$$

where $f(Y|\Gamma,H,n)$ is computed using Eq. (7). The GLRT approach in Eq. (8) avoids the summation in Eq. (6), but requires a search for the best correspondence $\Gamma$.

### 3.2. Conditional feature likelihood

To implement either Eq. (6) or (8), we require a probability model for $f(Y|\hat{X},\Gamma,H,n)$. To develop such a model, 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 validity of the independence assumption is, of course, application dependent; a justification for a radar target recognition example is presented in Section 5. The independence assumption significantly simplifies computation of the likelihood to yield

$$f(Y|\Gamma,H,n) = \prod_{j=1}^{n} f(Y_j|\Gamma,H,n). \tag{9}$$

In Eq. (9), each term in the product is a feature likelihood conditioned on a correspondence $\Gamma$ 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

$$f(Y_j|\Gamma_j = 0, H, n) = f_{FA}(Y_j). \tag{10}$$

If $Y_j$ corresponds to a predicted feature $X_i$, we write $\Gamma_j = i$ (for $i > 0$) and compute the feature match score from Eq. (7). In particular, from Eq. (7) it follows that for $i > 0$,

$$f(Y_j|\Gamma_j = i, H, n) = \int f(Y_j|\hat{X}_i, H, n) f(\hat{X}_i|X_i, H)\, \mathrm{d}\hat{X}_i. \tag{11}$$

For the special case of Gaussian uncertainties, Eq. (11) admits a closed-form solution. Let the extracted feature attributes have Gaussian uncertainty with zero mean and covariance $\Sigma_e$

$$f(Y_j|\hat{X}_i, H, n) \sim \mathcal{N}(\hat{X}_i, \Sigma_e) \tag{12}$$

and let the predicted feature have uncertainty

$$f(\hat{X}_i|X_i, H, n) \sim \mathcal{N}(X_i, \Sigma_p). \tag{13}$$

Then from Eqs. (11)–(13) we have

$$f(Y_j|\Gamma_j = i, H, n) = f(Y_j|X_i, H, n) \sim \mathcal{N}(X_i, \Sigma_p + \Sigma_e). \tag{14}$$

Thus, the conditional log-likelihood is proportional to the quadratic distance $d(X_i, Y_j) = (Y_j - X_i)^{\mathrm{T}}(\Sigma_p + \Sigma_e)^{-1}(Y_j - X_i)$. Similarly, for features whose attributes are discrete-valued, the likelihood is the sum

$$P(Y_j|X_i, H, n) = \sum_{\hat{X}_i} P(Y_j|\hat{X}_i, H, n) P(\hat{X}_i|X_i, H, n). \tag{15}$$

Extensions to non-Gaussian features or mixed continuous and discrete feature attributes is a straightforward extension of Eqs. (7) and (15); see Ref. [10]. For these cases the integral in Eq. (7) is evaluated either analytically or numerically, giving a function $g(X_i, Y_j)$ to be evaluated. The computational complexity increase over the Gaussian case is minor unless the function $g$ is significantly more computationally expensive to evaluate than is the quadratic distance $d(X_i, Y_j)$ in the Gaussian case.

## 4. Correspondence match scores

We next address the computation of a likelihood match score between two sets of features from Eq. (6) or (8). We propose four match scores in this section. The first is a many-to-many score in which every predicted feature may correspond to every extracted feature; thus, the admissible set $\mathcal{G}$ of possible correspondences is the set of all possible links as shown in Fig. 3. While either a probabilistic or deterministic correspondence model can be used, we choose a probabilistic model, because an assumption on the correspondence priors leads to a particularly simple form for the match score. The second match score admits only one-to-one correspondences between the feature sets. For this case we adopt a deterministic unknown correspondence assumption and maximize Eq. (8) over $\Gamma$. Finally, we develop two computationally efficient but suboptimal likelihood scores for the one-to-one correspondence.

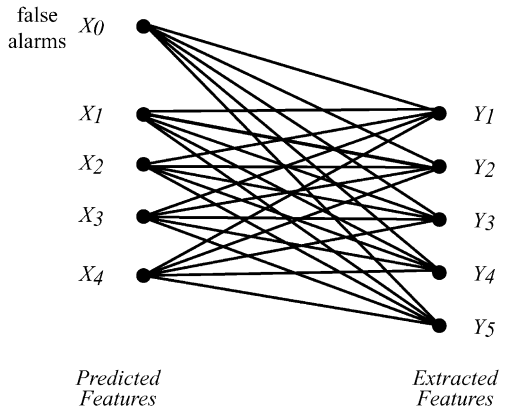


Fig. 3. The many-to-many correspondence map between predicted and extracted features. Every correspondence occurs with some probability.

### 4.1. Many-to-many likelihood score

Two main difficulties are faced when implementing a probabilistic correspondence in Eq. (6): (i) knowledge of the priors $P(\Gamma|H, n)$, and (ii) 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 appears intractable for most applications.

The sum over all correspondences in Eq. (6) can be simplified if the correspondence priors $P(\Gamma_j = i)$ are independent of $j$. Following Refs. [8,10], let $\lambda$ be the average number of false alarm features present in $Y$, and let $P_i(H)$, $1 \leqslant i \leqslant m$ denote the probability of detecting the $i$th predicted feature under hypothesis $H$. It follows that [8]

$$B = \frac{\lambda}{\lambda + \sum_{k=1}^{m} P_k(H)} \tag{16}$$

Table 1
The likelihood matrix for the many-to-many matcher in Eq. (19)

| | $Y_1$ | $\cdots$ | $Y_n$ |
|---|---|---|---|
| $X_1$ | $D_1(H)f(Y_1\|X_1)$ | $\cdots$ | $D_1(H)f(Y_n\|X_1)$ |
| $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| $X_m$ | $D_m(H)f(Y_1\|X_m)$ | $\cdots$ | $D_m(H)f(Y_n\|X_m)$ |
| False alarms | $Bf_{FA}(Y_1)$ | $\ddots$ | 0 |
| | 0 | | $Bf_{FA}(Y_n)$ |

is the probability that an extracted feature is a false alarm and

$$D_i(H) = (1 - B)\frac{P_i(H)}{\sum_{k=1}^{m} P_k(H)} \tag{17}$$

is the probability that an extracted feature comes from the $i$th predicted feature. With these assumptions, the probabilistic many-to-many likelihood score is

$$f(Y|X) = f(Y|H) = \prod_{j=1}^{n} f(Y_j|X, H) \tag{18}$$

$$= \prod_{j=1}^{n} \left[ Bf_{FA}(Y_j) + \sum_{i=1}^{m} D_i(H) f(Y_j|X_i, H) \right] \tag{19}$$

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

The many-to-many likelihood score in Eq. (19) can be conveniently computed using the cost function array in Table 1. The elements of the array are computed using Eqs. (14), (16), and (17). The likelihood score $f(Y|X)$ is then computed by summing each column of the array, then taking the product of the resulting column sums.

### 4.2. One-to-one likelihood score

For many applications, each extracted feature can correspond to at most one predicted feature and conversely. An example one-to-one correspondence map is shown in Fig. 4. For this case, a given correspondence map $\Gamma$ 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 likelihood score is then given by

$$f(Y|\Gamma, H, n) = \left\{ P(n_F \text{ false alarms}) \prod_{\{j:\Gamma_j = 0\}} f_{FA}(Y_j) \right\} \times \left\{ \prod_{\{j:\Gamma_j = i > 0\}} P_i(H) \right\} \times \left\{ f(Y_j|\Gamma_j = i, H, n) \prod_{\{i:\Gamma_j \neq i, \forall j\}} (1 - P_i(H)) \right\}, \tag{20}$$

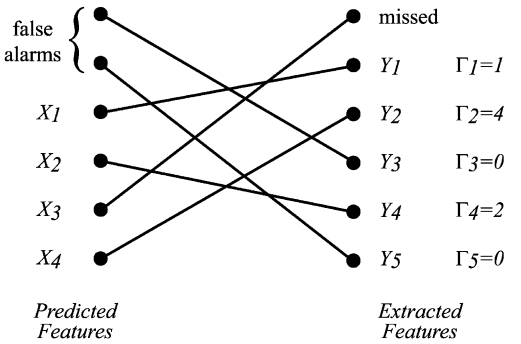


Fig. 4. An example one-to-one correspondence mapping for $m = 4$ and $n = 5$. Extracted features $Y_3$ and $Y_5$ are false alarms, and predicted feature $X_3$ is missed.

Table 2
The cost matrix for the one-to-one matcher in Eq. (20). Here, $c_{ij} = -\log[P_i(H)f(Y_j|\Gamma_j = i, H)]$, $F_j = -\log[\beta f_{FA}(Y_j)]$, and $M_k = -\log[1 - P_k(H)]$

| | $Y_1$ | $\cdots$ | $Y_n$ | Misses | | |
|---|---|---|---|---|---|---|
| $X_1$ | $c_{11}$ | $\cdots$ | $c_{1n}$ | $M_1$ | | $\infty$ |
| $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | | $\ddots$ | |
| $X_m$ | $c_{m1}$ | $\cdots$ | $c_{mn}$ | $\infty$ | | $M_m$ |
| false alarms | $F_1$ | $\ddots$ | $\infty$ | 0 | $\cdots$ | 0 |
| | | | | $\vdots$ | $\ddots$ | $\vdots$ |
| | $\infty$ | | $F_n$ | 0 | $\cdots$ | 0 |

where $P_i(H)$ is the detection probability of the $i$th predicted feature under hypothesis $H$. The first braced term in Eq. (20) 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 Eq. (8) finds the correspondence $\Gamma$ that maximizes $f(Y|\Gamma, H, n)$ in Eq. (20) for each candidate hypothesis $H \in \mathscr{H}$. For the case that $P(n_F$ false alarms) obeys an exponential rule

$$P(n_F \text{ false alarms}) = c\mathrm{e}^{-\beta n_F} \tag{21}$$

for some constants $c$ and $\beta$, the search can be efficiently implemented using the Hungarian algorithm [11].[2] The Hungarian algorithm finds, in $O(k^3)$ computations, the one-to-one correspondence between the elements of the $k \times 1$ vectors $[x_1, \ldots, x_k]^{\mathrm{T}}$ and $[y_1, \ldots, y_k]^{\mathrm{T}}$ that minimizes the cost of the correspondence, where the cost of corresponding $x_i$ with $y_j$ is given by the $ij$th entry of the $k \times k$ matrix $C$. The correspondence is equivalent to

[2] We thank Dr. William Irving for noting the application of the Hungarian algorithm to this search problem.

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 2. From Eqs. (20) and (21) we observe

$$\begin{aligned} -\log f(Y|X,\Gamma,H,n) &= -\sum_{\{j:\Gamma_j=0\}} \log[\beta f_{FA}(Y_j)] \\ &\quad -\sum_{\{j:\Gamma_j=i>0\}} \log[P_i(H) f(Y_j|\Gamma_j=i,H,n)] \\ &\quad -\sum_{\{i:\Gamma_j\neq i,\forall j\}} \log[1-P_i(H)] + \text{constant}. \end{aligned} \tag{22}$$

The elements on the right-hand side of Eq. (22) appear in the cost matrix in Table 2. Assume the correspondence resulting from applying the Hungarian algorithm with this cost matrix is $\{i, j=\Gamma_i\}$ where $\Gamma_i$ is a permutation of the integers $1, \ldots, (m+n)$. Consider $i \leqslant m$. If $\Gamma_i = j$ for some $j \in [1, \ldots, n]$, then $X_i$ corresponds to $Y_j$ with cost $c_{ij} = -\log[P_i(H) f(Y_j|\Gamma_j = i, H, n)]$. If $\Gamma_i > n$, then no $Y_j$ corresponds to $X_i$. In this case $j = i+n$ and $c_{i,i+n} = -\log[1-P_i(H)]$ is the miss probability cost for $X_i$. Note that $j$ cannot be any other integer greater than $n$, because the corresponding cost $c_{ij} = \infty$. Similarly, if $i > m$ and $\Gamma_i = j \in [1, \ldots, n]$, then $i = j+m$ with cost $c_{(m+j),j} = -\log[\beta f_{FA}(Y_j)]$. Here, feature $Y_j$ corresponds to no $X_i$, and is thus labeled as a false alarm. Finally, correspondences $\Gamma_i = j$ for $i > m$ and $j > n$ incur zero cost. The Hungarian algorithm thus finds the correspondence that minimizes the log-likelihood score (22) in $O((m+n)^3)$ computations.

### 4.3. Relationship between one-to-one and many-to-many likelihood scores

The one-to-one and many-to-many likelihood scores are derived under different assumptions about correspondence maps, but give similar performance in some cases. Because of its computational simplicity, the many-to-many score may be preferred even for problems in which it is known that only one-to-one correspondences are admissible. In this section we relate the two scores.

The relationship between the one-to-one and many-to-many likelihood scores can be seen by comparing Tables 1 and 2. The match score in Table 1 is found by summing the columns and then multiplying the sums; if one term in each column dominates, then the result is approximately the same as the product of the largest element of each column. In addition, the elements in Table 1 are similar to the first $n$ columns of Table 2 except for logarithms and constants that precede the entries.

Let $X$ and $Y$ be the predicted and extracted feature vectors. Because the features are unordered, suppose (without loss of generality) that $Y_j$ corresponds to $X_j$ for, $j = 1, \ldots, t$. Assume further that the feature uncertainty is sufficiently small such that

$$f(Y_j|X_j) \gg f(Y_j|X_i), \quad j \in \{1, \ldots, t\}, \ \forall i \neq j, \tag{23}$$

$$f_{FA}(Y_j) \gg f(Y_j|X_i), \quad j \in \{t+1, \ldots, n\}, \ \forall i. \tag{24}$$

These assumptions hold, for example, if the feature covariances are small compared to the squared distances $\|Y_j - X_i\|^2$, $i \neq j$ between the features. Under these assumptions, the many-to-many likelihood in Eq. (19) can be approximated as

$$\begin{aligned} f_{m-m}(Y|\Gamma,H,n) &\approx \prod_{j=1}^{t} \left[ \frac{(1-B)}{\sum_{k=1}^{m} P_k(H)} P_j(H) f(Y_j|X_j) \right] \\ &\quad \times \prod_{j=t+1}^{n} [B f_{FA}(Y_j)]. \end{aligned} \tag{25}$$

Furthermore, if the correct correspondence map is selected (which is the case with high probability if the assumptions (23)–(24) hold), then the one-to-one likelihood in Eq. (20) can be written as

$$\begin{aligned} f_{1-1}(Y|\Gamma,H,n) &= c \prod_{j=t+1}^{m} [1-P_j(H)] \\ &\quad \times \prod_{j=1}^{t} [P_j(H) f(Y_j|X_j)] \prod_{j=t+1}^{n} [\beta f_{FA}(Y_j)] \end{aligned} \tag{26}$$

where $c$ and $\beta$ are defined in Eq. (21).

From Eqs. (25) and (26) we see that if $m$, $n$ and $t$ are approximately equal across likely hypotheses, then the two match scores differ only in the relative weights applied to correspondences and false alarms. If one wishes to use the many-to-many score to approximate the one-to-one score, then $\lambda$ and $P_i(H)$ in the many-to-many match score can be chosen to give similar weights to the one-to-one score using Eqs. (19) and (25). Such an approximation is motivated by the much lower computational cost of the many-to-many score, which avoids the search over all one-to-one correspondences.

### 4.4. Suboptimal one-to-one likelihood scoring

We consider two suboptimal one-to-one correspondence procedures that search over only a subset of correspondence maps. Referring to the one-to-one cost matrix in Table 2, the suboptimal strategies are illustrated in Fig. 5.

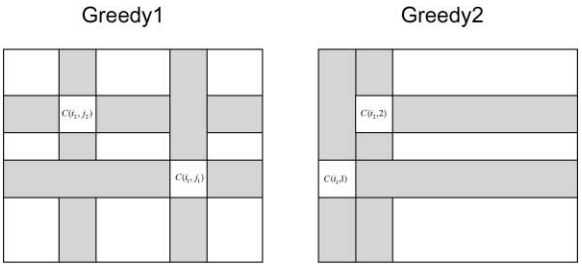


Fig. 5. Two suboptimal approaches for assigning correspondences from the one-to-one cost matrix in Table 2.

In Fig. 5, Greedy1 searches the first $n$ columns of the cost matrix in Table 2. The smallest entry is $C(i_1, j_1)$ is found, and $(i_1, j_1)$ is labeled as a correspondence. Then the $i_1$th row and $j_1$th column are removed, and the process continues until columns 1 to $n$ are labeled. Any predicted features that remain unmatched are labeled as missed.

A second suboptimal search, called Greedy2, searches the first column of the cost matrix for the smallest entry $C(i_1, 1)$, and labels $(i_1, 1)$ as a correspondence. The $i_1$th row and first column of the matrix are removed, and the process continues.

The Greedy1 method generally identifies a correspondence with a higher likelihood score than does Greedy2, but is a computationally more expensive search than is Greedy2. The classification performance resulting from these two suboptimal search strategies are compared to the one-to-one and many-to-many match score methods in Section 6.

## 5. Application to synthetic aperture radar classification

Synthetic aperture radar (SAR) provides all-weather, day-or-night remote sensing for mapping, search-and-rescue, mine detection, and target recognition [12]. Many SAR classification methods rely on image-based metrics, such as cross-correlation of the measured SAR image with empirically or synthetically derived image templates [13]. Staged classification is often used to efficiently search the hypothesis space. However, problem complexity has motivated recent interest in applying model-based classification to SAR imagery [8,14–16].

In this section we apply the proposed model-based Bayesian feature matching approach to the problem of object classification for SAR imagery. We adopt a parametric scattering center model to describe the SAR imagery; the parameters of the model become feature sets for feature-based matching. Each scattering center is a feature characterized by a vector of attributes describing the location, amplitude, curvature, length, and pose. The scattering centers are unordered, and matching of scattering centers naturally fits the feature set matching problem described above.

### 5.1. Data collection and image formation

The radar data are collected over a range of frequencies and aspect angles. Typically, the measured data are uniformly sampled in both $f \in [f_{\min}, f_{\max}]$ and $\phi \in [\phi_{\min}, \phi_{\max}]$. The SAR image formation process involves first resampling the data $E(f, \phi)$ to a uniform grid on the Cartesian coordinate space $(f_x, f_y) = (f\cos\phi, f\sin\phi)$. The rectilinear data $E(f_x, f_y)$ are then multiplied by a two-dimensional window function $W(f_x, f_y)$ and zero padded. Finally, a two-dimensional inverse discrete Fourier transform gives a complex-valued image $E(x, y)$ sampled on the image plane. The image is a spatial map of the microwave reflectivity.

### 5.2. An attributed scattering center feature model

We adopt a parametric data model which is based on approximate radar scattering physics [17,18]. From the geometric theory of diffraction [19,20], 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 [17]

$$E(f, \phi) = \sum_{k=1}^{p} E_k(f, \phi), \tag{27}$$

where each scattering center is modeled as

$$E_k^s(f, \phi) = A_k \exp\left(j\frac{4\pi f}{c}(Rx_k \cos\phi + Ry_k \sin\phi)\right) \times \left(j\frac{f}{f_c}\right)^{\alpha_k} \operatorname{sinc}\left(\frac{2\pi f}{c} L_k \sin(\phi - \bar{\phi}_k)\right). \tag{28}$$

In Eq. (28), $f_c$ is the center frequency of the radar bandwidth 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 $\alpha_k$ characterizes curvature of the scattering center. The vector

$$\theta_k = [Rx_k, Ry_k, A_k, \alpha_k, L_k, \bar{\phi}_k], \quad k = 1, \ldots, p \tag{29}$$

is the attribute vector for each feature, and is $X_i$ or $Y_j$ in the Bayes classification notation of Section 2.

The model in Eqs. (27)–(28) 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 $(\alpha, L)$ parameters [17].

### 5.3. Parameter estimation and parameter uncertainty

We assume the measured data $D(f, \phi)$ are well modeled by Eqs. (27)–(28) with an additive perturbation

$$D(f, \phi) = \sum_{k=1}^{p} E_k(f, \phi) + N(f, \phi). \tag{30}$$

Here, $N(f, \phi)$ represents the modeling error (background clutter, sensor noise, model mismatch, incomplete motion compensation, antenna calibration errors, etc.) and is modeled as a Gaussian noise process with known covariance.

To estimate the parameter vector $\theta$ from measured SAR imagery, we exploit the property that the energy of a given scattering center is localized in the image. We segment high-energy regions $R_k$ in the image and for each region obtain approximate maximum-likelihood (AML) estimates of the parameters of a low-order scattering center model [21]. Because the noise is Gaussian and the imaging process is linear, AML estimates are found by solving a nonlinear minimization problem of the form

$$\hat{\theta}_{k,AML} = \underset{\theta}{\operatorname{argmin}} \, [d - s(\theta)]^H \Sigma^{-1} [d - s(\theta)], \tag{31}$$

where $d$ is a vector of image pixels in the region $R_k$, $s(\theta)$ is the scattering model vector for these image pixels, and $\Sigma$ is the covariance matrix of the noise vector for this region. By estimating model parameters on regions, we decouple the high-order parameter estimation problem into a set of smaller estimation problems, providing a large gain in computational speed with only a slight degradation in estimation bias and variance. In addition, we gain robustness of the Gaussian noise assumption by requiring that assumption to hold locally around the scattering centers; thus, the estimator is robust to large clutter terms in the image that are not well modeled as Gaussian noise.

Use of estimated model parameters for Bayesian hypothesis testing requires that uncertainty be associated with each estimate. Motivated by the near statistical efficiency of the estimated parameters, we use the Cramér–Rao lower bound (CRB) to predict the feature uncertainty. The Cramér–Rao lower bound is derived in Ref. [7].

## 6. Performance evaluation

In this section we present synthetic classification performance predictions by applying the Bayes matcher to a 10-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 ($Rx$, $Ry$, $|A|$, $\alpha$, 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 10 targets in the MSTAR Public Targets data set [22]. The data set contains 0.3 m resolution SAR images of 10 targets at 17° depression angle. Each image is $128 \times 128$ complex-valued pixels. For each target, approximately 270 images are available covering the full 360° 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. Examples of the SAR image chips are shown in Fig. 6.

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 10 largest amplitude peaks. The remaining parameters are not provided by publicly available

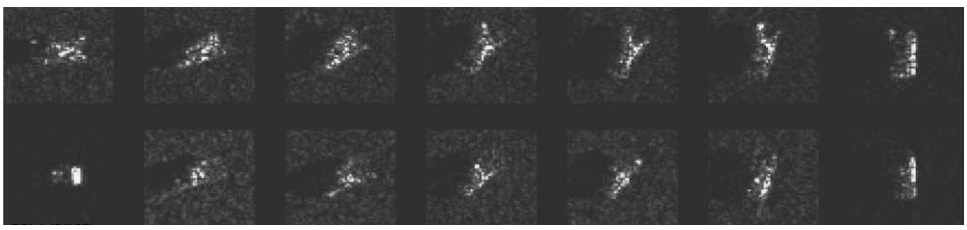

Fig. 6. Examples of the MSTAR SAR image chips used in the performance studies. Seven T-72 (top) and BMP-2 (bottom) are shown.

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 $P(L > 0) = 0.3$. These form the 2747 class mean vectors for the ten composite target classes. The uncertainty models for the feature attribute means, and the false alarm feature density function, are

- **Predict feature uncertainties $f(\hat{X}_i|H) - X_i$:**
  $Rx$, $Ry$ $n(0, 1\ ft^2)$; $\log_{10}(|A|) \sim \mathcal{N}(0, 0.25)$;
  no uncertainty on $\alpha$ 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: $P(n_F \text{ false alarms}) = e^{-\lambda}\lambda^{n_F}/(n_F!)$ with rate $\lambda = 3$;
  $(Rx, Ry) \sim$ uniform over the image;
  $\log_{10}(|A|) \sim \mathcal{N}(\mu, 0.25)$, with $\mu = \log_{10}$(median amplitude of predicted scattering centers);
  $\alpha \sim \mathcal{N}(0.5, 1)$; $L$: Bernoulli with $P(L > 0) = 0.3$.

An example realization of predicted and extracted location features is shown overlaid on the measured SAR image in Fig. 7.

In the above uncertainty model we have assumed that both the predicted and extracted features have conditionally independent uncertainties. Uncertainties in feature predictions arise from uncertainty in the locations and orientations of the facets in a CAD model of the vehicle. Because predicted scattering centers result from electromagnetic energy reflecting from these facets, and because different scattering centers generally correspond to different sets of facets, the scattering prediction errors can be assumed to be independent. If they are dependent, no model is available to describe the dependencies. Independence of extracted feature uncertainties is justified because the feature uncertainty covariance matrix is Gaussian and nearly diagonal. Gaussianity is reasonable, from the central limit theorem, due to the extensive averaging inherent in data collection and image formation. For example, if an approximate maximum-likelihood method is used to estimate features, then the feature uncertainty is given (approximately) by the Cramér–Rao bound, which is nearly diagonal if the separation between scattering centers is greater than one resolution cell of the SAR system [17].

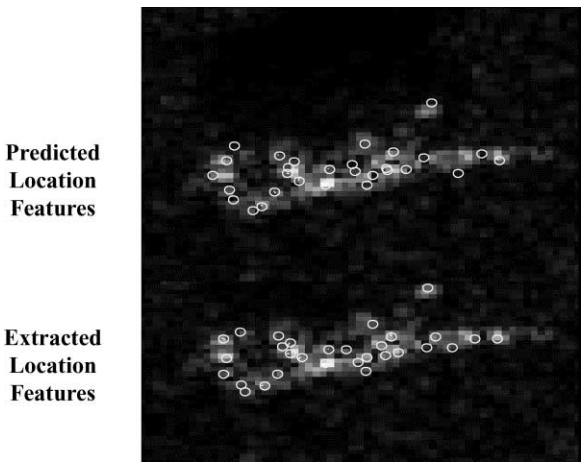


Fig. 7. Example predicted and extracted location features for a measured SAR image.

We emulate the index stage in Fig. 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 $P_d = 0.5$ or 0.9, so not all scattering centers are present in the extracted feature vector. The $P_d = 0.9$ choice is proposed in Ref. [8] and has been used in a fielded system; the $P_d = 0.5$ value is used to model scintillation of scattering centers (i.e., rapid change in scattering amplitude as a function of target pose). We also add clutter scattering centers to the extract feature vector. We then compute the match scores and posterior likelihoods assuming equally likely priors ($P(n|H)P(H) =$ 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 $27{,}470 \times 50$ matches.

We summarize the overall performance as an average probability of correct classification $P_c$. Fig. 8 presents the results of the above experiment when each scattering center has a detection probability $P_i(H) = 0.5$. Fig. 9 presents results when the detection probability is 0.9. Results are shown for the many-to-many, one-to-one, and two suboptimal one-to-one likelihood scores. In addition, classification performance is shown when two, three, four, or five of the available scattering center attributes are used in the match score. Comparing Figs. 8 and 9, we see significant improvement in classification performance when the scattering center detection probability increases; this is not surprising, because in the case

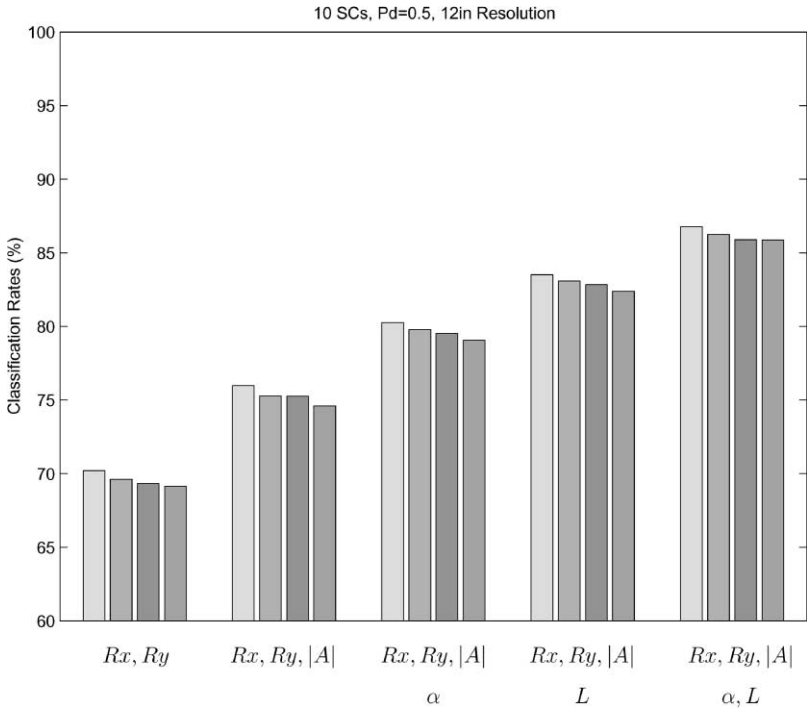


Fig. 8. Classification performance as a function of number of feature attributes and match score used. Feature detection probability is $P_i(H) = 0.5$. For each case, the four bars are: one-to-one (left), Greedy1, Greedy2, and many-to-many (right) match score.

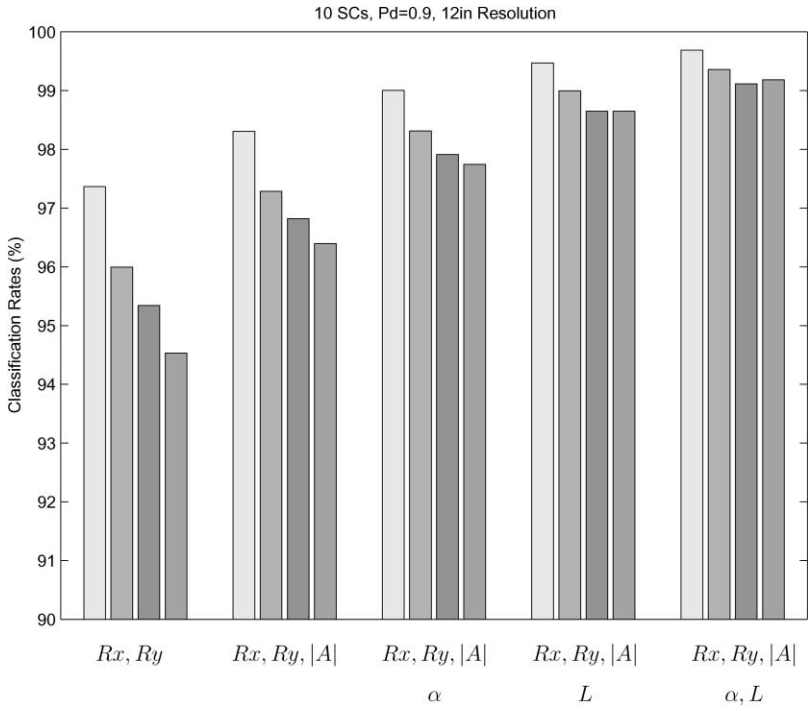


Fig. 9. Classification performance as a function of number of feature attributes and match score used. For each case, the four bars are: one-to-one (left), Greedy1, Greedy2, and many-to-many (right) match score. Feature detection probability is $P_i(H) = 0.9$.

of Fig. 8 there are on average $10 \times 0.5 = 5$ correctly detected scattering centers that can be corresponded in computing a match score, whereas in Fig. 9 the average number of correct scattering centers increases to 9. Within each figure, we see a substantial improvement in performance when the additional feature attributes are used. On the other hand, classification performance depends only mildly on the match metric used; the one-to-one metric gives the best performance and the many–many metric generally gives the worst performance of the four metrics, but the performance difference between the extremes is only a few percent.

Figs. 10 and 11 illustrate the sensitivity of classification performance to the assumed priors and uncertainty models. Fig. 10 shows classification performance when the assumed location uncertainty has standard deviation 0.5, 1, and 2 times the correct location uncertainty; the other assumed priors and uncertainties agree with the true ones. The scattering center detection probability is $P_i(H) = 0.9$ and a one-to-one match metric is used in these experiments. We see that the classification performance rates decrease by 1–3 dB due to mismatch in the location uncertainty model, with higher-performance degradation as the correct classification rate becomes closer to one. The results are shown for the one-to-one likelihood score, but performances for the other likelihood score methods show similar trends.

Fig. 11 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. 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 case shown in Figs. 8 (in Fig. 8 there are on average three false alarm scattering centers that lie anywhere in the image, whereas in Fig. 11 there are five 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

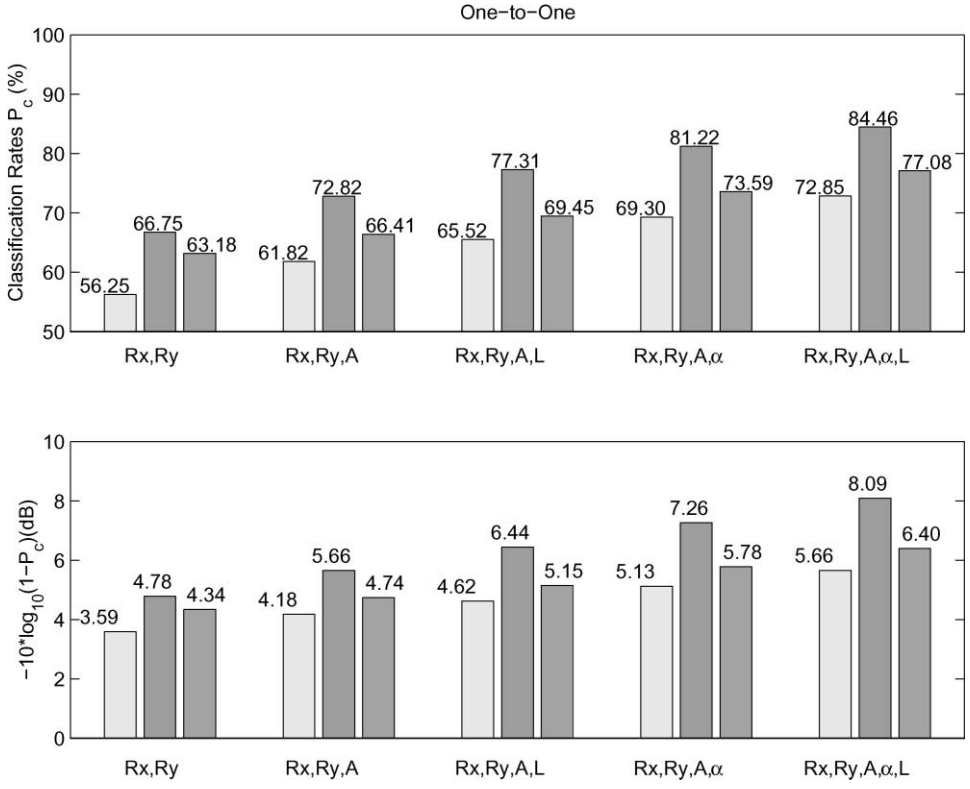


Fig. 10. Classification performance using correct (center) and erroneous location uncertainties in the one-to-one match score. The left (right) bars assume 0.5(2) times the true location uncertainty. The top figure shows average probability of correct classification ($P_c$); the bottom figure shows the same data plotted as average probability of error ($1 - P_c$) in dB.

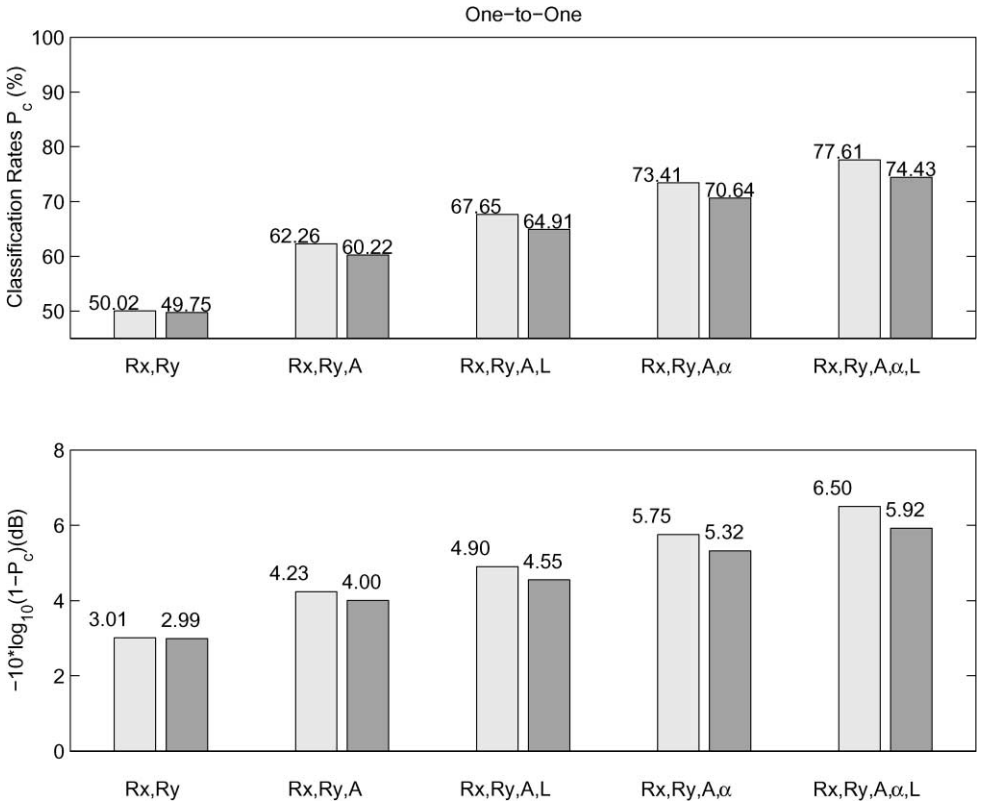


Fig. 11. Classification performance using correct (left) and erroneous (right) detection probabilities and false alarm rates in the one-to-one match score.

Table 3
Average CPU time (ms) 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

| | One-to-one | Greedy1 | Greedy2 | Many-to-many |
|---|---|---|---|---|
| Using ($R_x$, $R_y$) | 73.28 | 31.63 | 28.40 | 14.78 |
| Using five attributes | 92.00 | 49.21 | 45.31 | 32.02 |

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.

The CPU time needed to compute likelihood scores are summarized in Table 3. Shown are the average computation times for 1000 matches using the many-to-many, one-to-one, and two suboptimal 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. All four match scores are computed in polynomial time; $O(mn)$ computations are needed for the many-to-many and suboptimal one-to-one match scores, and $O((m+n)^3)$ computations for the one-to-one match score. The computation of the entries in Tables 1 or 2 comprise a significant fraction of the total compute time for this example.

Although class means are based on empirical data, the predicted classification performance is highly dependent on the uncertainty model used for the features; raising or lowering uncertainties of a particular feature impact significantly the overall classification performance. The actual uncertainties of the scattering attributes (especially the curvature and length attributes) have not been empirically characterized at present.

## 7. Conclusions

A Bayesian formalism leads to tractable hypothesis testing in model-based object recognition using unordered feature sets. Generically, a feature is merely an

ordered list; for example, a feature may be a point with associated attributes. Further, attributes may be either real-valued statistics or categorical variables. 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.

Complexity of the Bayes matching is polynomial in the problem size. Let $n$ and $m$ denote the number of measured and predicted features, respectively. Computation of a likelihood requires an $O(mn)$ construction of a tableaux (see Table 2) and an $O((m+n)^3)$ selection, using the Hungarian algorithm, of the optimal one-to-one feature correspondence. Heuristics provide an $O(mn)$ construction of two suboptimal one-to-one correspondences.

Tractability of the likelihood calculations in the Bayes approach requires only conditional independence assumptions. First, feature uncertainties are assumed to be conditionally independent given the hypothesis. Second, the omissions of features in a measurement and the appearance of spurious features are likewise assumed to be conditionally independent events. 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.

## References


[1] H.-C. Chiang, R.L. Moses, L.C. Potter, Model-based classification of radar images, IEEE Trans. Inform. Theory 46 (5) (2000) 1842–1854.

[2] L.P. Shapiro, R.M. Haralick, Structural description and inexact matching, IEEE Trans. Pattern Anal. Mach. Intell. 3 (1981) 504–519.

[3] K. Boyer, A. Kak, Structural stereopsis for 3d vision, IEEE Trans. Pattern Anal. Mach. Intell. 10 (1988) 144–166.

[4] W. Christmas, J. Kittler, M. Petrou, Structural matching in computer vision using probabilistic relaxation, IEEE Trans. Pattern Anal. Mach. Intell. 17 (1995) 749–764.

[5] I. Rigoutsos, R. Hummel, A Bayesian approach to model matching with geometric hashing, Comput. Vision Image Understanding 62 (1995) 11–26.

[6] R.C. Wilson, E.R. Hancock, Structural matching by discrete relaxation, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 634–648.

[7] R. Myers, R.C. Wilson, E.R. Hancock, Bayesian graph edit distance, Proceedings of the Tenth International Conference on Image Analysis and Processing, Venice, Italy, September 27–29, 1999, pp. 1166–1171.

[8] G.J. Ettinger, G.A. Klanderman, W.M. Wells, W.E.L. Grimson, A probabilistic optimization approach to SAR feature matching, SPIE 2757 (1996) 318–329.

[9] L.L. Scharf, Statistical Signal Processing, Detection, Estimation, and Time Series Analysis, Addison-Wesley, Reading, MA, 1991.

[10] H.-C. Chiang, Feature based classification with application to synthetic aperture radar, Ph.D. Thesis, The Ohio State University, Columbus, OH, 1999.

[11] C.H. Papadimitriou, K. Steiglitz, Combinatorial Optimization Algorithm and Complexity, Prentice-Hall, Englewood Cliffs, NJ, 1982.

[12] C. Curlander, R.N. McDonough, Synthetic Aperture Radar: Systems and Signal Processing, Wiley, New York, 1991.

[13] L.M. Novak, G.J. Owirka, C.M. Netishen, Radar target identification using spatial matched filters, Pattern Recognition 27 (4) (1994) 607–617.

[14] J. Wissinger, R. Washburn, D. Morgan, C. Chong, N. Friedland, A. Nowicki, R. Fung, Search algorithms for model-based SAR ATR, SPIE 2757 (1996) 279–293.

[15] E.R. Keydel, S.W. Lee, Signature prediction for model-based automatic target recognition, SPIE 2757 (1996) 306–317.

[16] G. Franceschetti, A. Iodice, M. Tesauro, From image processing to feature processing, Signal Processing 60 (1997) 51–63.

[17] M.J. Gerry, Two-dimensional inverse scattering based on the GTD model, Ph.D. Thesis, The Ohio State University, Columbus, OH, 1997.

[18] M.J. Gerry, L.C. Potter, I.J. Gupta, A. van der Merwe, A parametric model for synthetic aperture radar measurements, IEEE Trans. Antennas Propagation 47 (1999) 1179–1188.

[19] J.B. Keller, Geometrical theory of diffraction, J. Opt. Soc. Am. 52 (1962) 116–130.

[20] R.G. Kouyoumjian, P.H. Pathak, A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface, Proc. IEEE 62 (1974) 1448–1461.

[21] M.A. Koets, R.L. Moses, Image domain feature extraction from synthetic aperture imagery, Proceedings of the 1999 International Conference on Acoustics, Speech, and Signal Processing, Phoenix, AZ, March 15–19, 1999, pp. 2319–2322.

[22] MSTAR Public Target Database. http://www.mbvlab.wpafb.af.mil/public/sdms/datasets/mstar.


**About the Author**—HUNG-CHIH CHIANG received his B.S. degree in Communication Engineering from National Chiao-Tung University, Taiwan, in 1989, and his M.S. and Ph.D. degrees from the Ohio State University in 1994 and 1999, respectively. Since 1999 he has been employed by Etrend Electronics Inc., Taiwan, where he participates in the design of integrated circuits for consumer electronics.

**About the Author**—RANDOLPH L. MOSES received the B.S., M.S., and Ph.D. degrees in Electrical Engineering from Virginia Polytechnic Institute and State University in 1979, 1980, and 1984, respectively. During the summer of 1983 he was a SCEEE Summer

Faculty Research Fellow at Rome Air Development Center, Rome, NY. From 1984 to 1985 he was with the Eindhoven University of Technology, Eindhoven, The Netherlands, as a NATO Postdoctoral Fellow. Since 1985 he has been with the Department of Electrical Engineering, The Ohio State University, and is currently a Professor there. During 1994–95 he was on sabbatical leave as a visiting researcher at the System and Control Group at Uppsala University in Sweden. His research interests are in digital signal processing, and include parametric time series analysis, radar signal processing, sensor array processing, and communications systems. Dr. Moses served on the Technical Committee on Statistical Signal and Array Processing of the IEEE Signal Processing Society from 1991–94. He is a member of Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi, and Sigma Xi.

**About the Author**—LEE C. POTTER received the B.E. degree from Vanderbilt University and M.S. and Ph.D. degrees from the University of Illnois, Urbana, all in Electrical Enginnering. Since 1991 he has been with the Department of Electrical Engineering at The Ohio State University where he is currently Associate Professor. His research interests include statistical signal processing, inverse problems, detection, and estimation, with applications in radar target identification and ultra wide-band systems. Dr. Potter is a 1993 recipient of the Ohio State College of Engineering MacQuigg Award for Outstanding Teaching.