A Parametric Attributed Scattering Center Model for SAR Automatic Target Recognition*

Randolph L. Moses, Lee C. Potter
Hung-Chih Chiang, Michael A. Koets, and Ashutosh Sabharwal
Department of Electrical Engineering
The Ohio State University
Columbus, Ohio 43210
E-MAIL: {moses,potter,chiangh,koetsm,sabhara}@ee.eng.ohio-state.edu
HOMEPAGE: http://er4www.eng.ohio-state.edu/ips/research/Projects/darpa/darpa.html

Abstract

We present a parametric attributed scattering model for Synthetic Aperture Radar imagery. The model characterizes both frequency and aspect dependence of scattering centers. We present algorithms for estimating the model parameters from SAR image chips, and propose model order estimation algorithms that exploit nested model structures. We develop a Bayes classifier for the extracted model parameters; the classifier uses uncertainty in both extracted and predicted features. Numerical results on synthetic and measured SAR data validate the model and show encouraging results in both the ability to accurately extract scattering attributes and the utility of these attributes for improved discriminability of target classes.

1 Introduction

This paper describes our research on the use of a parametric scattering model for Synthetic Aperture Radar (SAR) Automatic Target Recognition (ATR). In this work we propose a parametric attributed scattering center model to describe the backscattered response from a region measured by a SAR sensor.

*This work was sponsored by the Defense Advanced Research Projects Agency under contract F33615-97-1020 monitored by Wright Laboratory. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the United States Government.

At high frequencies, the scattering response of an object is well approximated as a sum of responses from individual scattering centers[Keller, 1962]. These scatterers provide a concise, physically relevant description of the object and are thus good candidates for use in target recognition, radar data compression, and scattering phenomenology studies.

The proposed scattering model incorporates both frequency and aspect dependence of scattering centers. The model is based on dominant responses of monostatic scattering solutions from both physical optics and the geometric theory of diffraction. We model each scattering center by a set of parameters describing its location, shape, orientation (pose) and amplitude. This extends current SAR scattering center models, most of which implicitly assume scattering centers are localized or isolated points [Tu *et al.*, 1997, Sacchini *et al.*, 1993]. The proposed model generalizes the point scattering model, and provides a richer description of scattering behavior.

We develop algorithms for estimating the model parameters, or features, from measured SAR image chips. The algorithms estimate parameters directly from high energy regions in the image; this facilitates insertion into SAR ATR data processing streams and also provides robustness to any assumed clutter model (for example, it reduces degradation of estimates due to a nearby tree in the image chip). We have developed an approximate maximum likelihood algorithm that uses an iterative descent tech-

nique, and also a suboptimal but computationally fast estimator. We characterize feature uncertainty both by lower bounds and by simulations on algorithm performance.

We also address the problem of model order estimation. We propose a computationally efficient and statistically consistent method that exploits a nested model structure; such a structure appears in the scattering center models we consider in this work.

We develop a Bayes classifier that operates on extracted model features. The classifier assumes uncertainty in both the extracted features and the predicted catalog features. We use the Bayes classifier to predict ATR performance, and we describe our initial ATR performance estimation results using attributed scattering features.

2 Attributed Scattering Model

We have developed a parametric model that describes backscattering from objects measured at high frequencies [Gerry, 1997]. The model attempts to achieve high fidelity of scattering while remaining sufficiently simple and parsimonious to permit robust inference from estimated parameters.

The attributed scattering model assumes that the total scattered field as a function of frequency f and aspect ϕ is a sum of p individual scattering terms:

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

Each scattering center is modeled parameterized as

$$\begin{aligned} E_k^s(f,\phi) &= A_k \left(j\frac{f}{f_c}\right)^{\alpha_k} \exp(-2\pi f \gamma_k \sin\phi) \\ &\quad \cdot \mathrm{sinc}\left(\frac{2\pi f}{c} L_k \sin(\phi - \bar{\phi}_k)\right) \\ &\quad \cdot \exp\left(j\frac{4\pi f}{c}(x_k \cos\phi + y_k \sin\phi)\right) \end{aligned}$$

where (x_k, y_k) denote the scatterer location, A_k is its amplitude, α_k is its frequency dependence, L_k and $\bar{\phi}_k$ are the length and orientation of distributed scatterers, and γ_k is the aspect dependence of localized scatterers. The scattering model is thus described by the parameter set $(A_k, x_k, y_k, \alpha_k, \gamma_k, L_k, \bar{\phi}_k)$ for $k = 1, ..., p$. Each scattering center is either localized or distributed; for localized scattering centers $L_k = \bar{\phi}_k = 0$ and γ_k characterizes the (mild) aspect dependence of the scattering center. For distributed scattering centers aspect dependence is described by the pair $(L_k, \bar{\phi}_k)$ with $L_k > 0$, and we set $\gamma_k = 0$.

Table 1: Parameters α and L serve to discriminate many scattering geometries.

Example scattering geometries	α	L
dihedral	1	$L \neq 0$
corner reflector	1	0
cylinder	$\frac{1}{2}$	$L \neq 0$
sphere	0	0
edge broadside	0	$L \neq 0$
corner diffraction	-1	0
double corner diffraction	-2	0

The point scattering model can be seen as a special case of the attributed scattering center model with $\gamma_k = L_k = \alpha_k = 0$. The point scattering model contains no aspect and frequency dependence description of scatterers. Several canonical scattering geometries are, however, distinguishable by differences in frequency and aspect dependence of their backscattered response; example geometries distinguishable by their (α, L) parameters are shown in Table 1. The richer physical description afforded by these scattering attributes motivates an investigation of their utility for improved discriminability of targets and improved ATR performance.

3 Feature Extraction

3.1 Estimation Algorithms

We have developed two algorithms for estimating the attributed scattering center model parameters from SAR image chips. One is an approximate maximum likelihood (AML) algorithm, and is based on weighted least squares

fitting of the model to regions of high energy in the measured SAR image. The algorithm recursively estimates and subtracts contributions from small clusters of scattering centers, using an iterative descent from initial parameter estimates to minimize a nonlinear cost function. It has the interpretation of being approximate ML for a Gaussian clutter assumption on the measured imagery. The second algorithm is computationally faster than the approximate ML algorithm at the expense of an increase in the variance of estimated parameters. This second algorithm is a variation of the initial estimation stage of the AML algorithm.

Significantly, both algorithms operate directly on SAR image chips and fit models only on regions of high backscattered energy. Processing on image chips facilitates insertion into SAR ATR data processing streams, because post-detection processing is often applied to small chips containing detected regions of interest. By model fitting only on regions of the image, we realize robustness to any assumed clutter model; for example, we reduce uncertainty or bias in feature estimates that might be caused by large nearby clutter scattering that is not well modeled as Gaussian noise.

To estimate parameters from image chip data, we need a model of the SAR image formation process. We assume scattering data is collected as a function of frequency f and aspect ϕ as in equation (1). The data are interpolated to a rectangular grid using linear filtering, zero padded, multiplied by a window to reduce sidelobes, then inverse Fourier transformed [Koets, 1998].

We formulate the estimation problem in a maximum likelihood framework. Let D denote the measured SAR image pixels in the (x, y) plane, stacked as a vector, and let d be its corresponding vector in the (f, ϕ) domain; thus, $D = Bd$ where B is a linear operator encompassing the image formation procedure. We assume the measured vector d is the sum of the model in equation (1) and an additive Gaussian noise vector n with covariance Σ:

$$d = \sum_{k=1}^{p} m(f, \phi; \theta_k) + n, \quad n \sim \mathcal{N}(0, \Sigma) \qquad (2)$$

where $m(f, \phi; \theta_k)$ denotes the kth scattering center model as a function of (f, ϕ). Then it follows that

$$D = \sum_{i=k}^{p} M(x, y; \theta_k) + N \qquad (3)$$

where $N \sim \mathcal{N}(0, B\Sigma B^H)$ and $M(x, y; \theta_k)$ denotes the kth scattering center model after transformation to the image domain (x, y). Since D is a Gaussian random vector whose mean depends nonlinearly on the parameter vector $\theta = [\theta_1^T, \ldots, \theta_p^T]^T$, it follows that the maximum likelihood estimate of θ can be found by a weighted least squares minimization

$$\hat{\theta}_{ML} = \arg\min_{\theta} \|D - M(x, y; \theta)\|_W^2 \qquad (4)$$

where $M(x, y; \theta) = \sum_{i=k}^{p} M(x, y; \theta_k)$ and $W = (B\Sigma B^H)^{-1}$.

The above minimization is difficult because the minimization is performed over all image pixels and on a parameter vector θ of high dimension. We can effectively approximate this minimization by exploiting the fact that scattering centers have nearly all of their signal energy concentrated in a small region of the image plane. We realize this approximation by fitting models of small order to regions of high energy in the image plane as an approximation to the minimization in equation (4) [Koets, 1998].

Our algorithm consists of two stages. The first stage isolates regions of high energy in SAR images, classifies those regions as distributed or localized scattering centers, and generates initial estimates of the parameter values. The second stage of the algorithm uses these estimates to initialize an iterative nonlinear descent routine that generates approximate maximum likelihood estimates for the parameter values. We have implemented a second algorithm which employs a variation of the first stage of parameter estimation and skips the computationally intensive optimization stage. This vari-

ation requires two orders of magnitude less computation, and gives less accurate parameter estimates.

Our algorithm is recursive and uses an approach based on the CLEAN algorithm [Tsao and Steinberg, 1988]. In each of several iterations we estimate the parameters of scattering centers in a region of the image, simulate an image using the parametric model and the estimated parameters, and subtract the simulated scattering center from the measured data, removing the most recently processed peak and its sidelobes from the image. The algorithm terminates when a specified number of scattering centers have been processed, when a specified amount of the energy in the original images has been removed, or when the peak in the residual data falls a specified level below the original peak.

The faster variant on the algorithm incorporates only slight modifications to the first stage of the algorithm. We modify the segmentation and order selection procedures to produce regions that contain only a single scattering center. This variation of the algorithm removes a processed peak in the image by setting the pixels of the corresponding region to zero since the parameter estimates are sometimes not accurate enough for subtractive removal to be effective.

We use a segmentation process based on a watershed algorithm [Stach and LeBaron, 1996] to select regions in the image that contain a small number of scattering centers. Segmentation is carried out in two stages. We generate an initial segmentation which isolates each local peak in the image. Further processing combines these regions where it is necessary to construct a region that accurately reflects the extent of a small number of scattering centers. Model order and scattering center structure (localized or distributed) are estimated from the shape of the combined region and the number of initial segments combined [Koets, 1998].

We compute initial estimates for the parameter values from the measured data in the image region or assign initial values based on knowledge of the range of values that are reasonable to represent the scattering mechanisms. We initialize the x and y parameters using a center of mass of the segments. We estimate L from a best fit quadratic to the sinc function found in the FFT of a one dimensional slice of the image data through the center of mass of the selected image region. The γ parameter is initialized to zero. We initialize $\overline{\phi}$ as the tilt angle of the best fit ellipse to the pixels of the image region. The α parameter is drawn from a small set of discrete values, and is initialized by exhaustive search over $\alpha \in [-1, -1/2, 0, 1/2, 1]$. The amplitude estimate is initialized using a linear least squares fit over the pixels of the image region. For the fast variation of the algorithm we estimate only $|A|$ by fitting to the magnitude of the measured data.

We use a standard iterative descent technique to minimize (4) starting from the initial estimates. These methods will not determine the globally optimum parameter set for an arbitrary initial guess, so it is important that our initial parameter estimate be near the global minimum.

We have computed the Cramér-Rao bound (CRB) on estimator variance for the parameters of our model [Gerry, 1997, Koets, 1998]. The CRB provides a lower bound on the variance of unbiased estimators. Since our estimator is an approximate ML estimator we expect it to be at least asymptotically unbiased and efficient.

3.2 Experimental Results

We present feature extraction experimental results on three data sets: 1) synthesized model data, 2) XpatchF synthesized data of the SLICY target, and 3) MSTAR measurements of a T72 image chip.

First, to validate the estimation algorithms and to test the relative statistical efficiency of the approximate maximum likelihood method, we applied the algorithm to both localized and distributed scattering centers at several resolutions using the parametric model. The scattering center models were chosen to closely approximate scattering centers on the SLICY target described below. We then added correlated Gaussian noise to the images and estimated the parameters of the scattering centers using the ap-

proximate ML variation of the algorithm. We performed 50 such trials for each combination of scattering center type, resolution, and signal to noise ratio. Our parameter estimators essentially achieve the CRB in every case.

We next present results of feature estimation from synthetic images of the SLICY geometric test target that were generated using the XpatchF electromagnetic prediction package. These images were generated at several resolutions. When applied to this data, the algorithms accurately estimate the physically characteristics of the localized and distributed scattering centers present on the target object.

We then added noise to the synthetic images and estimated the scattering center parameters with both the fast and the approximate ML variations of the algorithm. Example results are shown in Figures 1 and 2 for the trihedral scatterer on the SLICY target. The observed variances of the parameter estimates and the CRB for the x, y, (in units of inch^2) and α parameters of this scattering center in the synthetic data are shown in Figure 1. This experiment was conducted with 30 dB SNR where SNR is defined as the ratio of the peak pixel in the noiseless data to the standard deviation of the noise. Figure 2 shows the observed variances and the CRB for the x, y, and α parameters of a localized scattering center in the 6" resolution data at several SNR values. In both cases we see good agreement with the CRB for the AML algorithm, and somewhat higher variances for the fast estimation algorithm; however, even in the fast algorithm the standard deviations of scattering center locations are less than 0.1 inch for all noise levels and resolutions.

Finally, we have applied our algorithms to measured SAR images of vehicles measured under the MSTAR program. Figure 3 shows a measured SAR image of a T-72 tank. We estimated the parameters of 33 scattering centers in this image. We used the estimated parameters with the parametric model to form the reconstruction of the image shown in Figure 4.

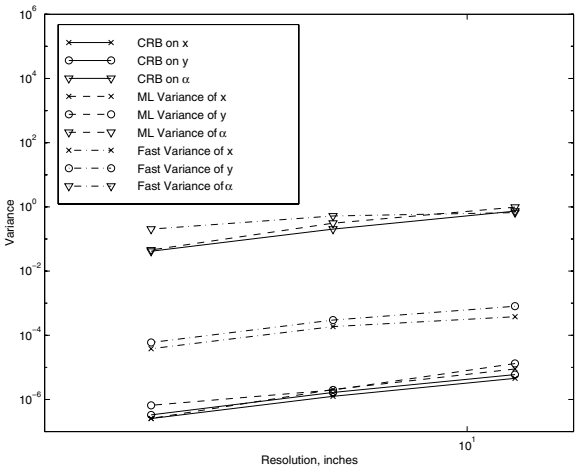

Figure 1: Scattering center parameter variances (inch^2) and CRB for synthetic SLICY trihedral, for 3", 6", and 12" SAR resolutions.

4 Model Selection for Nested Model Classes

4.1 Introduction

An important problem in parametric model fitting is that of estimating the model order. The maximum likelihood (ML) principle is well-known for estimating unknown fixed parameters when the dimension of the model is known. In applications where the parameter dimension is also unknown, minimum description length (MDL) [Rissanen, 1978, Rissanen, 1996] has been widely proposed [Wax and Kailath, 1985, Barron and Cover, 1991]. The implementation aspects of MDL-based methods, such as computational complexity, have received little attention. For instance, MDL-based methods require maximum likelihood estimates of model parameters for each of the hypothesized models. However, many engineering problems involve nested models, where the simpler model can be embedded in the more complex one to form a nesting.

We present an parameter estimation technique that uses the Wald statistic [Wald, 1943] to estimate MDL cost of each of the individual models in the hypothesized class of nested models. The major advantage in employing the Wald statistic is that only the parameter estimates of the

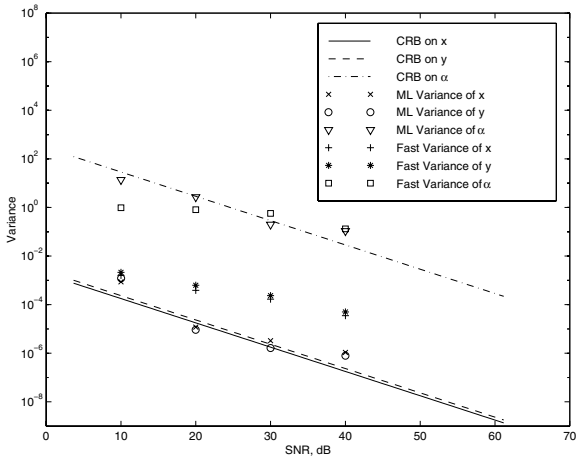

Figure 2: Scattering center parameter variances and CRB for synthetic SLICY trihedral; 6" resolution and varying SNR.

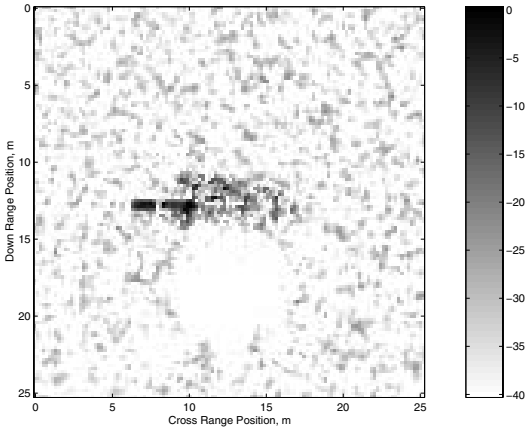

Figure 3: Measured SAR Image of T-72 Tank, in dB.

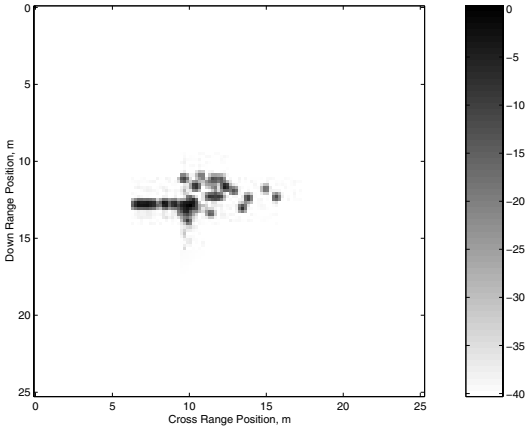

Figure 4: Reconstruction of T-72 tank image from estimated model parameters, in dB.

most complex model are required. The Wald statistic can intuitively be understood as a distance of ML estimates from the parameter set of the simpler model.

We establish the statistical consistency of the proposed order selection method and the consistency of the corresponding parameter estimates. Also, a dynamic programming approach is used with the Wald statistic to minimize the expected computational cost of order selection. Computed performance is presented for estimating scattering centers. Results are presented for nonlinear regression models in Gaussian noise, but are also applicable to the i.i.d. or non-Gaussian noise cases.

4.2 Notation

Consider nonlinear regression models in additive white Gaussian noise

$$x_t = g_t(\theta) + \epsilon_t, \quad t = 1, 2, \ldots \tag{5}$$

where $g_t(\theta)$ is a known real-valued continuous function defined on a compact set $\overline{\Theta}_k \subset \mathcal{R}^k$. The noise ϵ_t is assumed to be Gaussian with zero mean and unknown variance σ^2. The parameter space for each model order j is denoted by Θ_j and is assumed to be a bounded open subset of $\mathcal{R}^j$. For each $j \leq k$, $\Theta_{j,k}$ represents the embedding of Θ_j in $\mathcal{R}^k$. Further, we assume that $\Theta_{j,k} \cap \Theta_{l,k} = \emptyset, \forall\ l \neq j$, and $\Theta_{j,k} \subset \overline{\Theta}_{l,k}$ for all $j \leq l$. Also, assume that the parameter set of a simpler model $j < k$, $\Theta_{j,k}$, can be obtained by imposing $k-j$ continuous restrictions, $r_{j,k}(\theta_k) = 0$, on the parameters of model k, *i.e.*,

$$\Theta_{j,k} = \{\theta_k : r_{j,k}(\theta_k) = 0, \theta_k \in \overline{\Theta}_k\}. \tag{6}$$

The restrictions are such that $R_{j,k}(\theta) = \frac{\partial r_{j,k}}{\partial \theta_k^T}$ is full rank for all $\theta_k \in \overline{\Theta}_k$. If the parameter sets for different model orders form a nesting, then the restrictions defining them are also nested. Where k is clear from the context, we will use θ_j, Θ_j, r_j and R_j to represent $\theta_{j,k}$, $\Theta_{j,k}$, $r_{j,k}$ and

$R_{j,k}$, respectively.

We assume that the model order is bounded by K. The true model order and parameters are denoted by k^o and $\theta^o_{k^o,K}$, respectively. The prior probability density on model order is denoted by $\pi(k)$, and the prior on model parameters for model order k is denoted by $w_k(\theta), \theta \in \Theta_k$. The model, $f(x|\theta), \theta \in \overline{\Theta}_K$ is defined on a measurable space with respect to a fixed sigma-finite measure $\lambda(dx)$. The model is assumed to be regular, *i.e.*, the Fisher information matrix is full rank, at all $\theta_k \in \Theta_k$. For clarity, we use $\pi(\cdot)$ to denote pdf for a discrete random variable and $f(\cdot)$ for a continuous random variable. Finally, we let $x^N = [x_1, x_2, \ldots, x_N]^T$ denote the observed data vector and $\hat{\theta}_K \in \overline{\Theta}_K$ denote the maximum likelihood (ML) estimate of $\theta^o_{k,K} \in \Theta_{k,K}$.

4.3 Proposed Method

Consider the large sample form of Stochastic Complexity [Rissanen, 1978],

$$SC(k) = -\log f\left(x|\hat{\theta}_k\right) + \frac{k}{2}\log N + O(\log N)$$

where N is the number of data samples. For $k > j$,

$$\begin{aligned} SC(j) - SC(k) &= -\log \frac{f\left(x|\hat{\theta}_j\right)}{f\left(x|\hat{\theta}_k\right)} \\ &\quad + \frac{j-k}{2}\log N + O(\log N) \end{aligned} \tag{7}$$

We replace the first term in equation (7) with the generalized Wald statistic [Hadi and Wells, 1990], yielding an estimate of $SC(j)$

$$\begin{aligned} \widehat{SC}_k(j) &= SC(k) + r_j^T(\hat{\theta}_k)B(\hat{\theta}_k)r_j(\hat{\theta}_k) \\ &\quad + \frac{j-k}{2}\log N + O(\log N) \\ &= SC(k) + W_{j,k} \\ &\quad + \frac{j-k}{2}\log N + O(\log N) \end{aligned} \tag{8}$$

Here $B_j(\hat{\theta}_k) = \left(R_j^T(\hat{\theta}_k)\tilde{I}^+(\hat{\theta}_k)R_j(\hat{\theta}_k)\right)^+$ where $\tilde{I}(\hat{\theta}_k)$ is an estimate of the Fisher information matrix computed at the true parameter value, $I(\theta^o_{k_o})$. Further, A^+ is the Moore-Penrose pseudo-inverse of A. Based on (8), the model order is estimated as

$$\begin{aligned} \hat{k}_N &= \arg \min_{j=1,\ldots,K} \widehat{SC}_K(j) \\ &= \arg \min_{j=1,\ldots,K} W_{j,K} + \frac{j}{2}\log N \end{aligned} \tag{9}$$

Note from (8) that the Wald statistic requires the ML parameter estimates of only the more complex model k. Thus, given the ML parameter estimates for model k, the stochastic complexity of all the lower dimensional models can be estimated using (8). We note that $W_{j,k}$ is a non-increasing sequence in j [Sabharwal and Potter, 1998].

4.4 Consistency of Proposed Order Selection

To prove that the proposed method (9) produces consistent model order estimates, we require the consistency of the overparameterized ML estimates. Since we are interested in the case where the Fisher information at true parameter value, $\theta^o_{j,k}$, may have rank less than k, the known results on ML consistency [Jennrich, 1969] cannot be directly applied. We extend the results of [Jennrich, 1969] to singular Fisher information case, in nonlinear regression, under the assumption of compactness of the parameter space. The case with singular Fisher information is of practical utility rather than mere mathematical generalization. A number of engineering models when overparameterized have singular Fisher information at the true parameter; for instance, damped exponential model, ARMA [Veres, 1985], ARMAX [Klein and Spreij, 1996] and multilayer perceptron networks [Fukumizu, 1996]. We note that the results in this section on consistency of overparameterized ML estimates are directly applicable for complex valued g_t, and to zero mean finite variance non-Gaussian errors, ϵ_t in equation (5).

Theorem 1 (*Consistency*) *Let the true model order be denoted by k^o. If either of the following conditions are satisfied, then* $\lim_{N\to\infty} Pr\left(\hat{k}^N = k^o\right) = 1$.

1. *the Fisher information matrix is full rank at all* $\theta_K \in \overline{\Theta}_K$ *and* $\hat{\theta}_K \xrightarrow{a.s.} \theta_K^o$. *Further, the derivatives*

$$g_t^i = \frac{\partial g_t(\theta)}{\partial \theta^i}, \text{ and } g_t^{i,j} = \frac{\partial^2 g_t(\theta)}{\partial \theta^i \partial \theta^j}$$

exist, and are continuous on $\overline{\Theta}_K$. *Lastly, all the tail cross products of form* $[v, u]$ *exist, where* $v, u = g, g^i, g^{i,j}$.

2. *the Fisher information matrix is rank deficient at* $\theta_{k^o,K} \in \overline{\Theta}_K$, $K > k^o$, *such that* $rank(I(\theta_{k^o,K})) = k^o$. *Let* $\tilde{I}(\hat{\theta}_K)$ *denote an estimator of* $I(\theta_{k^o}^o)$, *with* $\hat{\theta}_K \xrightarrow{a.s.1} \theta_{k^o}^o \in \mathcal{E}_o$. *Assume that the Fisher information for each of the identifiable parameter grows as* $O(N^\alpha)$, $\alpha \geq 1$. *Further, assume that* $\frac{1}{N^\alpha}\tilde{I}(\hat{\theta}_k) \rightarrow \frac{1}{N^\alpha} I(\theta_{k^o}^o)$, $\tilde{I}(\theta)$ *is continuous function of* θ, *and*

$$\lim_{N\to\infty} Pr\left(rank(\tilde{I}(\hat{\theta}_K)) = rank(I(\theta_{k^o}^o))\right) = 1.$$

Proof [Sabharwal and Potter, 1998].

4.5 Asymptotic Rules for Cost Minimization

The Wald statistic asymptotically reveals model order for overparametrized models; implementation requires selection of a hypothesized model order and computation of the corresponding ML parameter estimates. A dynamic programming approach is used to select candidate model orders such that average computational cost is minimized. Let k^o denote the true model order which is distributed with prior probability distribution, $\pi(k^o)$. At a cost $c(j)$, we can compute ML estimates for order j. Using the Wald statistic, we (asymptotically) receive one of the two answers regarding model order:

A1: the true value $k^o < j$, or

A2: the knowledge that k^o equals or exceeds j.

We seek the decision strategy to minimize the expected cost. Let π' denote the prior probability distribution on $[1, \ldots, K]$ when parameters are estimated for order j. Let $R(j|\pi')$ denote the expected risk of arriving at the correct answer if order j is tried, then the following recursion is obtained using dynamic programming principles:

$$\begin{aligned} R(j|\pi') &= \Pr(k < j|\pi')c(j) + \Pr(k \geq j|\pi') \\ &\quad \cdot \left(c(j) + \min_{m>j} R(m|\rho(\pi')\right) \end{aligned}$$

where $R(K|\pi') = c(K)$ and $\rho(\pi')$ is the posterior probability distribution on k given that order j estimates result in answer (A2).

4.6 Superimposed Exponential Signals

We apply the Wald-statistic based approach for order selection and parameter estimation in superimposed exponential models,

$$x_t = \sum_{i=1}^{k^o} a_i e^{\jmath\phi_i} \left(\upsilon_i e^{-\jmath\omega_i}\right)^t + \epsilon_t \qquad (10)$$

for $t = 1, \ldots, N$, where ϵ_t is normally distributed with mean zero and variance σ^2 and $\upsilon_i \equiv 1$. The model order k^o, the model parameters $\{(a_i, \phi_i, \omega_i)\}_{i=1}^{k^o}$, and the noise variance σ^2 are the unknown parameters in (10). For 1 foot resolution X-band and K-band radars, equation (10) is an excellent approximation for far-field electromagnetic scattering predicted by the Geometric Theory of Diffraction [Gerry, 1997]. To apply the Wald statistic, we have verified the required assumptions for model (10) and have derived restrictions $r_{j,k}(\theta_k) = 0$.

Monte-Carlo simulations are computed for a controlled comparison of the proposed order selection procedure to MDL and MAP. In all the results, the true model order was uniformly distributed on $[1, \ldots, 7]$. As a comparison, model order estimates using MDL and MAP were computed using

$$\begin{aligned} \hat{k}_{MDL} &= \arg\min_k f(x|\hat{\theta}_k) + \frac{3k}{2}\log n \\ \hat{k}_{MAP} &= \arg\min_k f(x|\hat{\theta}_k) + \frac{5k}{2}\log n \end{aligned}$$

A low computation SVD-based procedure from [**?**] is used for computing ML parameter estimates. With uniform prior on the true model

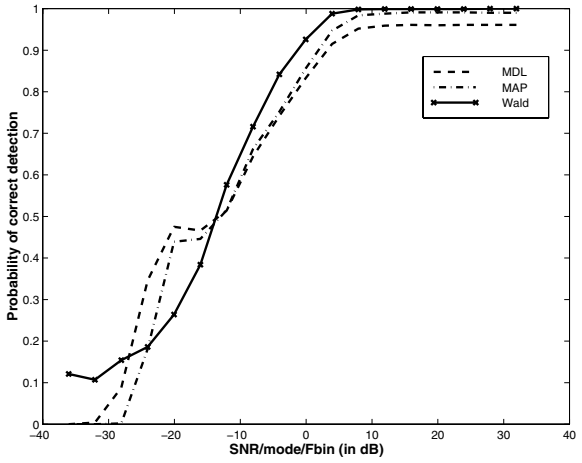

Figure 5: Probability of correct detection of model order as a function of SNR

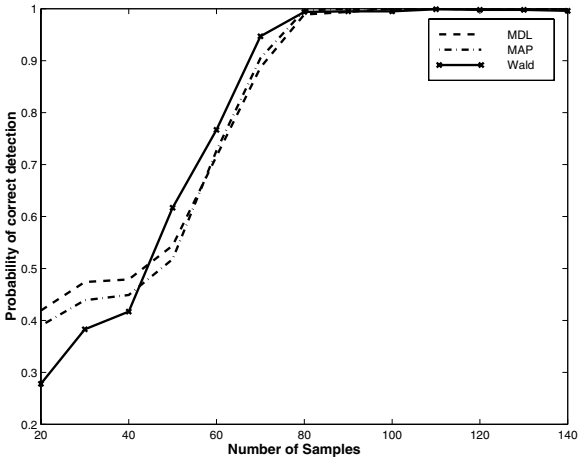

Figure 6: Probability of correct detection of model order as a function of data length

order, the asymptotically optimal rule for model selection using the Wald statistic is to compute ML estimates for model order $k = 7$. If, for example, $K = 15$, then the asymptotically optimal rule is to: (a) compute ML estimates for model order 13, (b) if $\hat{k} < 13$, stop else compute ML estimates for model order 15 and estimate $\hat{k}$.

In Figure 5, the performance of the order selection using the three methods is shown as a function of SNR/mode/Fbin. The average computational cost of MDL and MAP is more than three times the proposed method, with similar detection performance.

In Figure 6, the detection performance is shown as a function of number of samples, N, with SNR/mode $= -15$ dB. The ratio of the average computational cost of MDL and MAP to the average cost of the proposed method is shown in Figure 7.

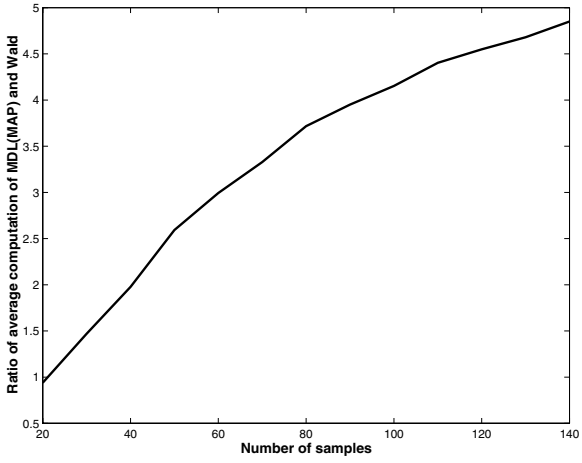

Figure 7: Ratio of average computational complexity $\left(\frac{C_{MDL/MAP}}{C_{Wald}}\right)$ as a function of data length

5 Classification

We are interested in evaluating ATR performance when attributed scattering center model parameters are used in a feature-based classification scheme. In particular, we are interested in quantifying the ATR performance improvement of frequency and aspect dependence of the model, relative to a baseline of point scattering

features.

In order to evaluate ATR performance, we have developed a Bayes classifier that operates on extracted attributed scattering center model features, and compares to a catalog of predicted model features. The classifier assumes feature uncertainty in both the extracted and predicted model parameters.

We propose the statistical models for attributed scattering features as follows: Let H denote a set of hypotheses (including type, pose, articulation, and configuration), x denote a predicted feature vector, y denote an extracted feature vector, x_i (y_i) denote the i^{th} scatterer in x (y), x_{ik} (y_{ik}) denote the k^{th} attribute of x_i (y_i). Let $\mathcal{N}(M, \Sigma)$ denote a Gaussian random vector with mean M and covariance matrix Σ. Define correspondence Γ such that $\Gamma_j = i$ provided Γ maps y_j from x_i; $\Gamma_j = 0$ provided that Γ maps y_j from clutter points. We denote P_i as the probability of detection of predicted scattering center x_i.

We assume all extracted attributes are mutually independent conditioned on the corresponding predicted attributes; that is,

$$\begin{aligned} f(y_j|x_{\Gamma_j}) &= \prod_k f(y_{jk}|x_{\Gamma_j}) \\ &= \prod_k f(y_{jk}|x_{\Gamma_j k}) \end{aligned} \quad (11)$$

We also assume all extracted scatterers with correspondence are mutually independent conditioned on hypothesis H:

$$f(y, \Gamma|H) = \prod_j f(y_j, \Gamma|H)$$

We adopt a one-to-one correspondence [Ettinger *et al.*, 1996, Irving *et al.*, 1997]. Our goal is to find the conditional likelihood $f(y, \Gamma|H)$. We adopt the following likelihood equation

$$\begin{aligned} f(y, \Gamma|H) &= e^{-\lambda}\frac{\lambda^f}{m!}\prod_{\{j:\Gamma_j=0\}} p_f(y_j) \\ &\cdot \left(\prod_{\{j:\Gamma_j\neq 0\}} P_{\Gamma_j}\left(\int p(y_j|x_{\Gamma_j})p(x_{\Gamma_j})dx_{\Gamma_j}\right)\right) \\ &\cdot \prod_{k,\Gamma_j=k,\forall j}(1-P_k) \end{aligned} \quad (12)$$

where m is the total number of extracted scattering centers, and f is the total number of false scattering centers in Γ.

To compute the above probabilities, we need prior probabilities on the parameters, and we need to determine the probability density functions for the predicted features $f(x_{ik})$, the extracted features conditioned on predicted features $f(y_{jk}|x_{\Gamma_j k}), \forall j, k$, and the extracted features when extracted from clutter $f_c(y_{jk})$. We model these densities as follows:

- Amplitude: Log Normal for $|A|$, Gaussian for A.
- Location: Gaussian for $f(x_{ik})$ and $f(y_{jk}|x_{\Gamma_j k})$, 2D Poisson with rate λ for $f_c(y_{jk})$.
- Geometric Type: Since $\alpha \in [-1 \ -1/2 \ 0 \ 1/2 \ 1]$, the densities are probability mass functions.
- Length: $f_c(y_{jL}) = e \cdot \delta(\bar{y}_{jL})$ for $\bar{y}_{jL} = 0$ and $f_c(y_{jL}) = 2(1-e) \cdot g(\bar{y}_{jL}, 0, \sigma_{y_{jL}})$ for $\bar{y}_{jL} > 0$, and similarly for the other densities.
- Tilt Angle: Gaussian

From the above densities, we can compute $f(y_j|x_{\Gamma_j k})$ for various attributes assuming $\Gamma_j = i$ in a straightforward manner; for details, see [Chiang, 1998].

5.1 ATR Performance Estimation

The above likelihood formulation provides a mechanism for matching extracted to predicted model parameters. We can compute the Bayes error of a match as a function of the uncertainty of the attributes. We thus use the Bayes match formulation to predict ATR performance, and to quantify the performance as a function of feature uncertainty. An initial illustrative example is presented below.

Figure 8 shows the result of a simple experiment that motivates the use of more scattering

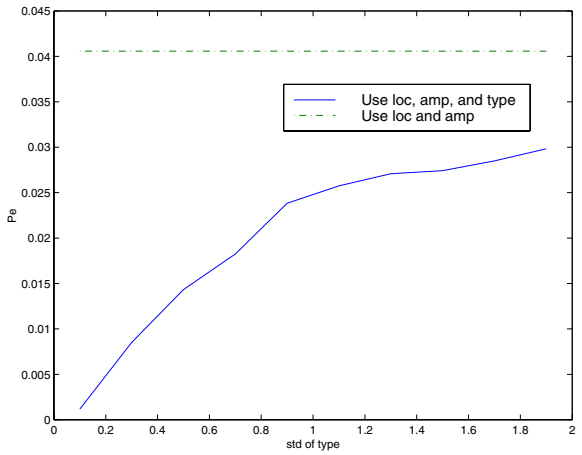

Figure 8: Bayes classification error as a function of uncertainty in frequency dependence parameter α

features in a target identification task. The experiment is a two-class classification task. The two targets are a T72 and a BMP2 from the MSTAR 1995 Public Release database. Each target catalog contains 60 images of different pose angles from 0 to 90 degrees with a resolution of 1 foot by 1 foot. For each image we extract the locations and amplitudes of the three strongest peaks using an FFT-based CLEAN technique [Tsao and Steinberg, 1988]. For each peak we randomly set the type parameter $\alpha \in \{-1, -1/2, 0, 1/2, 1\}$ with equal probabilities. The standard deviation of the location uncertainty is set to 5 feet, the standard deviation of the amplitude uncertainty is set to be equal to the extracted amplitude, and the standard deviation of the type uncertainty is varied from 0.1 to 1.9. We assume equal priors on the images. A standard two-class Bayesian classifier (a likelihood ratio test) is performed [Van Trees, 1968]. The likelihood score for a test pattern given a true pattern is computed as the maximum of the likelihood score among all possible one-to-one correspondences. For each image we generate 100 realizations. Figure 8 shows that the performance is improved by 30 % with the additional type attribute even if the relative uncertainty of the type attribute is twice as large as that of the amplitude attribute.

6 Conclusion

We have presented a GTD-based parametric scattering model for the extraction of scattering centers from radar data measured as a function of frequency and aspect angle. We have developed algorithms for extracting these features from measured SAR image chips. We have bounded feature uncertainty using the Cramér-Rao bound, and have shown that our approximate maximum likelihood estimation algorithm nearly achieves this bound in most cases. We have also presented a model order estimation method that uses the Wald statistic to provide computational savings while retaining good statistical performance. Finally, we have developed a Bayes classifier that operates on attributed scattering features. An initial ATR performance experiment shows improved classification performance, when compared to performance using only point scattering features.

Acknowledgments

The authors thank Dr. William Irving of Alphatech for his valuable input in likelihood function formulation, and to Dr. Ed Friel of Sverdrup Technology Inc. for generating XPatchF SLICY imagery.

References

[Barron and Cover, 1991] A. R. Barron and T. M. Cover. Minimum complexity density estimation. *IEEE Trans. Info. Th.*, 37(4):1034–1054, July 1991.

[Chiang and Potter, 1996] Da-Ming Chiang and Lee C. Potter. On the Equivalence of Exponential Signal Estimators. *IEEE Transactions on Signal Processing*, August 1996. submitted.

[Chiang, 1998] Hung-Chih Chiang. Bayesian matching for attributed scattering centers. Technical Report Ph.D. Research Proposal, The Ohio State University, 1998.

[Ettinger *et al.*, 1996] Gil J. Ettinger, Gregory A. Klanderman, William M. Wells, and W. Eric L. Grimson. A Probabilistic Opti-

mization Approach To SAR Feature Matching. *SPIE*, 2757:318–329, April 1996.

[Fukumizu, 1996] K. Fukumizu. A regularity condition of the information matrix of a multilayer perceptron network. *Neu. Net.*, 9:871–9, July 1996.

[Gerry, 1997] M. Gerry. *Two-dimensional inverse scattering based on the GTD model.* PhD thesis, The Ohio State University, Columbus, OH, 1997.

[Hadi and Wells, 1990] A. S. Hadi and M. T. Wells. A note on generalized Wald's method. *Metrika*, 37:309–315, 1990.

[Irving *et al.*, 1997] William Irving, John Wissinger, Gil Ettinger, Ronald Chaney, and Greg Klanderman. Analysis of atr performance degradation in preference of bayesian modeling errors. Technical report, June 1997.

[Jennrich, 1969] R. I. Jennrich. Asymptotic properties of non-linear least squares estimators. *Ann. Math. Stat.*, 40(2):633–643, 1969.

[Keller, 1962] J. B. Keller. Geometrical theory of diffraction. *J. Opt. Soc. Am.*, 52(2):116–130, 1962.

[Klein and Spreij, 1996] A. Klein and P. Spreij. On Fisher's information matrix of an ARMAX process and Sylvester's resultant matrices. *Lin. Alg. Appl.*, 237/238:579–590, 1996 1996.

[Koets, 1998] Michael Koets. Automated algorithms for extraction of physically relevant features from synthetic aperture radar imagery. Master's thesis, The Ohio State University, Columbus, OH, 1998.

[Rissanen, 1978] J. Rissanen. Modeling by the shortest data description. *Automatica*, 14:465–471, 1978.

[Rissanen, 1996] J. J. Rissanen. Fisher information and stochastic complexity. *IEEE Trans. Info. Theory*, 42(1):40–47, 1996.

[Sabharwal and Potter, 1998] Ashutosh Sabharwal and Lee C. Potter. Model selection for nested model classes with complexity constraints. *IEEE Transactions on Information Theory*, 1998. in review.

[Sacchini *et al.*, 1993] Joseph J. Sacchini, William M. Steedly, and Randolph L. Moses. Two-dimensional Prony modeling and parameter estimation. *IEEE Transactions on Signal Processing*, 41(11):3127–3137, November 1993.

[Stach and LeBaron, 1996] J. Stach and E. LeBaron. Enhanced image editing by peak region segmentation. In *Proceedings of the 1996 AMTA Symposium*, October 1996.

[Tsao and Steinberg, 1988] J. Tsao and B.D. Steinberg. Reduction of Sidelob and Speckle Artifacts in Microwave Imaging: The CLEAN technique,. *IEEE Trans. Antennas and Propgation*, 36:543–556, April 1988.

[Tu *et al.*, 1997] M.-W. Tu, I.J. Gupta, and E.K. Walton. Application of maximum likelihood estimation to radar imaging. *IEEE Transactions on Antennas and Propagation*, 45(1):20–27, 1997.

[Van Trees, 1968] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory: Part I Detection, Estimation, and Linear Modulation Theory.* Wiley, New York, 1968.

[Veres, 1985] S. Veres. On the overparametrized maximum likelihood estimation of ARMA processes. In *Proc. 5th Pannonian Symp. on Math. Stat.*, pages 381–400, Visegrad, Hungary, 1985.

[Wald, 1943] A. Wald. Tests of statistical hypothesis concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.*, 54(3):426–482, Nov. 1943.

[Wax and Kailath, 1985] Mati Wax and Thomas Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. ASSP*, ASSP-33(2):387–392, April 1985.