

## REFERENCES

- [1] Lord, R. D. (1954)  
The use of the Hankel transform in statistics.  
*Biometrika*, **41** (1954), 44–55.
- [2] Drumheller, D. M. (1999)  
Padé approximations to matched filter amplitude probability functions.  
*IEEE Transactions on Aerospace and Electronic Systems*, **35**, 3 (1999), 1033–1045.
- [3] Trunk, G. V. (1972)  
Radar properties of non-Rayleigh sea clutter.  
*IEEE Transactions on Aerospace and Electronic Systems*, **AES-8**, 2 (1972), 196–204.
- [4] Dugelay, S., Pace, N. G., Heald, G. J. and Brothers, R. J. (2000)  
Statistical analysis of high frequency acoustic scatter: What makes a statistical distribution?  
*Proceedings of the Fifth European Conference on Underwater Acoustics*, Lyon, France, 2000.
- [5] Eaves, J. L., and Reedy, E. K. (1987)  
*Principles of Modern Radar*.  
Princeton, NJ: Van Nostrand Reinhold, 1987, ch. 11.
- [6] Sangston, K. J., and Gerlach, K. R. (1992)  
Non-Gaussian noise models and coherent detection of radar targets.  
NRL Report 5341-92-9367, Naval Research Laboratory, Washington, DC, Nov. 1992.
- [7] Kay, S. M. (1993)  
*Fundamentals of Statistical Signal Processing*.  
Saddle River, NJ: Prentice-Hall, 1993, ch. 15.
- [8] Wooding, R. A. (1956)  
The multivariate distribution of complex normal random variables.  
*Biometrika*, **43** (1956), 212–215.
- [9] Turin, G. L. (1960)  
The characteristic function of Hermitian quadratic forms in complex normal variables.  
*Biometrika*, **47** (1960), 199–201.
- [10] Conte, E., and Longo, M. (1987)  
Characterization of radar clutter as a spherically invariant random process.  
*IEE Proceedings*, **134**, Pt. F, 2 (1987), 191–197.
- [11] Conte, E., DiBisceglie, M., Longo, M., and Lops, M. (1995)  
Canonical detection in spherically invariance noise.  
*IEEE Transactions on Communications*, **43**, 2/3/4 (1995), 416–424.
- [12] Goldman, J. (1976)  
Detection in the presence of spherically symmetric random vectors.  
*IEEE Transactions on Information Theory*, **IT-22**, 1 (1976), 52–59.
- [13] Rangaswamy, M., Weiner, D., and Öztürk, A. (1993)  
Non-Gaussian random vector identification using spherically invariant random processes.  
*IEEE Transactions on Aerospace and Electronic Systems*, **29**, 1 (1993), 111–124.
- [14] Yao, K. (1973)  
A representation theorem and its applications to spherically-invariant random processes.  
*IEEE Transactions on Information Theory*, **IT-19**, 5 (1973), 600–608.
- [15] Watson, G. N. (1944)  
*A Treatise on the Theory of Bessel Functions*, (2nd ed.).  
Cambridge, England: Cambridge University Press, 1944.
- [16] Luke, Y. L. (1962)  
*Integrals of Bessel Functions*.  
New York: McGraw-Hill, 1962, ch. 2.

## Performance Complexity Study of Several Approaches to Automatic Target Recognition from SAR Images

A framework which allows for the direct comparison of alternate approaches to automatic target recognition (ATR) from synthetic aperture radar (SAR) images is described and applied to variants of several ATR algorithms. This framework allows comparisons to be made on an even footing while minimizing the impact of implementation details and accounts for variation in image sizes, in angular resolution, and in the sizes of orientation windows used for training. Alternate approaches to ATR are characterized in terms of the best achievable performance as a function of the complexity of the model parameter database. Several approaches to ATR from SAR images are described and the performance achievable by each for a range of database complexities is studied and compared. These approaches are based on a likelihood test under a conditionally Gaussian model, log-magnitude least squared error, and quarter power least squared error. All approaches are evaluated for a wide range of parameterizations and the dependence on these parameters of both the resulting performance and the resulting database complexity is explored. Databases for all of the approaches are trained using identical sets of images and their performance is assessed under identical testing scenarios in terms of probability of correct classification, confusion matrices, and orientation estimation error. The results indicate that the conditionally Gaussian approach outperforms the other two approaches on average for both target recognition and orientation estimation, that accounting for radar power fluctuation improves performance for all three methods, and that the conditionally Gaussian approach normalized for power delivers average performance that is equal or superior to all other considered approaches.

### I. INTRODUCTION

Many approaches to performing automatic target recognition (ATR) from synthetic aperture radar (SAR) data have been proposed in the literature

Manuscript received December 21, 1999; revised November 29, 2000, September 20, 2001, and January 15, 2002; released for publication January 16, 2002.

IEEE Log No. T-AES/38/2/11448.

Refereeing of this contribution was handled by L. M. Kaplan.

This work was supported in part by the U.S. Army Research Office, Grant DAAH04-95-1-0494, by the Office of Naval Research, Grant N00014-98-1-06-06, and by the Boeing-McDonnell Foundation.

0018-9251/02/\$17.00 © 2002 IEEE

and their performance is reported under training and testing scenarios with sample data. The relative performance of the various approaches, however, can be difficult to assess based upon these empirical performance evaluations. This difficulty is due to implementation details which may vary from one report to another and which may not achieve optimal results in the studies to which they are applied. These details include methods of segmenting target, shadow, and clutter regions, methods of feature extraction, selections of data to use for training and testing the algorithms, and parameterizations such as angular resolution and the widths of intervals from which training data are selected.

Here we seek to provide a framework for the direct comparison of ATR algorithms and which points to the best parameterization of each approach. Building upon the work of O'Sullivan, et al. [20], several variants of three different approaches to ATR from SAR data are compared within this framework. The algorithms compared are a likelihood approach based on a conditionally Gaussian model [11, 19, 20], a least squared error approach based on log-magnitude images [15, 22], and a least squared error approach based on quarter power images [28]. The comparison is performed in terms of the relationship between algorithm performance and database complexity afforded by each approach using sets of training and testing images defined identically for all approaches.

As described in [1], ATR performance is highly dependent upon the range of orientation angles used for training. The influence of these parameter choices on a comparison of algorithms is reduced by considering a wide range of angular resolutions and interval widths from which data is selected for training each angular cell. For each approach, the performance results and overall database complexity under each combination of parameters is analyzed to determine the best performance that can be achieved for any given database complexity. The resulting performance-complexity curves for all approaches are then directly compared. Database complexity is reported as the logarithm of the number of floating point values per target type which must be trained and stored in order for the recognition algorithm to operate. Performance of the various approaches is empirically assessed using publicly released data from the Moving and Stationary Target Acquisition and Recognition (MSTAR) program. The performance is reported in terms of average orientation estimation error, percentage of correctly recognized targets, and confusion matrices.

This method of comparing approaches to ATR is motivated by information-theoretic notions of the tradeoffs between communication rates and data distortion. The database complexity, indicating the quantity of data which must be stored and processed, plays a role similar to communication rate and

performance measurements, which indicate the disparity between the true and inferred targets, play a role similar to distortion. The study is aimed at exploring the capability of the models underlying the algorithms employed, so the comparison of approaches to ATR is limited in the sense that algorithm testing is performed only with images taken under conditions which were very similar to the images used to train. This comparison method can also be applied to tests in which the operating conditions vary.

The study is also limited in the sense that optimal segmentation of the SAR image chips into target, shadow, and background is not considered and the various approaches are required to adequately model pixels in all three region types. Because the approaches to ATR may vary in their response to the relative amounts of these types of pixels, the SAR images are truncated into square subimages of six different sizes, the largest of which includes all target and shadow pixels and the smallest of which eliminates substantial portions of the shadow and background so the target occupies a larger portion of the image chip. Performance results may be expected to improve for carefully and properly selected target segmentations and optimal segmentation depends, in part, on the recognition algorithm used. Algorithms which tend to deliver superior performance with very small or very large images may be good candidates for segmentation strategies which err on the side of removing or retaining, respectively, pixels whose classification is in doubt. Many approaches to the segmentation of SAR images have been reported including those based on statistical based methods [26], fractals [12], and neural networks [17]. Information-theoretic based approaches include simple thresholding [6, 7], Rissanen's minimum description length (MDL) principle [24], and optimal component selection [5, 21]. The performance-complexity method of comparing alternate approaches to ATR can accommodate such methods of segmentation, alternate model representations, etc., and this work establishes a performance baseline against which performance improvements of such alternates can be measured. A direct extension of this framework to other complexity measures, such as chip processing rate [4], and to the comparison of other approaches to ATR such as scattering center approaches [2, 27], neural network based approaches [25], and enhanced-resolution-based approaches [16] is also possible.

Robustness to deviations from model assumptions are an important issue in ATR. It is desirable for an algorithm to be robust with respect to many types of deviation including the presence of clutter, radar gain, partial occlusion, vehicle modifications, and the presence of confuser vehicles (target classes present in the set of testing images but for which training data does not exist). ATR algorithms have

the potential to be robust with respect to some of these and not to others. Preference for one algorithm over another is likely to be influenced by the choice of robustness dimensions. A comparison in terms of the performance-complexity behavior of alternate algorithms can be carried out with test data selected to typify the dimensions of concern.

Section II describes in detail the models for SAR data which are considered. Section III details the algorithms for target recognition and orientation estimation that are derived from the SAR data models. Section IV looks at the performance of each of the algorithms when orientation is finely discretized. Section V contains a full performance versus complexity study for all of the algorithms. Conclusions follow in Section VI.

## II. MODELS FOR SYNTHETIC APERTURE RADAR DATA

This section contains a discussion of the models underlying the conditionally Gaussian, log-magnitude least squared error, and quarter power least squared error approaches. Estimation of the parameters of each model from training data is also considered. Common to all three approaches is the use of a dictionary of estimated values, or templates, which characterize properties of the SAR images of each target class at multiple orientations. The notation employs a representation of a complex-valued SAR image of size  $m \times n$  as a column vector of length  $mn$  obtained by concatenating the values of successive rows and columns of the image.

### A. Conditionally Gaussian Model

In the conditionally Gaussian model [20], the pixels in a SAR image are modeled as independent, complex, conditionally Gaussian random variables given the target class  $a$  and orientation  $\theta \in [0, 2\pi)$ . Specifically, the SAR data are modeled as the vector  $\mathbf{r} = \mathbf{s}(\theta, a) + \mathbf{w}$ , where the signal  $\mathbf{s}$  is a complex Gaussian signal vector conditioned on  $\{\theta, a\}$  with mean  $\mathbf{0}$  and diagonal covariance matrix  $\mathbf{K}(\theta, a)$ , and the receiver noise  $\mathbf{w}$  is also complex Gaussian with mean  $\mathbf{0}$  and covariance  $N_0\mathbf{I}$  and is independent of the signal  $\mathbf{s}$ . The received vector  $\mathbf{r}$  is complex conditionally Gaussian with mean  $\mathbf{0}$  and diagonal covariance matrix  $\sigma^2(\theta, a) = \mathbf{K}(\theta, a) + N_0\mathbf{I}$ . The log-likelihood of  $\mathbf{r}$  given  $\{\theta, a\}$  is a summation over all pixels as

$$l(\mathbf{r} | \theta, a) = \sum_i \left[ -\ln(\sigma_i^2(\theta, a)) - \frac{|r_i|^2}{\sigma_i^2(\theta, a)} \right] \quad (1)$$

where  $\sigma_i^2(\theta, a)$  is the  $i$ th diagonal element of  $\sigma^2(\theta, a)$ .

Model training amounts to estimating the pixel variances  $\sigma_i^2(\theta, a)$  which are the only unknowns in (1).

Because the mean value of every pixel is assumed to be zero, an unbiased estimate for pixel variance given a target and orientation is the sample mean of the squared magnitude of pixel values given  $\{\theta, a\}$ . Since sufficient training data is not likely to be available at each possible value of the continuous variable  $\theta$ , the variance functions  $\sigma_i^2(\theta, a)$  are approximated as piecewise constant over some finite number of windows  $N_w$ . The variance over the  $k$ th window is estimated from the training data which comes from images of targets with azimuth angles lying in an interval  $W_k$  centered at  $2\pi k/N_w$  and having a width  $d$ ,  $W_k = [2\pi k/N_w - d/2, 2\pi k/N_w + d/2]$ . Note that for  $d > 2\pi/N_w$  the training intervals will overlap and a pixel from a given training image may contribute to the variance estimate for more than one azimuth window. Given sets of training data  $\mathcal{I}_{k,a}$  which consist of SAR images of target  $a$  at azimuth  $\theta \in W_k$ , the variance of the  $i$ th pixel over the  $k$ th window for the  $l$ th target is estimated as

$$\hat{\sigma}_i^2(\theta_k, a_l) = \frac{1}{|\mathcal{I}_{k,a}|} \sum_{\mathbf{r} \in \mathcal{I}_{k,a}} |r_i|^2, \quad 1 \leq k \leq N_w, \quad 1 \leq l \leq t \quad (2)$$

where  $t$  is the number of targets in the dictionary.

The power emitted by a SAR platform is subject to fluctuation so that the pixel values in a SAR image of a target may be uniformly scaled by some unknown quantity. The model for a received SAR image can be extended to incorporate such fluctuation via the maximum-likelihood estimate of the power. Specifically, we model the SAR image obtained under nominal transmitter power as  $\mathbf{y}$ ; let  $\mathbf{r} = c\mathbf{y}$  denote the SAR image obtained with some other transmitter power, where  $c$  is an unknown deterministic scale factor. The received vector  $\mathbf{r}$  is a complex Gaussian random vector with mean  $\mathbf{0}$  and diagonal covariance matrix  $c^2\sigma^2(\theta, a)$  and the maximum-likelihood estimate for  $c^2$  is

$$\hat{c}^2(\theta, a) = \frac{1}{N} \sum_i \frac{|r_i|^2}{\sigma_i^2(\theta, a)}. \quad (3)$$

Substituting this estimate into the log-likelihood function for  $\mathbf{r}$  yields

$$l(\mathbf{r} | \theta, a) = \sum_i \left[ -\ln(\hat{c}^2\sigma_i^2(\theta, a)) - \frac{|r_i|^2}{\hat{c}^2\sigma_i^2(\theta, a)} \right] \\ = -2N \ln \hat{c}(\theta, a) - \sum_i \ln \sigma_i^2(\theta, a) - N. \quad (4)$$

Note that what constitutes the nominal transmitter power is quite arbitrary since it is accommodated by the estimated multiplicative factor  $\hat{c}$ . In practice we can take this nominal power to be that of the SAR platform from which the training images  $\mathcal{I}_{k,a}$  were collected.

## B. Log-Magnitude Model

In the log-magnitude model, the complex valued pixels in a SAR image are modeled as independent and having a squared magnitude which follows a log-normal distribution conditioned on target type and orientation. The log-normal distribution has been suggested for use in radar applications, for example [10] and [23, p. 183]. This is analogous to the advanced methods employed in [15, 16, 22] in the sense that the likelihood function for this distribution yields the same decision function as employed there, though these also incorporate hand segmentation of models, superresolution techniques, and threshold-based confuser rejection. According to this model, the image in decibel representation has mean  $\mu_{\text{dB}}(\theta, a)$  and the variance of each pixel is constant across all pixels, targets, and orientations. Then for a received SAR image  $\mathbf{r}$ , the pixels in  $20\log|\mathbf{r}| - \mu_{\text{dB}}(\theta, a)$ , where  $|\cdot|$  is the pixel-wise magnitude and the logarithm is applied pixel-wise, are Gaussian random variables that are independent and identically distributed having mean zero and variance  $\sigma^2$ . In this case, the log-likelihood of  $\mathbf{r}_{\text{dB}} = 20\log|\mathbf{r}|$  given  $\{\theta, a\}$  is the negative of the sum over all pixels of the squared difference between  $\mathbf{r}_{\text{dB}}$  and  $\mu_{\text{dB}}(\theta, a)$ . Maximizing this log-likelihood over  $\{\theta, a\}$  is equivalent to minimizing the squared distance between a received SAR image  $\mathbf{r}_{\text{dB}}$  and the mean vector for some  $\{\theta, a\}$  defined by

$$d^2(\mathbf{r}_{\text{dB}}, \mu_{\text{dB}}(\theta, a)) = \|\mathbf{r}_{\text{dB}} - \mu_{\text{dB}}(\theta, a)\|^2 \quad (5)$$

where  $\|\cdot\|^2$  equals the sum of the squared components.

With the goal of producing a recognition system which is less sensitive to fluctuations in the radar transmitter and receiver on successive uses, the use of image normalization prior to the distance computation in (5) has been suggested. In the geometric mean approach to normalization, the squared magnitude vector  $|\mathbf{r}|^2$  is divided by its geometric mean prior to representation in decibels. This is equivalent to subtracting the arithmetic mean in the decibel scale. The resulting pixels are no longer independent, however in keeping with the form of (5), the squared distance used as a score function indicating how much the received vector appears to come from target  $a$  at orientation  $\theta$  is

$$\begin{aligned} d^2(\mathbf{r}_{\text{dB}} - \overline{\mathbf{r}_{\text{dB}}}, \mu_{\text{dB}}(\theta, a) - \overline{\mu_{\text{dB}}(\theta, a)}) \\ = \|(\mathbf{r}_{\text{dB}} - \overline{\mathbf{r}_{\text{dB}}}) - (\mu_{\text{dB}}(\theta, a) - \overline{\mu_{\text{dB}}(\theta, a)})\|^2 \end{aligned} \quad (6)$$

where  $\overline{\mathbf{x}}$  is the average value across all components of the vector  $\mathbf{x}$ .

Model training for these approaches consists of estimating the mean vectors  $\mu_{\text{dB}}(\theta, a)$ . Estimates for these mean vector functions can be obtained from the sample mean of the magnitude in decibels

given  $\{\theta, a\}$ . As with the variance functions in the conditionally Gaussian model, it is useful to approximate the mean vector functions as piecewise constant in  $\theta$  over some finite number of windows  $N_w$ . The mean over the  $k$ th window is estimated by averaging sample vectors that come from a target having an azimuth angle in the interval  $W_k = [2\pi k/N_w - d/2, 2\pi k/N_w + d/2]$  for some interval width  $d$ . Given sets of training images  $\mathcal{I}_{k,a}$  which consist of SAR images of target  $a$  at azimuth  $\theta \in W_k$ , the mean of the  $i$ th pixel over the  $k$ th window for the  $l$ th target is estimated as

$$\begin{aligned} \hat{\mu}_i(\theta_k, a_l) = \frac{1}{|\mathcal{I}_{k,a}|} \sum_{\mathbf{r} \in \mathcal{I}_{k,a}} 20\log|r_i|, \\ 1 \leq k \leq N_w, \quad 1 \leq l \leq t \end{aligned} \quad (7)$$

where  $t$  is the number of targets in the dictionary.

## C. Quarter Power Model

In the quarter power approach, the magnitude of the complex pixels in a SAR image are modeled as gamma distributed random variables. It has been noted [8] that for a gamma random variable  $g$ , the transformation  $z = g^\alpha$  for some  $\alpha \in (0, 1)$  yields a random variable which is similar to a Gaussian random variable in the sense that the kurtosis of  $z$ ,  $E[(z - \mu_z)^4]/E^2[(z - \mu_z)^2]$ , is approximately 3 which is the kurtosis for a Gaussian random variable. In the quarter power approach, the value used is  $\alpha = 1/2$  and the square root of the magnitude of each pixel in a SAR image is modeled as having some mean value  $\mu_{\text{QP}}(\theta, a)$  which is a function of both orientation and target. The squared distance between the quarter power representation of a received SAR image  $\mathbf{r}_{\text{QP}} = |\mathbf{r}|^{1/2}$ , where  $|\cdot|^{1/2}$  indicates the pixel-wise square root of the magnitude, and the mean vector for some  $\{\theta, a\}$ ,

$$d^2(\mathbf{r}_{\text{QP}}, \mu_{\text{QP}}(\theta, a)) = \|\mathbf{r}_{\text{QP}} - \mu_{\text{QP}}(\theta, a)\|^2 \quad (8)$$

is used as a score function indicating how much the image appears to be that of target  $a$  at orientation  $\theta$ .

As with the other approaches, a form of normalization of  $\mathbf{r}$  can be invoked in an attempt to reduce the sensitivity of the recognition algorithm to fluctuations in radar performance. Here, the vectors  $\mathbf{r}_{\text{QP}}$  and  $\mu_{\text{QP}}(\theta, a)$  are normalized to have unit magnitude. In keeping with the form of (8), the squared distance

$$\begin{aligned} d^2(\mathbf{r}_{\text{QP}}/\|\mathbf{r}_{\text{QP}}\|, \mu_{\text{QP}}(\theta, a)/\|\mu_{\text{QP}}(\theta, a)\|) \\ = \left\| \frac{\mathbf{r}_{\text{QP}}}{\|\mathbf{r}_{\text{QP}}\|} - \frac{\mu_{\text{QP}}(\theta, a)}{\|\mu_{\text{QP}}(\theta, a)\|} \right\|^2 \end{aligned} \quad (9)$$

is used as a score function.

Model training in this case consists of estimating the mean vectors  $\mu_{\text{QP}}(\theta, a)$ . Estimates for these mean vector functions can be obtained from the sample mean of the quarter power given  $\{\theta, a\}$ . Again, we approximate the mean vector functions as piecewise constant in  $\theta$ . Given sets of training images  $\mathcal{I}_{k,a}$  which consist of SAR images of target  $a$  at azimuth  $\theta \in W_k$ , the mean of the  $i$ th pixel over the  $k$ th window for the  $l$ th target is estimated as

$$\hat{\mu}_i(\theta_k, a_l) = \frac{1}{|\mathcal{I}_{k,a}|} \sum_{\mathbf{r} \in \mathcal{I}_{k,a}} |r_i|^{1/2}, \quad 1 \leq k \leq N_w, \quad 1 \leq l \leq t \quad (10)$$

where  $t$  is the number of targets in the dictionary.

### III. ESTIMATION AND RECOGNITION ALGORITHMS

This section contains a brief discussion of the measure of orientation error used in this work and then considers various methods for inferring a target class  $a(\mathbf{r})$  corresponding to a SAR image  $\mathbf{r}$  and estimating the orientation of the target in the image given the trained models discussed in Section II. This section concludes with a description of the SAR image data used for assessing the performance of the various approaches and the empirical measures of performance which are employed.

#### A. Orientation Distance Measures

At any given time, a target has an orientation in 3-dimensional space that can be represented by a rotation matrix which is an element of  $\mathbf{SO}(3)$ , the special orthogonal group of dimension 3. For ground-based targets on a horizontal surface, the orientation can be represented by a rotation matrix which is an element of  $\mathbf{SO}(2)$ . The orientation of a target in  $\mathbf{SO}(2)$  can be thought of as a function of the azimuth angle  $\theta$  of the target relative to a radar platform with known depression angle. The matrix representation in  $\mathbf{SO}(2)$  corresponding to angle  $\theta$  is

$$\mathbf{O}(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}. \quad (11)$$

The squared Hilbert–Schmidt distance between two matrices is defined as the sum of the squared difference between the entries of the two matrices. For any two angles  $\theta_1$  and  $\theta_2$ ,

$$\begin{aligned} d_{\text{HS}}^2(\mathbf{O}(\theta_1), \mathbf{O}(\theta_2)) &= \|\mathbf{O}(\theta_1) - \mathbf{O}(\theta_2)\|_{\text{HS}}^2 \\ &= 4 - 4 \cos(\theta_1 - \theta_2). \end{aligned} \quad (12)$$

We use the squared Hilbert–Schmidt distance  $d_{\text{HS}}^2(\mathbf{O}_{\text{true}}, \hat{\mathbf{O}}(\mathbf{r}))$  between the true orientation of a target,  $\mathbf{O}_{\text{true}}$ , and any estimate,  $\hat{\mathbf{O}}(\mathbf{r})$ , as a measure of the squared error and report the average of this squared distance over the testing set as the average squared error  $e_{\text{HS}}^2$ . The average estimation error

is also reported as the equivalent error in degrees,  $\cos^{-1}(1 - e_{\text{HS}}^2/4)$ , to aid in interpreting the results.

#### B. Algorithms for the Conditionally Gaussian Model

Orientation estimates under the conditionally Gaussian model can be obtained through the Hilbert–Schmidt estimator which is the orientation that minimizes the mean of the squared estimation error conditioned on the received vector  $\mathbf{r}$  and the target class  $a$ . Since the conditional probability density function  $p(\mathbf{O} | \mathbf{r}, a) \propto p(\mathbf{r} | \mathbf{O}, a)p(\mathbf{O} | a)$ , the Hilbert–Schmidt estimator can be written as

$$\begin{aligned} \hat{\mathbf{O}}_{\text{HS}}(\mathbf{r}) &= \underset{\mathbf{O} \in \mathbf{SO}(2)}{\operatorname{argmin}} \int_{\mathbf{SO}(2)} \|\mathbf{O}' - \mathbf{O}\|_{\text{HS}}^2 \\ &\quad \times p(\mathbf{r} | \mathbf{O}', a)p(\mathbf{O}' | a)\gamma(d\mathbf{O}') \end{aligned} \quad (13)$$

where  $\gamma$  is the base measure on  $\mathbf{SO}(2)$  and the form of  $p(\mathbf{r} | \mathbf{O}, a)$  is known from Section IIA. Expressing this integral in terms of  $\theta$  and simplifying yields an expression for the Hilbert–Schmidt estimator for target azimuth

$$\hat{\theta}_{\text{HS}}(\mathbf{r}) = \tan^{-1} \left[ \frac{\int \sin(\theta)p(\mathbf{r} | \theta, a)p(\theta | a)d\theta}{\int \cos(\theta)p(\mathbf{r} | \theta, a)p(\theta | a)d\theta} \right]. \quad (14)$$

Finally, since the covariance function  $\sigma^2(\theta, a)$  is not known but approximated as piecewise constant with estimated values  $\hat{\sigma}_i^2(\theta_k, a)$  from (2), the probability density function used is also piecewise constant in  $\theta$  and proportional to the exponential of the log-likelihood in (1) or (4), depending on whether power normalization is employed. Then the above integrals can be reduced to summations over the azimuth windows  $W_k$ , and with a uniform prior on orientation and with azimuth windows of uniform width, the estimator becomes

$$\hat{\theta}_{\text{HS}}(\mathbf{r}) = \tan^{-1} \left[ \frac{\sum_k \sin(\theta_k) \exp[\hat{l}(\mathbf{r} | \theta_k, a)]}{\sum_k \cos(\theta_k) \exp[\hat{l}(\mathbf{r} | \theta_k, a)]} \right]. \quad (15)$$

A Bayesian approach to ATR can be utilized under the conditionally Gaussian model by selecting the target class  $\hat{a}(\mathbf{r})$  which maximizes the conditional probability  $P(a | \mathbf{r}) \propto p(\mathbf{r} | a)P(a)$ . We already have an expression for  $p(\mathbf{r} | \mathbf{O}, a)$  so we write  $p(\mathbf{r} | a)$  as a marginal probability density over all orientations and select a target class according to

$$\hat{a}_{\text{Bayes}}(\mathbf{r}) = \underset{a}{\operatorname{argmax}} P(a) \int_{\mathbf{SO}(2)} p(\mathbf{r} | \mathbf{O}, a)p(\mathbf{O})\gamma(d\mathbf{O}). \quad (16)$$

Given an estimated log-likelihood function  $\hat{l}(\mathbf{r} | \mathbf{O}, a)$  that is piecewise constant in  $\theta$  from either (1) or (4),

the above integral can be reduced to a summation over all azimuth windows  $W_k$ , and with uniform priors on orientation and target type and with uniformly spaced windows, we can select a target class according to

$$\hat{a}_{\text{Bayes}}(\mathbf{r}) = \underset{a}{\operatorname{argmax}} \sum_k \exp[\hat{l}(\mathbf{r} | \theta_k, a)]. \quad (17)$$

### C. Algorithms for the Log-Magnitude Model

Orientation estimates in the log-magnitude model are obtained from SAR data  $\mathbf{r}$  for a given target  $a$  by selecting the azimuth angle  $\theta$  that minimizes the squared distance between  $\mathbf{r}_{\text{dB}}$  and the mean vector function for  $a$ . Given the estimated mean vector function  $\hat{\boldsymbol{\mu}}_{\text{dB}}(\theta, a)$  from (10), which is piecewise constant in  $\theta$ , the estimate is given as

$$\hat{\theta}_{\text{LM}}(\mathbf{r}, a) = \underset{\theta}{\operatorname{argmin}} d^2(\mathbf{r}_{\text{dB}}, \hat{\boldsymbol{\mu}}_{\text{dB}}(\theta, a)). \quad (18)$$

In the case that normalization is used, the distance measure minimized is  $d^2(\mathbf{r}_{\text{dB}} - \overline{\mathbf{r}_{\text{dB}}}, \boldsymbol{\mu}_{\text{dB}}(\theta, a) - \overline{\boldsymbol{\mu}_{\text{dB}}(\theta, a)})$ .

ATR in the log-magnitude model from  $\mathbf{r}$  is performed by finding the combination of target type  $a$  and azimuth angle  $\theta$  which minimizes the squared distance between  $\mathbf{r}_{\text{dB}}$  and the corresponding mean vector. That is,

$$\hat{a}_{\text{LM}}(\mathbf{r}) = \underset{a}{\operatorname{argmin}} \min_{\theta} d^2(\mathbf{r}_{\text{dB}}, \hat{\boldsymbol{\mu}}_{\text{dB}}(\theta, a)). \quad (19)$$

In the case that normalization is used, the distance measure minimized is  $d^2(\mathbf{r}_{\text{dB}} - \overline{\mathbf{r}_{\text{dB}}}, \hat{\boldsymbol{\mu}}_{\text{dB}}(\theta, a) - \overline{\hat{\boldsymbol{\mu}}_{\text{dB}}(\theta, a)})$ .

### D. Algorithms for the Quarter Power Model

Orientation estimates in the quarter power model are obtained from SAR data  $\mathbf{r}$  for a given target  $a$  by selecting the azimuth angle  $\theta$  that minimizes the squared distance between  $\mathbf{r}_{\text{QP}}$  and the mean vector function for  $a$ . Given the estimated mean vector function,  $\hat{\boldsymbol{\mu}}_{\text{QP}}(\theta, a)$  from (10), which is piecewise constant in  $\theta$ , the estimate is given as

$$\hat{\theta}_{\text{QP}}(\mathbf{r}, a) = \underset{\theta}{\operatorname{argmin}} d^2(\mathbf{r}_{\text{QP}}, \hat{\boldsymbol{\mu}}_{\text{QP}}(\theta, a)). \quad (20)$$

If the normalization form is used, the distance measure minimized is  $d^2(\mathbf{r}_{\text{QP}}/\|\mathbf{r}_{\text{QP}}\|, \boldsymbol{\mu}_{\text{QP}}(\theta, a)/\|\boldsymbol{\mu}_{\text{QP}}(\theta, a)\|)$ .

ATR in the quarter power model from  $\mathbf{r}$  is performed by finding the combination of target type  $a$  and azimuth angle  $\theta$  which minimizes the squared distance between  $\mathbf{r}_{\text{QP}}$  and the corresponding mean vector. That is,

$$\hat{a}_{\text{QP}}(\mathbf{r}) = \underset{a}{\operatorname{argmin}} \min_{\theta} d^2(\mathbf{r}_{\text{QP}}, \hat{\boldsymbol{\mu}}_{\text{QP}}(\theta, a)). \quad (21)$$

If the normalization form is used, the distance measure minimized is  $d^2(\mathbf{r}_{\text{QP}}/\|\mathbf{r}_{\text{QP}}\|, \hat{\boldsymbol{\mu}}_{\text{QP}}(\theta, a)/\|\hat{\boldsymbol{\mu}}_{\text{QP}}(\theta, a)\|)$ .

### E. MSTAR Dataset and Empirical Evaluation

The SAR image data used for the results we present were collected under the MSTAR<sup>1</sup> program conducted under DARPA funding. The image set contains X band SAR data for a number of targets imaged at two or more depression angles. The SAR images in the dataset are of 1 ft by 1 ft resolution. The target lies in the central portion of the chip and occupies a small portion of it. The rest of the chip consists of background clutter and target shadow. The results we present include an assessment of how the performance of the algorithms varies as a function of the target chip size which is intended to accommodate different sensitivities to shadow removal and excess clutter as the approaches may vary in their ability to model target, shadow, and clutter regions. Smaller image chips of sizes  $128 \times 128$ ,  $112 \times 112$ ,  $96 \times 96$ ,  $80 \times 80$ ,  $64 \times 64$ , and  $48 \times 48$  were constructed by extracting a square region from the center of the MSTAR images. Variance images from the conditionally Gaussian model are shown in Fig. 1 for several of the targets at each of the six image sizes. As can be seen from the figure, for some targets and some orientations such as the BRDM-2 and BTR-70 shown, decreasing the size of images used from  $128 \times 128$  down to  $80 \times 80$  reduces background clutter in the image without infringing upon either the target or shadow. Further reduction in image size eliminates more of the clutter pixels but also eliminates much of the shadow region. For some targets and orientations such as the ZIL131 shown, any reduction of the image size eliminates both target shadow and image artifacts along with the background clutter. As will be shown, each algorithm tends to favor its own particular tradeoff between clutter reduction and retention of shadow and artifact pixels which may have implications for segmentation strategies likely to be useful for each approach.

For each target and depression angle, the MSTAR set contains SAR image data for two to three hundred values of target aspect pose. We have divided these images into two nonoverlapping sets, one for model training and the other for assessing the performance of the various approaches. A summary of the datasets used is provided in Table I. Orientation estimation performance is characterized in terms of the average squared Hilbert–Schmidt distance between the actual and estimated orientations. This average is computed as a function of target type and as an overall value across all 3192 images of the testing set. Recognition performance is characterized in terms of confusion

<sup>1</sup>The MSTAR dataset can be requested through the Sensor Data Management System (SDMS) web page of Wright Laboratory at the URL <http://www.mbvlab.wpafb.af.mil/public/sdms/> or by contacting Kelly Miller at AFRL/SNAS, Building 23, 2010 Fifth Street, WPAFB, OH 45433.

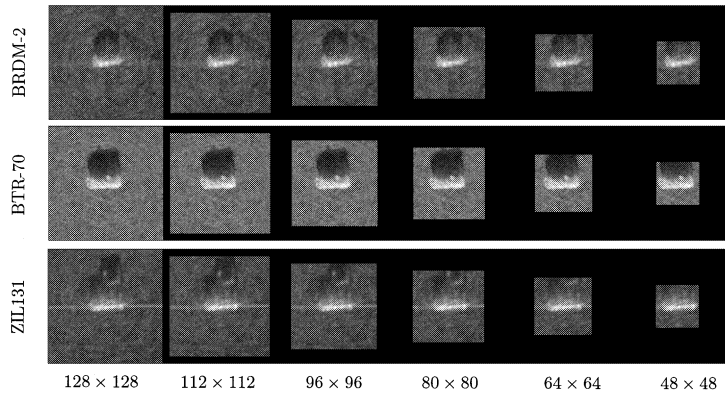


Fig. 1. Variance images for broadside ( $90^\circ$  azimuth) targets BRDM-2, BTR-70, ZIL131. Each image shows variance of corresponding target for each of six image sizes.

TABLE I  
MSTAR Dataset Used in the Performance Analysis of All Models

Target	Train			Test		
	Vehicles	Images	Depression	Vehicles	Images	Depression
2S1	b01	299	$17^\circ$	b01	274	$15^\circ$
BMP-2	9563, 9566, c21	697	$17^\circ$	9563, 9566, c21	587	$15^\circ$
BRDM-2	E-71	298	$17^\circ$	E-71	263	$15^\circ$
BTR-60	k10yt7532	256	$17^\circ$	k10yt7532	195	$15^\circ$
BTR-70	c71	233	$17^\circ$	c71	196	$15^\circ$
D7	92v13015	299	$17^\circ$	92v13015	274	$15^\circ$
T62	A51	299	$17^\circ$	A51	273	$15^\circ$
T-72	132, 812, s7	691	$17^\circ$	132, 812, s7	582	$15^\circ$
ZIL131	E12	299	$17^\circ$	E12	274	$15^\circ$
ZSU 23 4	d08	299	$17^\circ$	d08	274	$15^\circ$

Note: In total, 3670 images are used for training and 3192 images are used for testing.

matrices, percentage of correctly recognized images of each target, and as the percentage of correctly recognized images over all images of the testing set.

#### IV. PERFORMANCE USING 72 WINDOWS

This section presents the performance results for the orientation estimation and target recognition algorithms of Section III when the databases of trained models are relatively large with the model parameters approximated as piecewise constant over  $N_w = 72$  windows and trained over intervals of width  $d = 10^\circ$  each.

##### A. Orientation Estimation Using 72 Windows

For 72 windows, orientation estimation results are indicated in Fig. 2(a) which shows the average orientation estimation error as a function of SAR image size for all approaches. The normalized conditionally Gaussian approach performed significantly better than all other approaches and achieved an average orientation estimation error of  $0.0267$ , equivalent to  $6.62^\circ$ , with  $128 \times 128$  images. The nonnormalized conditionally Gaussian approach

performed better than even the normalized versions of the other algorithms for all but  $128 \times 128$  images. The figure also demonstrates that normalization improves the orientation estimates of the log-magnitude approach for images of size  $64 \times 64$  and above and that normalization improved the orientation estimates of the quarter power approach at all image sizes.

The contribution to the average orientation estimation error is not evenly distributed among the ten targets, but some targets result in a disproportionately large average error as indicated in Fig. 3. This figure shows the average squared Hilbert–Schmidt distance between the actual and estimated orientations on a target-by-target basis for each of the approaches. The top row, from left to right, contains charts for the conditionally Gaussian, log-magnitude, and quarter power approaches. The bottom row contains charts for the corresponding normalized versions. For all approaches, the orientation estimation error for the targets BTR-60, ZIL131, and ZSU-23-4 was very low whereas the error for the targets 2S1, BMP2, and T72 was consistently large. The use of normalization decreased the average error for most of the targets in all three approaches though the targets that benefited most varied by approach. Under the conditionally Gaussian

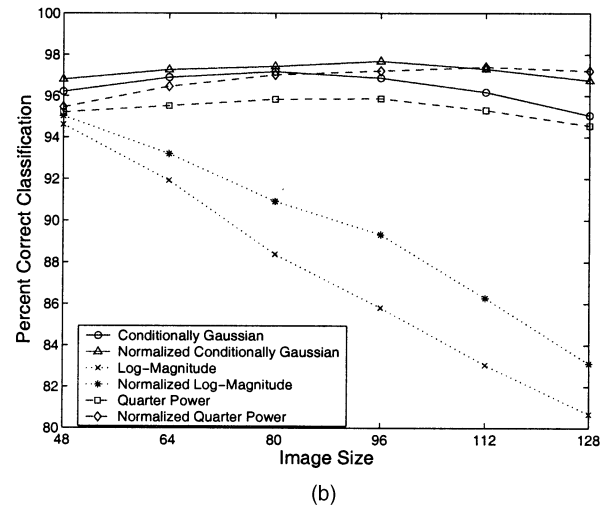
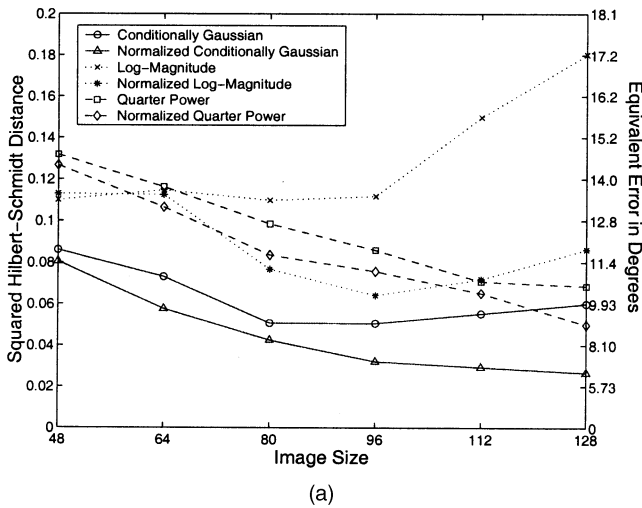


Fig. 2. Performance comparison for each of the approaches when model parameters are approximated as piecewise constant over 72 windows and trained over intervals of width  $10^\circ$  each. (a) Average squared Hilbert-Schmidt distance between actual and estimated orientations as function of target chip size; smaller values denote better performance. (b) Overall percentage of correctly classified test images as function of target chip size; larger values denote better performance.

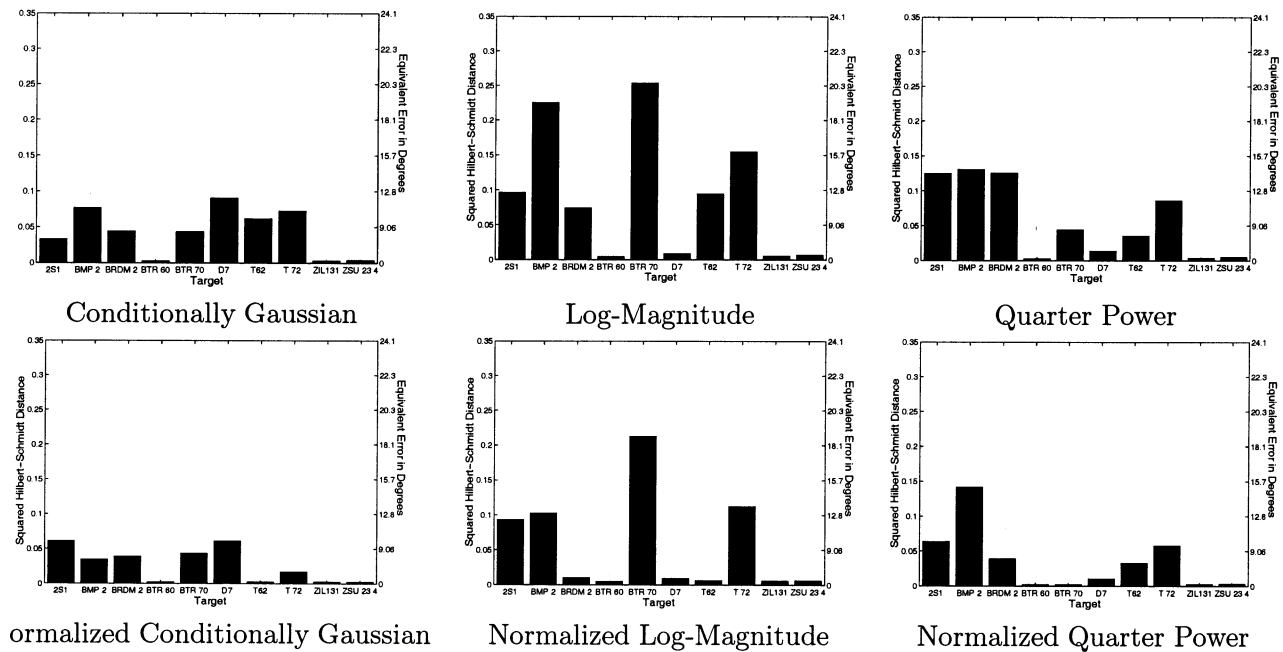


Fig. 3. Average orientation error by target at best performing image size for each of the models; smaller values denote better performance. From left to right, top row shows performance for conditionally Gaussian approach with  $96 \times 96$  images, log-magnitude approach at  $80 \times 80$ , and quarter power approach at  $128 \times 128$ . Bottom row shows results from normalized versions of these approaches at  $128 \times 128$ ,  $96 \times 96$ , and  $128 \times 128$  images, respectively.

approach, the 2S1 performed slightly worse with normalization but still better than for any variant of the log-magnitude or quarter power approaches.

Much of the average orientation estimation error is due to a few estimates which are nearly  $180^\circ$  from the true orientation, implying that some images of targets are occasionally interpreted as facing the opposite direction. This is indicated in Fig. 4(a) which shows a relative frequency histogram of orientation estimation errors for the normalized quarter power

approach. This chart shows that over 95% of the orientation estimates are within  $5^\circ$  of the actual orientation and that over 98% are within  $10^\circ$ . There are almost no estimates that are between  $20^\circ$  and  $170^\circ$  from the true orientation. The small bar at the extreme right edge of the chart indicates that around 1/2% of the estimates are over  $175^\circ$  from the actual orientation. These results are typical for all approaches as can be seen in Table II which shows the percentages of estimates which differ from the true orientation by more than  $170^\circ$  and of those differing

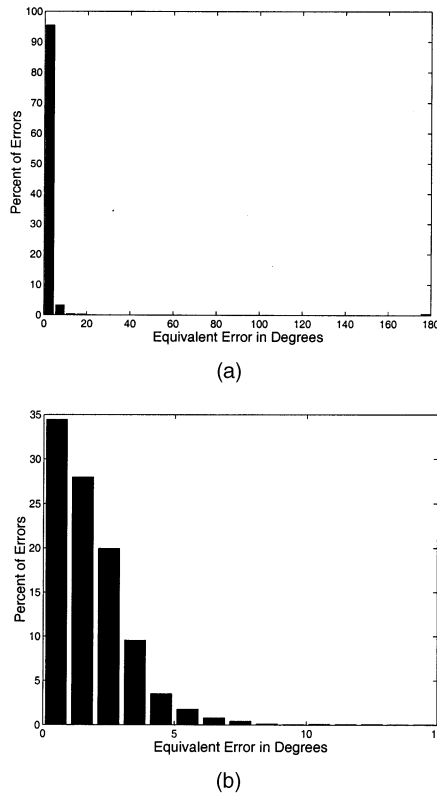


Fig. 4. Relative frequency histogram showing percentage or orientation estimates that fall within a given distance of the true orientation for magnitude normalized quarter power approach when optimal image size of  $128 \times 128$  is used. The x axes have been labeled in terms of the angle, in degrees, which is equivalent to the resulting average squared Hilbert-Schmidt distance. (a) Histogram data collected over 36 successive  $5^\circ$  intervals ranging from  $0^\circ$  to  $180^\circ$ . (b) More detailed look at histogram data from  $0^\circ$  to  $15^\circ$  in 15 successive  $1^\circ$  intervals.

by more than  $175^\circ$ . The normalized conditionally Gaussian approach results in far fewer errors of this type.

Fig. 4(b) shows the details of the relative frequency histogram for orientation estimation errors from  $0^\circ$  to  $15^\circ$  for the magnitude normalized quarter power approach. This chart shows that most of the target orientation estimates are within  $2^\circ$  of the actual orientation. This result and the general shape of the histogram chart are common to all approaches.

## B. Target Recognition Using 72 Windows

The performance results for target recognition with 72 windows are shown in Fig. 2(b). This graph shows the percentage of correctly classified images versus image size for all of the approaches to target recognition. It can be seen that the normalized conditionally Gaussian approach produced correct recognition rates higher than the other approaches for image sizes less than or equal to  $96 \times 96$  while the magnitude normalized quarter power approach produced higher rates for image sizes greater than or equal to  $112 \times 112$ . Furthermore, at  $96 \times 96$  images the normalized conditionally Gaussian approach achieved a correct recognition rate of 97.7% which was the highest rate of all the approaches at any image size. The log-magnitude approach and its normalized variant performed competitively for  $48 \times 48$  images but their performance dropped off considerably for larger image sizes. Finally note that all three normalized variants performed better at all image sizes than their unnormalized counterparts, with the largest increases occurring at the largest image sizes.

Exact details of the recognition performance on a target-by-target basis are provided in Tables III–VIII which give confusion matrices for all approaches at the best performing image size for each method. The targets BMP2, D7, and T72 were among the top performers for each method, whereas the targets 2S1, BRDM2, and T62 were among the worst. Recall from Section IVA that the targets BMP2 and T72 were consistently among the worst performers in terms of orientation estimation error, implying that in general there is not a direct relationship between the ability of these approaches to correctly recognize a target and their ability to estimate its orientation.

These confusion matrices for the log-magnitude and quarter power approaches also indicate that images of the targets 2S1 and BRDM2 were commonly misclassified as BMP2 and that images of targets 2S1 and T62 were commonly misclassified as T72. Both normalized approaches increased the recognition rate of the 2S1 by more than 10 percentage points and increased the recognition rate of the T62 by 2 to 4 percentage points. Under the log-magnitude approach, normalization also resulted

TABLE II  
Percentage of Orientation Estimates Which Differ From True Orientation By a Large Amount

Recognition Approach	Errors $> 170^\circ$	Errors $> 175^\circ$
Conditionally Gaussian	0.56%	0.47%
Conditionally Gaussian, Normalized	0.28%	0.22%
Log-Magnitude	1.07%	0.78%
Log-Magnitude, Normalized	0.60%	0.41%
Quarter Power	0.66%	0.56%
Quarter Power, Magnitude Normalized	0.53%	0.50%

Note: The results shown are for the optimum image size for each method.

TABLE III

Confusion Matrix for Conditionally Gaussian Approach Using  $80 \times 80$  Pixel SAR Images With 72 Windows Trained Over  $10^\circ$  Intervals

	2S1	BMP 2	BRDM2	BTR 60	BTR 70	D7	T62	T72	ZIL131	ZSU 23 4	
2S1	262	0	0	0	0	0	4	8	0	0	95.62%
BMP 2	0	581	0	0	0	0	0	6	0	0	98.98%
BRDM2	5	3	227	1	0	14	3	5	4	1	86.31%
BTR 60	1	0	0	193	0	0	0	0	0	1	98.97%
BTR 70	4	5	0	0	184	0	0	3	0	0	93.88%
D7	2	0	0	0	0	271	1	0	0	0	98.91%
T62	1	0	0	0	0	0	259	11	2	0	94.87%
T72	0	0	0	0	0	0	0	582	0	0	100%
ZIL131	0	0	0	0	0	0	2	0	272	0	99.27%
ZSU 23 4	0	0	0	0	0	2	0	1	0	271	98.91%

Note: Row headings give the true target class for each of the test images represented in the row. For each row, the number in each cell indicates the number of test images which were classified as the corresponding column heading. The rightmost column shows the overall percentage of correctly classified test images for each target.

TABLE IV

Confusion Matrix for Normalized Conditionally Gaussian Approach Using  $96 \times 96$  Pixel SAR Images With 72 Windows Trained Over  $10^\circ$  Intervals

	2S1	BMP 2	BRDM2	BTR 60	BTR 70	D7	T62	T72	ZIL131	ZSU 23 4	
2S1	264	1	0	0	0	0	1	7	1	0	96.35%
BMP 2	0	584	0	0	0	0	0	1	0	2	99.49%
BRDM2	5	3	232	1	0	8	3	9	1	0	88.21%
BTR 60	0	0	0	193	0	0	0	2	0	0	98.97%
BTR 70	0	2	0	0	188	0	0	6	0	0	95.92%
D7	0	0	0	0	0	270	2	1	0	1	98.54%
T62	0	0	0	0	0	0	259	12	0	2	94.87%
T72	0	0	0	0	0	0	0	582	0	0	100%
ZIL131	0	0	0	0	0	0	0	1	272	1	99.27%
ZSU 23 4	0	0	0	0	0	0	0	0	0	274	98.91%

TABLE V

Confusion Matrix for Log-Magnitude Approach Using Image Size  $48 \times 48$  Yielding Best Percentage of Correct Classification When Using 72 Windows Trained Over  $10^\circ$  Intervals

	2S1	BMP 2	BRDM2	BTR 60	BTR 70	D7	T62	T72	ZIL131	ZSU 23 4	
2S1	220	28	2	0	0	0	1	22	0	1	80.29%
BMP 2	0	579	0	0	0	0	0	7	0	1	98.64%
BRDM2	1	21	236	3	0	0	0	0	1	1	89.73%
BTR 60	0	3	0	190	2	0	0	0	0	0	97.44%
BTR 70	1	10	0	2	181	0	0	2	0	0	92.35%
D7	0	0	0	0	0	272	0	1	1	0	99.27%
T62	0	5	0	0	0	0	232	33	3	0	84.98%
T72	0	4	0	0	0	0	0	578	0	0	99.31%
ZIL131	0	2	0	0	0	0	0	7	263	2	95.99%
ZSU 23 4	0	0	0	0	0	0	0	4	1	269	98.18%

in a decrease of the recognition rate of the BTR60 by more than 6 percentage points. Under the quarter power approach, the use of normalization decreased the recognition rate of the BRDM2 by more than 2 percentage points.

## V. PERFORMANCE-COMPLEXITY TRADEOFFS

The relationship between algorithm performance and database complexity is explored in this section. To produce these results, all algorithms were run for 40

combinations of number of windows  $N_w$  and training interval width  $d$ , at 6 different image sizes for a total of 240 parameter combinations. Fig. 5 shows the combinations of  $N_w$  and  $d$  that were included along with the legend for the performance-complexity detail plots in this section. The use of databases with image sizes that vary by target was not considered. Training was performed using the set of 3670 training images described in Table I for all parameter combinations, and testing was performed twice, once as a check of proper operation using the training set itself and

TABLE VI  
Confusion Matrix for Normalized Log-Magnitude Approach Using Image Size  $48 \times 48$  Yielding Best Percentage of Correct Classification When Using 72 Windows Trained Over  $10^\circ$  Intervals

	2S1	BMP 2	BRDM2	BTR 60	BTR 70	D7	T62	T72	ZIL131	ZSU 23 4	
2S1	249	10	2	0	0	0	3	9	0	1	90.87%
BMP 2	0	581	0	0	1	0	0	5	0	0	98.98%
BRDM2	5	24	227	3	0	0	0	0	3	1	86.31%
BTR 60	1	7	1	177	0	3	1	2	1	2	90.77%
BTR 70	5	12	0	2	172	0	0	5	0	0	87.76%
D7	0	0	0	0	0	272	0	1	1	0	99.27%
T62	3	2	0	0	0	0	239	24	3	2	87.55%
T72	0	3	0	0	0	0	0	579	0	0	99.49%
ZIL131	0	2	0	0	0	0	0	3	268	1	97.81%
ZSU 23 4	0	0	0	0	0	2	0	2	1	269	98.18%

TABLE VII  
Confusion Matrix for Quarter Power Approach Using Image Size  $96 \times 96$  Yielding Best Percentage of Correct Classification When Using 72 Windows Trained Over  $10^\circ$  Intervals

	2S1	BMP 2	BRDM2	BTR 60	BTR 70	D7	T62	T72	ZIL131	ZSU 23 4	
2S1	229	22	4	0	0	0	2	16	1	0	83.58%
BMP 2	1	572	1	1	0	0	0	8	1	3	97.45%
BRDM2	0	5	248	3	0	0	0	3	3	1	94.30%
BTR 60	0	0	0	193	1	0	0	1	0	0	98.97%
BTR 70	0	2	0	0	193	0	0	1	0	0	98.47%
D7	0	0	0	0	0	271	1	1	1	0	98.91%
T62	0	0	1	0	0	0	235	31	3	3	86.08%
T72	0	4	0	0	1	0	0	576	0	1	98.97%
ZIL131	0	0	0	0	0	0	0	1	272	1	99.27%
ZSU 23 4	0	0	0	0	0	0	0	2	1	271	98.91%

TABLE VIII  
Confusion Matrix for Normalized Quarter Power Approach Using Image Size  $112 \times 112$  Yielding Best Percentage of Correct Classification When Using 72 Windows Trained Over  $10^\circ$  Intervals

	2S1	BMP 2	BRDM2	BTR 60	BTR 70	D7	T62	T72	ZIL131	ZSU 23 4	
2S1	263	7	1	0	0	0	1	2	0	0	95.99%
BMP 2	1	579	1	0	0	0	0	2	1	3	98.64%
BRDM2	0	19	241	2	0	0	0	0	0	1	91.64%
BTR 60	0	0	0	195	0	0	0	0	0	0	100%
BTR 70	0	4	0	0	192	0	0	0	0	0	97.96%
D7	0	0	0	0	0	272	0	1	1	0	99.27%
T62	1	2	0	0	0	0	247	21	1	1	90.48%
T72	0	2	0	0	0	0	0	580	0	0	99.66%
ZIL131	1	1	0	0	0	0	0	1	270	1	98.54%
ZSU 23 4	0	0	0	0	0	2	0	2	0	270	98.54%

once to produce the performance analysis using the nonoverlapping set of 3192 test images. This complete set of computations consumed approximately three weeks of processing time on a dual processor Sun Enterprise 250 for each of the six approaches. Only the results from the nonoverlapping set of test images are presented in this section.

#### A. Database Complexity

For each of the approaches to orientation estimation and target recognition discussed in Section III, there are three parameters that influence

the number of values that must be stored in the algorithm's database. These parameters are the number of target classes which are to be differentiated, the number of templates chosen to represent the possible orientations of the target  $N_w$ , and the number of pixels in each template. The width of the training interval  $d$  does not affect the number of values in the database but does play a role in the resulting performance [1]. All of the approaches considered require a single floating point value for each pixel of each template of each target type. For the purposes of this paper we define the complexity of a template database as the logarithm of the number

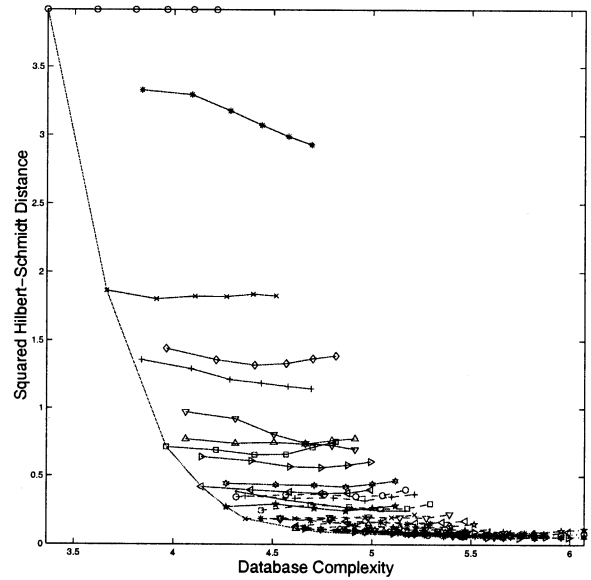
$\circ$ — $\circ$ $N_w = 1, d = 360^\circ$	$\triangleleft$ — $\triangleleft$ $N_w = 6, d = 60^\circ$	$*$ — $*$ $N_w = 12, d = 30^\circ$	$\star$ — $\star$ $N_w = 24, d = 15^\circ$	$\nabla$ — $\nabla$ $N_w = 45, d = 8^\circ$
$\times$ — $\times$ $N_w = 2, d = 180^\circ$	$\triangleright$ — $\triangleright$ $N_w = 6, d = 120^\circ$	$\square$ — $\square$ $N_w = 12, d = 60^\circ$	$\dashv$ — $\dashv$ $N_w = 24, d = 30^\circ$	$\triangle$ — $\triangle$ $N_w = 45, d = 16^\circ$
$+$ — $+$ $N_w = 3, d = 120^\circ$	$\star$ — $\star$ $N_w = 8, d = 45^\circ$	$\diamond$ — $\diamond$ $N_w = 15, d = 24^\circ$	$\circ$ — $\circ$ $N_w = 30, d = 12^\circ$	$\triangleleft$ — $\triangleleft$ $N_w = 60, d = 6^\circ$
$*$ — $*$ $N_w = 3, d = 240^\circ$	$\star$ — $\star$ $N_w = 8, d = 90^\circ$	$\nabla$ — $\nabla$ $N_w = 15, d = 48^\circ$	$\times$ — $\times$ $N_w = 30, d = 24^\circ$	$\triangleright$ — $\triangleright$ $N_w = 60, d = 12^\circ$
$\square$ — $\square$ $N_w = 4, d = 90^\circ$	$\dashv$ — $\dashv$ $N_w = 9, d = 40^\circ$	$\triangle$ — $\triangle$ $N_w = 18, d = 20^\circ$	$+$ — $+$ $N_w = 36, d = 10^\circ$	$*$ — $*$ $N_w = 72, d = 5^\circ$
$\diamond$ — $\diamond$ $N_w = 4, d = 180^\circ$	$\circ$ — $\circ$ $N_w = 9, d = 80^\circ$	$\triangleleft$ — $\triangleleft$ $N_w = 18, d = 40^\circ$	$*$ — $*$ $N_w = 36, d = 20^\circ$	$\star$ — $\star$ $N_w = 72, d = 10^\circ$
$\nabla$ — $\nabla$ $N_w = 5, d = 72^\circ$	$\times$ — $\times$ $N_w = 10, d = 36^\circ$	$\triangleright$ — $\triangleright$ $N_w = 20, d = 18^\circ$	$\square$ — $\square$ $N_w = 40, d = 9^\circ$	$\dashv$ — $\dashv$ $N_w = 72, d = 15^\circ$
$\triangle$ — $\triangle$ $N_w = 5, d = 144^\circ$	$+$ — $+$ $N_w = 10, d = 72^\circ$	$*$ — $*$ $N_w = 20, d = 36^\circ$	$\diamond$ — $\diamond$ $N_w = 40, d = 18^\circ$	$\circ$ — $\circ$ $N_w = 72, d = 20^\circ$

Fig. 5. Combinations of number of windows  $N_w$  and training interval widths  $d$  used in performance-complexity computation. Along with each combination are the symbols used to denote it in the detailed performance-complexity graphs.

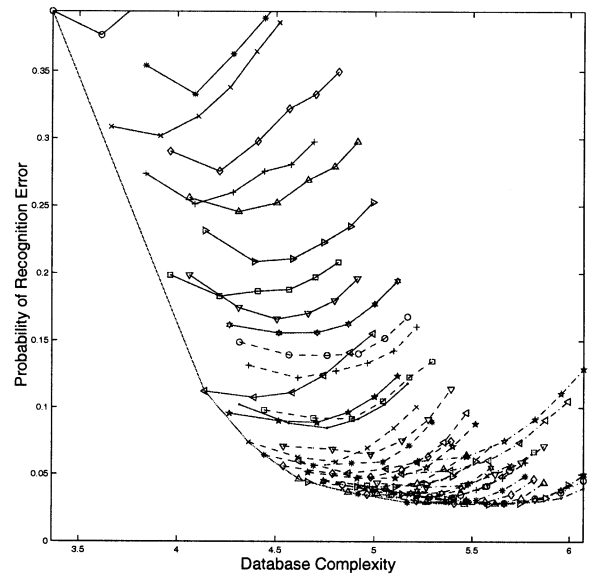
of floating point values per target type that must be permanently stored in order for the algorithm to operate. The logarithm of the number of floating point operations required per target type is roughly proportional to this quantity.

To obtain an empirical assessment of how the performance of each algorithm varies with database complexity, we applied each of the algorithms to the ten target classes given in Table I under 240 combinations of number of templates, widths of training intervals, and image size. The overall performance in terms of both orientation estimation error and percentage of falsely classified targets was noted for each combination of parameters and the curve representing the convex outer boundary of the corresponding performance-complexity pairs was recorded. This outer boundary represents the best performance that is achievable for each complexity value out of all combinations of parameters that were examined or, alternatively, the database complexity required to achieve any desired performance. The outer boundary terminates at that point where increases in database complexity do not yield any improvement in performance. The combination of performance and complexity represented by any point on this boundary between experimentally obtained data points can be achieved on average by randomly selecting, for each test, the parameters yielding one of the points with probability  $P$  and the parameters yielding the other point with probability  $(1 - P)$ .

Fig. 6 shows the results of this computation for the conditionally Gaussian approach. Fig. 6(a) shows the orientation estimation performance of the Hilbert–Schmidt estimator resulting from each combination of parameters. The curves corresponding to the line segments listed in the legend of Fig. 5 show how the performance varies with image size for the number of templates and width of training interval indicated. The image sizes corresponding to each marker along the curve from left to right are  $48 \times 48$ ,  $64 \times 64$ ,  $80 \times 80$ ,  $96 \times 96$ ,  $112 \times 112$ , and  $128 \times 128$ . For example, the solid line with “+” shaped point

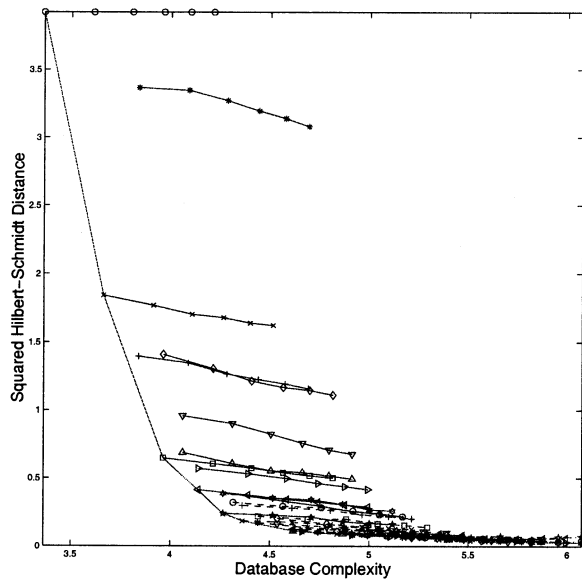


(a)

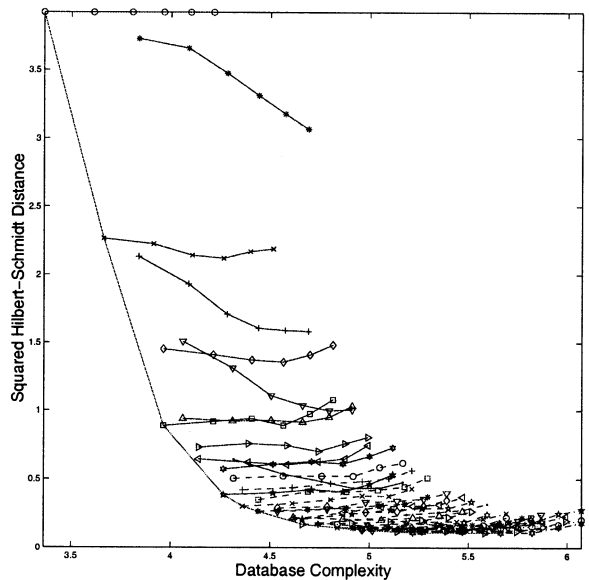


(b)

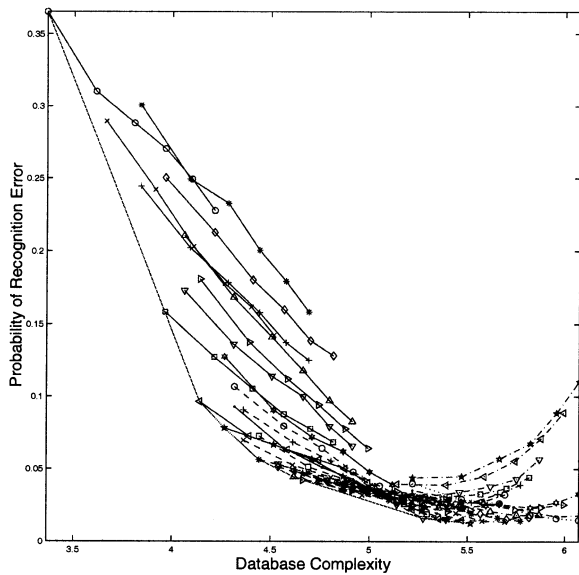
Fig. 6. Performance versus database complexity for conditionally Gaussian approach. Legend given in Fig. 5. (a) Average orientation estimation error. (b) Percentage of erroneously classified targets.



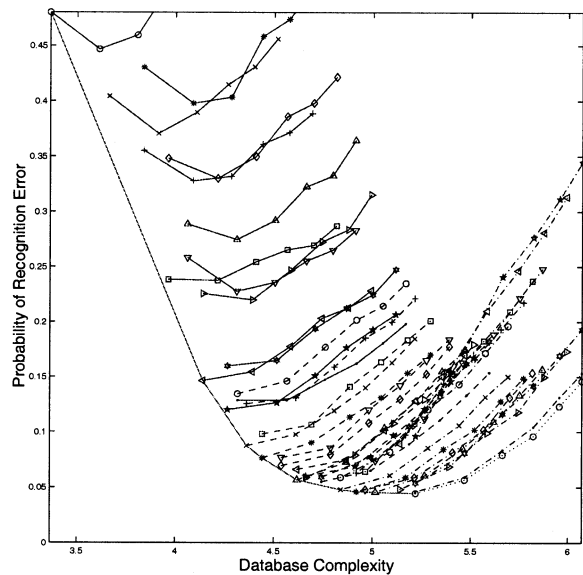
(a)



(a)



(b)



(b)

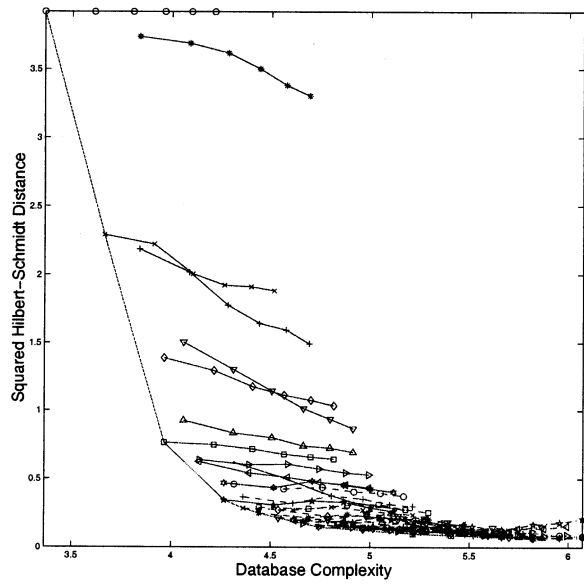
Fig. 7. Performance versus database complexity for conditionally Gaussian approach. Legend given in Fig. 5. (a) Average orientation estimation error. (b) Percentage of erroneously classified targets.

Fig. 8. Performance versus database complexity log-magnitude approach. Legend given in Fig. 5. (a) Average orientation estimation error. (b) Percentage of erroneously classified targets.

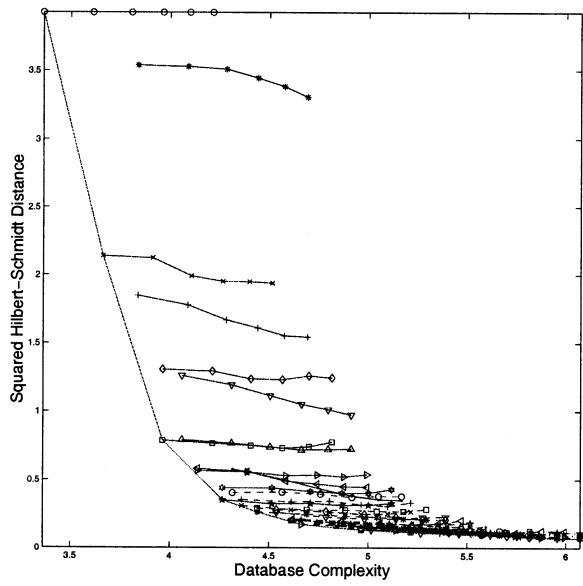
markers shows that for 3 templates each trained over an interval of  $120^\circ$ , the average orientation estimation error decreases uniformly with increasing image size from nearly 1.4 for  $48 \times 48$  images down to around 1.2 for  $128 \times 128$  images. The database complexity for this combination of parameters grows uniformly from around 3.8 for  $48 \times 48$  images to 4.7 for  $128 \times 128$  images. Fig. 6(b) shows the recognition error rate for each combination of parameters. Figs. 7–11 show the computation results for the other five approaches.

These figures show that for all of the six approaches and for both orientation estimation and target recognition, at small complexities the largest

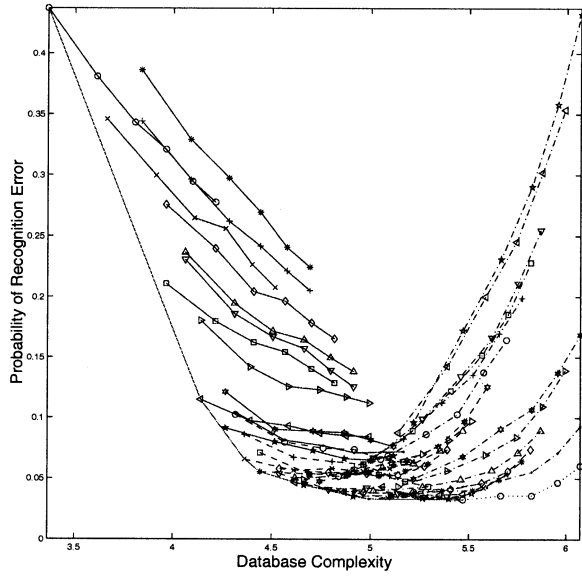
increases in performance come from increasing the number of templates in the database rather than increasing image size. Also for all approaches, increases in performance tend to taper off as the database complexity passes 4.5. The figures also indicate that at large database complexities it is the image size and not the number of templates that has the largest influence on performance. In all approaches orientation estimation performance is less sensitive to image size than is recognition performance. At high database complexities, the quarter power and normalized quarter power approaches are much less sensitive to variations in image size than are any of the other approaches.



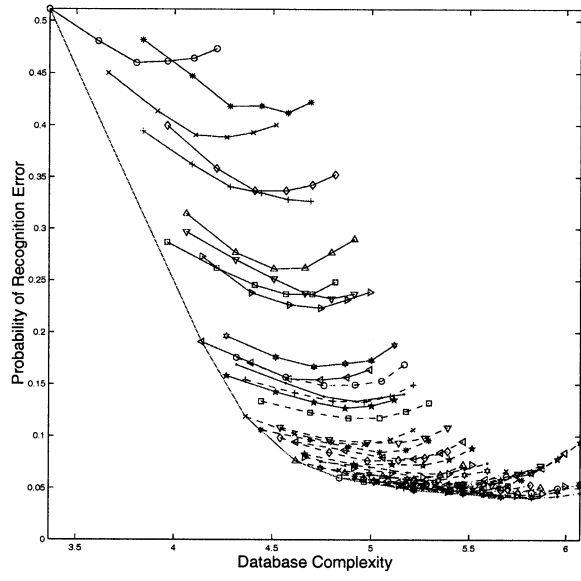
(a)



(a)



(b)



(b)

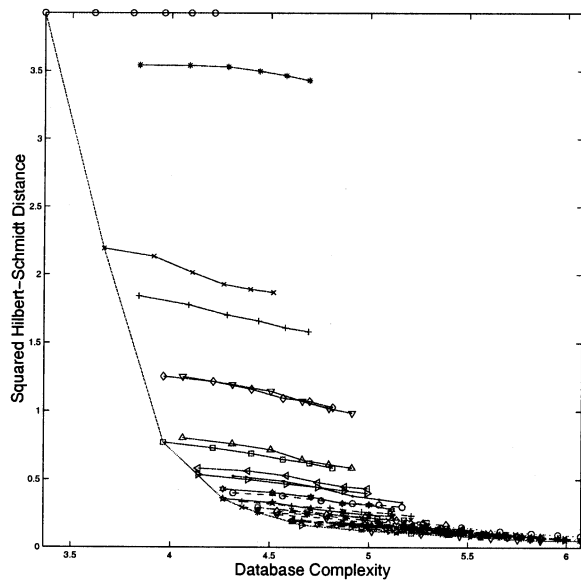
Fig. 9. Performance versus database complexity for normalized log-magnitude approach. Legend given in Fig. 5. (a) Average orientation estimation error. (b) Percentage of erroneously classified targets.

Fig. 10. Performance versus database complexity for quarter power approach. Legend given in Fig. 5. (a) Average orientation estimation error. (b) Percentage of erroneously classified targets.

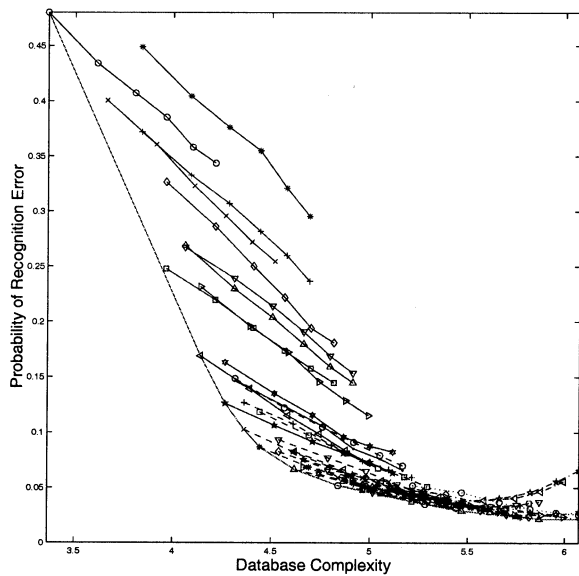
Finally, note that the particular combination of 72 templates per target class with parameters trained over  $5^\circ$  intervals (denoted by a five-pointed star on a dash-dotted line), which is occasionally reported in the literature, is for all six approaches one of the worst performing among the highly complex database parameterizations.

These figures also show some trends in the orientation estimation and recognition error rates when the number of templates per target type and width of the template training interval are held constant and the image sizes are varied. For many combinations of targets and pose, nearly all additional pixels at large image sizes contain clutter. This

could potentially bias the results for large images. Both the conditionally Gaussian and quarter power approaches tend to have orientation estimation performance which is relatively insensitive to image size and a recognition error rate which dips at the medium image sizes. This is consistent with clutter degrading performance. Normalization of the conditionally Gaussian and quarter power approaches yields performance which tends to increase with image size for both orientation estimation and target recognition regardless of the number of templates in the database. Though clutter pixels are not directly useful for target classification, their use does improve the estimates of power scale factors



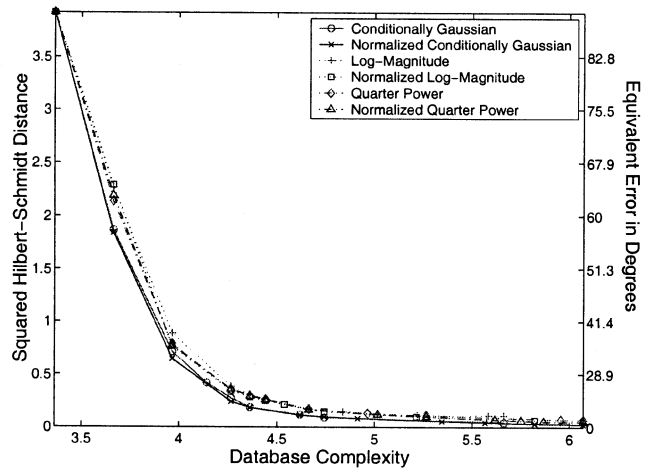
(a)



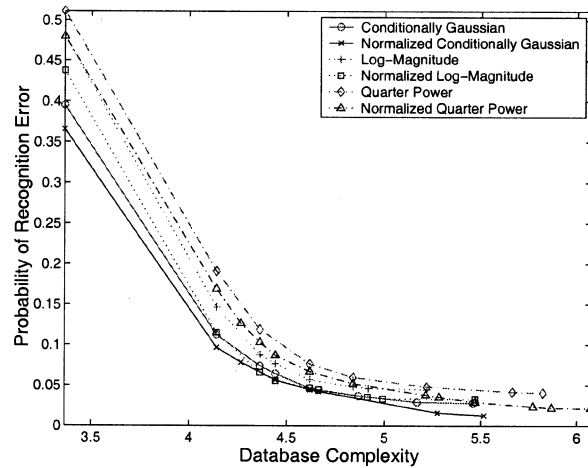
(b)

Fig. 11. Performance versus database complexity for normalized quarter power approach. Legend given in Fig. 5. (a) Average orientation estimation error. (b) Percentage of erroneously classified targets.

$c$ , and these in turn improve classification results. The log-magnitude approach tends to have slightly worse estimation performance with increasing image size and shows significant degradation in recognition performance with large image sizes. Normalization of the log-magnitude approach shows less sensitivity to image size when the number of templates per target type is not large, for both orientation estimation and target recognition, but continues to show a large degradation in recognition performance with increasing image size when a large number of templates is employed.



(a)



(b)

Fig. 12. Best achievable performance versus database complexity for all approaches. For both graphs, smaller values denote better performance.

## B. Side-by-Side Comparison

The best achievable performance for all of the methods as a function of database complexity is shown in Fig. 12. Fig. 12(a) shows the lowest achievable average orientation estimation error for each approach and indicates that the normalized conditionally Gaussian approach delivered better estimation performance than any of the other approaches at all database complexities. The unnormalized conditionally Gaussian variant was a close second. The other approaches are relatively undifferentiated in terms of orientation estimation except for database complexities below 4.5 where the log-magnitude approach performs slightly worse than the others.

Fig. 12(b) shows the lowest achievable recognition error rate for each of the methods. This figure shows that the normalized conditionally Gaussian approach delivered performance equal to or better than all other methods for all complexities. The normalized

log-magnitude approach delivers equivalent performance over a small range of complexities near 4.5. Normalization improved the performance of each of the three approaches for all complexities with a decrease in error rate of approximately a quarter of a percentage point for most complexities.

## VI. CONCLUSIONS

This study extends to ten classes the performance-complexity results described by O'Sullivan, DeVore, Kedia, and Miller [20], and it extends the study to five other approaches. Exact descriptions of all algorithms were presented along with details of the methods by which the empirical results were obtained. None of these six algorithms are suggested for a direct implementation because key elements, such as optimal segmentation strategy, effects of SAR image resolution, the effects of varying depression angles, and characterization of performance under extreme conditions are not yet included in this study. Also, this work does not address methods of rejecting confuser vehicles as do [14, 15]. A framework was presented which provides a direct means of quantifying the cost and benefit, in terms of database complexity and performance, of various approaches to incorporating these elements. This same framework allows for the direct comparison of differing approaches to ATR from SAR images and was demonstrated in a study of six different algorithms, providing baseline results for each.

Normalization of the received SAR image to account for radar power fluctuation resulted in a significant increase in target recognition performance but did not significantly improve orientation estimation capability. The normalized conditionally Gaussian approach delivered superior performance for every database complexity in both target recognition and orientation estimation. The approach yielded a best case average orientation estimation error of  $6.62^\circ$  for  $128 \times 128$  images with 72 orientation windows and parameters trained over  $10^\circ$  intervals. The best case percentage of correctly classified images was 98.75% for  $128 \times 128$  images with 20 orientation windows and parameters trained over  $36^\circ$  intervals. The normalized variants of the conditionally Gaussian and quarter power approaches delivered their best performance with larger image sizes, suggesting that improved estimates of normalization scale factors offset the presence of additional clutter, and segmentation algorithms for those approaches may do well to retain pixels whose classification is not clear. However, both variants of the log-magnitude approach delivered their best performance with much smaller images which suggests that segmentation algorithms for those algorithms may do well to disregard pixels whose classification is ambiguous. The normalization results suggest that careful use of the surrounding

scene to more accurately determine fluctuating radar characteristics may prove useful in classifying the contents of the scene.

Current efforts are underway to explore the performance of ATR algorithms as functions of alternate measures of complexity such as the time to classify a SAR image chip [4], to compare other likelihood-based approaches [3]. Future efforts include investigating the use of optimal segmentation strategies and extending the study to investigate algorithm behavior under extended operating conditions [20] in which testing and training data are gathered under dissimilar conditions. Finally, the incorporation of confuser vehicles into the likelihood-based approaches and a comparison of the confuser rejection capability of each of the methods is planned. One possible model-based approach to confuser rejection is to set a probability of false rejection for each target class, and reject an image if the likelihood falls in the resulting critical region for the most likely class. Regardless of the confuser rejection method chosen, a non-zero probability of false rejection will affect the probability of correct classification.

## ACKNOWLEDGMENTS

The authors wish to thank Michael Bryant of Wright Laboratory who provided details about the quarter power approaches evaluated in this paper.

**MICHAEL D. DeVORE**  
**JOSEPH A. O'SULLIVAN**  
Electronic Systems and Signals  
Research Laboratory  
Department of Electrical Engineering  
Box 1127  
Washington University  
One Brookings Dr.  
St. Louis, MO 63130-4899

## REFERENCES

- [1] Bryant, M. L., Worrell, S. W., and Dixon, A. C. (1998) MSE template size analysis for MSTAR data. In E. G. Zelnio (Ed.), *Algorithms for Synthetic Aperture Radar Imagery V, Proceedings of SPIE*, **3370** (1998), 396-405.
- [2] Chiang, H. C., Moses, R. L., and Irving, W. W. (1999) Performance estimation of model-based automatic target recognition using attributed scattering center features. In *Proceedings of the 10th International Conference on Image Analysis and Processing*, Sept. 1999.
- [3] DeVore, M. D., Lanterman, A. D., and O'Sullivan, J. A. (2000) ATR performance of a Rician model for SAR images. In F. A. Sadjadi (Ed.), *Automatic Target Recognition X, Proceedings of SPIE*, **4050** (2000).
- [4] DeVore, M. D., O'Sullivan, J. A., Chamberlain, R. D., and Franklin, M. A. (2001) Relationships between computational system performance and recognition system performance. In F. A. Sadjadi (Ed.), *Automatic Target Recognition XI, Proceedings of SPIE*, **4379** (2001).

- [5] DeVore, M. D., Schmid, N. A., and O'Sullivan, J. A. (2000) Analytical and experimental performance-complexity tradeoffs in ATR. In *Proceedings of the Thirty-Fourth Asilomar Conference on Signals, Systems, and Computers*, Oct. 2000.
- [6] Donoho, D. L. (1995) De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, **41**, 33 (May 1995), 613–627.
- [7] Donoho, D. L., and Johnstone, I. M. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 3 (1994), 425–455.
- [8] Fukunaga, K. (1990) *Introduction to Statistical Pattern Recognition* (2nd ed.). New York: Academic Press, 1990, 76–77.
- [9] Grenander, U., Miller, M. I., and Srivastava, A. (1998) Hilbert-Schmidt lower bounds for estimators on matrix lie groups for ATR. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, 8 (1998), 790–802.
- [10] Heidbreder, G. R., and Mitchell, R. L. (1967) Detection probabilities for log-normally distributed signals. *IEEE Transactions on Aerospace and Electronic Systems*, **AES-3**, 1 (Jan. 1967), 5–13.
- [11] Jacobs, S. P., and O'Sullivan, J. A. (2000) Automatic target recognition using sequences of high resolution radar range-profiles. *IEEE Transactions on Aerospace and Electronic Systems*, **36**, 2 (Apr. 2000), 364–382.
- [12] Kaplan, L. M. (1999) Extended fractal analysis for texture classification and segmentation. *IEEE Transactions on Image Processing*, **8**, 11 (Nov. 1999), 1572–1585.
- [13] Kaplan, L. M., Murenzi, R., Asika, E., and Namuduri, K. R. (1998) Effect of signal-to-clutter ratio on template-based ATR. In E. G. Zelnio (Ed.), *Algorithms for Synthetic Aperture Radar Imagery V, Proceedings of SPIE*, **3370** (1998), 408–419.
- [14] Mossing, J. C., and Ross, T. D. (1998) An evaluation of SAR ATR algorithm performance sensitivity to MSTAR extended operating conditions. In E. G. Zelnio (Ed.), *Algorithms for Synthetic Aperture Radar Imagery V, Proceedings of SPIE*, **3370** (1998), 554–565.
- [15] Novak, L. M., Owirka, G. J., Brower, W. S., and Weaver, A. L. (1997) The automatic target recognition system in SAIP. *Lincoln Laboratory Journal*, **10**, 2 (1997), 187–201.
- [16] Novak, L. M., Owirka, G. J., and Weaver, A. L. (1999) Automatic target recognition using enhanced resolution SAR data. *IEEE Transactions on Aerospace and Electronic Systems*, **35**, 1 (1999), 157–175.
- [17] Osman, H. M., and Blostein, S. D. (1996) SAR imagery segmentation using probabilistic winner-take-all clustering. In E. G. Zelnio and R. J. Douglass (Eds.), *Algorithms for Synthetic Aperture Radar Imagery III, Proceedings of SPIE*, **2757** (1996), 217–226.
- [18] O'Sullivan, J. A., Blahut, R. E., and Snyder, D. L. (1998) Information-theoretic image formation. *IEEE Transactions on Information Theory*, **44**, 6 (Oct. 1998), 2094–2123.
- [19] O'Sullivan, J. A., and DeVore, M. D. (1999) Performance analysis of ATR from SAR imagery. In *Proceedings of the 33rd Annual Conference on Information Sciences and Systems*, Apr. 1999.
- [20] O'Sullivan, J. A., DeVore, M. D., Kedia, V., and Miller, M. I. (2001) Automatic target recognition performance for SAR imagery using a conditionally Gaussian model. *IEEE Transactions on Aerospace and Electronic Systems*, **37**, 1 (Jan. 2001), 91–108.
- [21] O'Sullivan, J. A., and Ryleeva, N. A. (2000) Method for reducing dimensionality in ATR systems. In F. A. Sadjadi (Ed.), *Automatic Target Recognition X, Proceedings of SPIE*, **4050** (2000).
- [22] Owirka, G. J., and Novak, L. M. (1994) A new SAR ATR algorithm suite. In *Algorithms for Synthetic Aperture Radar Imagery, Proceedings of SPIE*, **2230** (1994), 336–343.
- [23] Raemer, H. R. (1997) *Radar System Principles*. Boca Raton, FL: CRC Press, 1997.
- [24] Rissanen, J. (1986) Stochastic complexity and modeling. *The Annals of Statistics*, **14**, 3 (1986), 1080–1100.
- [25] Theera-Umpon, N., Khabou, M. A., Gader, P. D., Keller, J. M., Shi, H., and Li, H. (1998) Detection and classification of MSTAR objects via morphological shared-weight neural networks. In E. G. Zelnio (Ed.), *Algorithms for Synthetic Aperture Radar Imagery V, Proceedings of SPIE*, **3370** (1998), 530–540.
- [26] Weisenseel, R. A., Karl, W. C., Castañón, D. A., Power, G. J., and Douville, P. (1999) Markov random field segmentation methods for SAR target chips. In E. G. Zelnio (Ed.), *Algorithms for Synthetic Aperture Radar Imagery VI, Proceedings of SPIE*, **3721** (1999), 462–473.
- [27] Wissinger, J., Washburn, R. B., Friedland, N. S., Nowicki, A., Morgan, D. R., Chong, C., and Fung, R. (1996) Search algorithms for model based SAR ATR. In E. G. Zelnio and R. J. Douglass (Eds.), *Algorithms for Synthetic Aperture Radar Imagery III, Proceedings of SPIE*, **2757** (1996), 279–293.
- [28] Worrell, S. W., Parker, S., and Bryant, M. L. (1997) Class separability assessments and MSE algorithm robustness. In E. G. Zelnio (Ed.), *Algorithms for Synthetic Aperture Radar Imagery IV, Proceedings of SPIE*, **3070** (1997), 294–304.