HRR Automatic Target Recognition from Super-Resolution Scattering Center Features

Jon Gudnason *Member, IEEE*, Jingjing Cui and Mike Brookes *Member, IEEE*

Abstract—This paper presents a procedure for the automatic recognition of ground-based targets from high range resolution (HRR) profile sequences that may be obtained from a synthetic aperture radar (SAR) platform. The procedure incorporates an adaptive target mask and uses a super-resolution algorithm to identify the cross-range positions of target scattering centers. These are used to generate a pseudo-image of the target whose low-order discrete cosine transform coefficients form the recognizer feature vector. Within the recognizer, the states of a hidden Markov model are used to represent the target orientation and a Gaussian mixture model is used for the feature vector distribution. In a closed-set identification experiment, the mis-classification rate for ten MSTAR targets was 2.8%. The paper also presents results from open-set experiments and investigates the effect on recognizer performance of variations in feature vector dimension, azimuth aperture and target variants.

Index Terms—Automatic Target Recognition, Synthetic Aperture Radar, Hidden Markov Models, Super-Resolution, High Range Resolution

I. Introduction

This paper addresses the automatic recognition of ground-based targets from high range resolution (HRR) profile sequences such as those that may be obtained from a synthetic aperture radar (SAR) platform. Our proposed procedure incorporates several novel features that result in high performance.

In conventional SAR processing [2], the target image is obtained by taking the discrete Fourier transform of a sequence of HRR profiles uniformly spaced in cross-range. Fig. 1(a), taken from the MSTAR database [32], shows a sequence of HRR profiles from a T72 tank and Fig. 1(b) shows the resultant image obtained by taking the Fourier transform in the horizontal, or cross-range, direction. At sufficiently short wavelengths, the radar return from a target may be represented as arising from a set of discrete scattering centers [19], [31] each of which corresponds to a single cross-range complex spectral component in the HRR profile sequence. In the recognition algorithm described below, we identify these spectral components using a super-resolution spectral analysis technique, the MUSIC (multiple signal classification) algorithm [35]. The use of a super-resolution peak-picking algorithm to identify spectral components in the HRR sequence provides inherent resistance to the low-level spectral noise arising from radar clutter. We further improve the recognition of compact targets by discarding any scattering centers that lie outside an adaptive target mask, shown in Fig. 1(b) as the inner rectangle, which we determine from the SAR image.
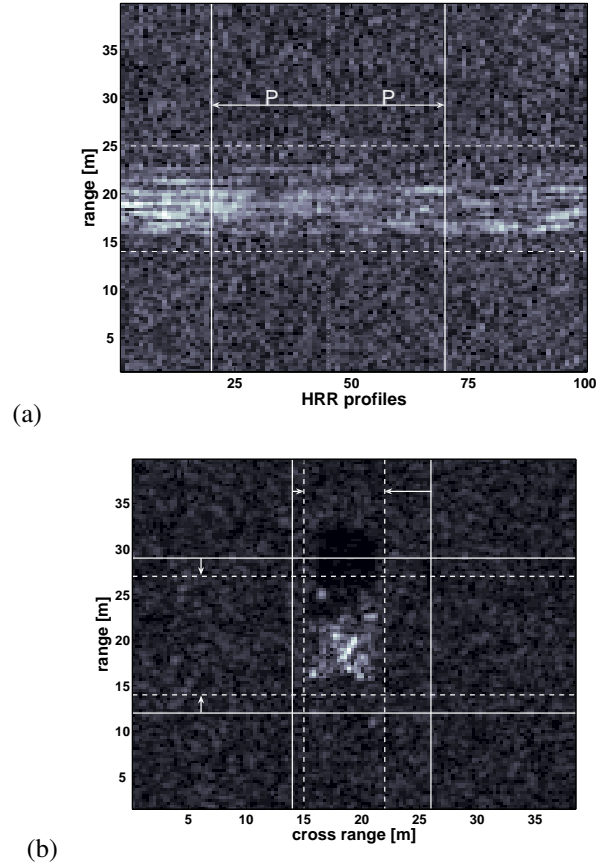
(a)

(b)

Fig. 1. (a) A sequence of 100 HRR profiles from a T72 tank in the MSTAR database and (b) the resultant SAR image. The solid lines in the SAR image show the initial processing region and the broken horizontal lines show the smaller adaptive mask. The vertical lines in the HRR profile image shows the limit of the HRR sequence used for extraction of a single feature vector.

Rather than performing recognition directly on the list of extracted target scattering centers, we instead use them to create a pseudo-image of the target and use the low-order discrete cosine transform coefficients of this pseudo-image as our feature vector. This indirect procedure circumvents problems arising from the ordering or incompleteness of the scattering center list. An additional benefit is that the cross-range resolution of the pseudo-image can be chosen independently of the azimuth aperture of the HRR profile sequence. Since any target motion during SAR data acquisition will result in image blurring [30], basing each feature vector on a reduced number of HRR profiles will reduce the sensitivity to target motion. The solid vertical lines in Fig. 1(a) indicate

the sequence of $2P + 1$ profiles used to generate a single feature vector.

Following a literature survey in Sec. II we present the proposed feature extraction method in Sec. III. The recognition architecture is described in Sec. IV and the experimental results are presented in Sec. V. The method is tested using the MSTAR database [32] and results are shown for closed-set identification, varying feature vector size, different number of hidden Markov Model (HMM) states, robustness of target variants, feature vector aspect aperture and open-set identification. The paper concludes in Sec. VI with a summary and conclusions. An earlier version of the algorithm was described in [7] and [8] which also investigated the use of shadow features for recognition.

II. FEATURES FOR AUTOMATIC TARGET RECOGNITION

The reliable recognition of ground-based targets using SAR data has been the subject of research for many years. The most straightforward approach is to compare the observed SAR image with a set of target templates in the image domain using either real or complex pixel intensities. The comparison may be based either on Euclidean distance [27] or on a likelihood measure derived from an assumed stochastic model [11], [29]. In some cases images are transformed prior to the comparison in order to make clutter distributions more Gaussian [12], remove inter-pixel correlations [39] or to obtain features that are invariant under rotations and/or translations [5], [20], [24].

An alternative approach is to model the radar return as emanating from a finite set of discrete scattering centers [19], [31]. Most commonly, these are treated as point scatterers but some authors have used more sophisticated models [41], [14]. Identifying point scattering centers is most conveniently performed in the frequency domain using super-resolution spectral analysis. Early workers generally used Prony's method [16], [3] but more recently, the RELAX [21] or MUSIC [15], [28] algorithms have been preferred. An extensive survey of super-resolution algorithms for SAR imaging is included in [9].

Advantages of the scattering center approach include improved noise immunity, since only the peaks in the spectral domain are modeled, and the opportunity to use compact target models consisting only of a list of scattering centers. A difficulty with the approach, however, is the need to compare two unordered scattering center lists that may each contain omissions or insertions. Methods of comparing such lists are discussed in [1], [6], [36].

The recognition of moving targets from SAR images is especially difficult since target motion during acquisition will result in image blurring. In response, a number of workers have proposed forming recognition features from individual HRR profiles [22], [23]. The approach is robust to target motion but, on static targets, performance is much worse than for SAR-based approaches because an individual HRR profile gives no information about the cross-range structure of the target.

Radar signatures can vary greatly with small changes in target orientation (also known as aspect or pose). Template-matching recognition systems generally cope with this by representing azimuth intervals with templates covering between $1°$ and $30°$ in azimuth [17]. In situations where a sequence of observations is available, such as with HRR-based recognizers, it is possible to impose a smoothly changing target orientation by modeling it with a Hidden Markov Model (HMM) [34], [22].

III. TARGET SCATTERING CENTER FEATURES

In this section we present the feature extraction process which consists of four stages: adaptive target mask formation, scattering center extraction, pseudo-image construction and feature vector calculation.

A. Definitions

The features that we use for target recognition are derived from the sequence of complex-valued HRR profiles, $x(n, k)$, which may be obtained from a complex SAR image chip by the procedure described in [22]. The profile index is denoted by $n \in \{1, \ldots, N\}$ and the range-bin index by $k \in \{1, \ldots, K\}$. Fig. 1(a) shows a typical plot of $|x(n, k)|$ and Fig. 1(b) shows the SAR image that results from windowing $x(n, k)$ and taking the DFT with respect to n. We do not normally use all N HRR profiles to derive a single feature vector. Instead we divide them into $N - 2P$ overlapping sets of $2P + 1$ profiles and determine a feature vector from each set. The feature vector centered on profile n is thus based on the profile subset, $x_n(p, k)$, defined by

$$x_n(p, k) = x(n + p, k) \qquad \text{where} \qquad p \in \{-P, \ldots, P\}. \quad (1)$$

We will frequently omit the subscript n when discussing the processing associated with an individual feature vector. In Fig. 1(a), the vertical lines indicate the range of profiles used to derive a single feature vector when $P = 25$. Selecting a small value for P will reduce the impact of target motion and azimuth-dependent signature variation but will result in increased coefficient noise; the effect of P on recognition performance is evaluated in Sec. V-F.

B. Adaptive Mask Identification

We assume that the target recognition is carried out after target detection. If the target detection procedure does not give an accurate location of the target center or the target scattering center region then an adaptive mask identification is necessary. This is also the case for the MSTAR database. The solid lines shown in the SAR image of Fig. 1(b) define the initial target region-of-interest and it can be seen that the target lies within this region but is offset slightly to the left. Using the procedure described below, we determine the rectangular target mask that is shown on the image using dashed lines. It can be seen that this mask fits snugly around the target and includes all its true scattering centers. This target mask is used to eliminate any false scattering centers that may be identified from the clutter region. The positions of the extracted target scattering centers are compared with the mask and any centers that lie outside the mask are assumed to arise from clutter and are discarded.

To identify the target mask shown with dashed lines in Fig. 1(b), we first form an image from the HRR profiles. As

described above, the number of profiles used to form each feature vector is $2P+1$ where P is a system parameter. These $2P+1$ profiles are used to form an image of the target whose central portion is shown in Fig. 2(a). To determine the target mask, we segment the pixels of this image into target and background classes using a maximum likelihood criterion and we then define the target mask to be the smallest bounding rectangle that encloses all pixels assigned to the target class.

To classify the pixels into a target class, ξ_1, and a background class, ξ_2, we assume that within either of these classes, the complex valued pixels are taken from a complex Gaussian distribution with zero mean and variance σ_κ^2 in both real and imaginary parts. Thus the log-likelihood of the complex-valued pixel $x(i,j)$, conditioned on the class ξ_κ ($\kappa = 1,2$), is given by

$$
\begin{aligned}
\ell(x(i,j)|\xi_\kappa) &= \log\big(\mathcal{N}(x_\Re(i,j);0,\sigma_\kappa^2)\big) \\
&\quad + \log\big(\mathcal{N}(x_\Im(i,j);0,\sigma_\kappa^2)\big) \\
&= -\log(2\pi\sigma_\kappa^2) - \frac{|x(i,j)|^2}{2\sigma_\kappa^2}
\end{aligned}
\tag{2}
$$

where $\mathcal{N}(x;\mu,\sigma_\kappa^2)$ is a Gaussian distribution with mean μ and variance σ_κ^2 and $x_\Re(i,j)$ and $x_\Im(i,j)$ are the real and imaginary parts respectively of the pixel $x(i,j)$. We initialize σ_1^2 by computing the variance of a small window in the middle of the target region and similarly we initialize σ_2^2 from pixels near the edge.

In the following expectation maximization [10] iterative procedure both the partitioning of the image pixels and the two variances, σ_κ^2, are determined. We use (2) for the expectation step. For the maximization step, we classify each image pixel and recalculate the variances from that classification. The classification is based on the smoothed log-likelihood

$$
\kappa(i,j) = \underset{\kappa}{\mathrm{argmax}}\,\big(w(i,j) * \ell(x(i,j)|\xi_\kappa)\big)
\tag{3}
$$

where $*$ denotes 2-dimensional convolution and $w(i,j)$ is the impulse response of a smoothing filter which introduces correlation into the classification of adjacent pixels. The choice of $w(i,j)$ is a compromise between the elimination of isolated strong responses from the clutter region and the preservation of small regions that truly belong to the target. In the results described below, we have used a uniform 3×3 window for the smoothing filter since it gives good correlation between neighbouring pixels but maintains spatial sharpness for isolated target pixels.

Using the partition defined by (3), we recalculate the class variances, σ_κ^2, and re-evaluate $\kappa(i,j)$ for 10 iterations or until the total expectation

$$
\sum_{i,j} \ell\big(x(i,j)|\xi_{\kappa(i,j)}\big)
\tag{4}
$$

ceases to increase.

Finally, we introduce class probabilities p_κ satisfying $p_1 + p_2 = 1$ and choose them so that the partition

$$
\kappa'(i,j) = \underset{\kappa'}{\mathrm{argmax}}\,\big(w(i,j) * \ell(x(i,j)|\xi_{\kappa'(i,j)}) + \log(p_{\kappa'})\big)
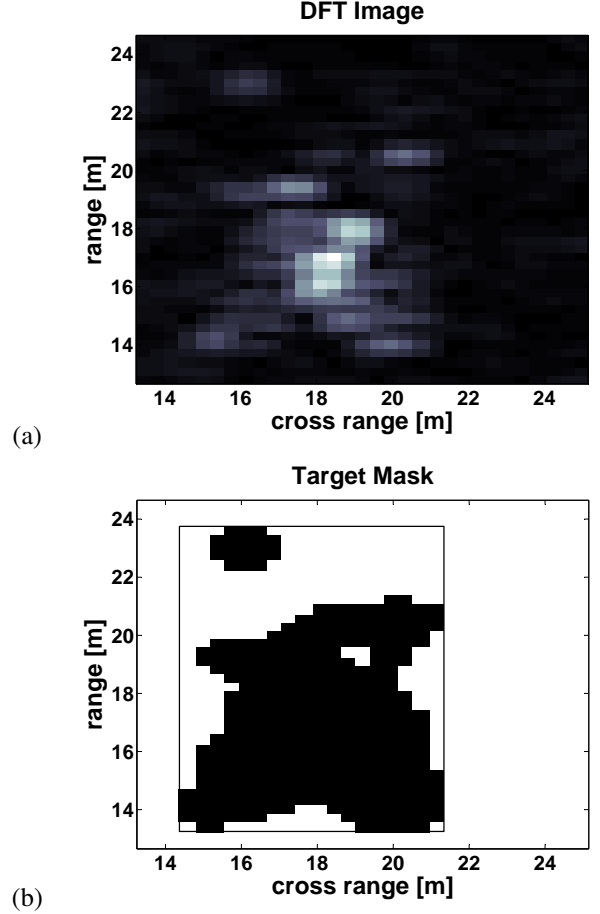\tag{5}
$$

(a)

(b)

Fig. 2. (a) SAR image formed using $2P+1 = 51$ HRR profiles and (b) the pixels in the "target" class with the final target mask as a bounding rectangle.

maximizes the smoothed log likelihood summed over all pixels.

Fig. 2(b) shows the resultant partitioning of the pixels together with the bounding rectangle that is used as the final target mask. It is clear that for this image the limits of the target have been identified correctly.

C. Forming the Feature Vector

The first step in forming the feature vector is to identify the target scattering centers within each range bin by means of the super-resolution MUSIC algorithm [35]. It works on the assumption that the signal comprises of a fixed number of exponentials in the presence of noise and the algorithm provides an estimate of the frequency location and magnitude of these exponentials using eigenvalue decomposition of the signal's autocorrelation matrix. The signal model is given by

$$
x(p,k) = \sum_{m=1}^{M_k} a_{m,k} e^{j\omega_{m,k}p} + v(p,k)
\tag{6}
$$

where k is the (constant) range-bin index, $p \in \{-P,\ldots,P\}$ is the HRR profile index (the signal index), M_k is the number of extracted scattering centers (exponentials) in the range bin, $a_{m,k}$ is the radar cross section (magnitude) of the mth scattering center of the kth range bin, $\omega_{m,k}$ corresponds to

its azimuth (frequency) location and $v(p,k)$ is assumed to be white noise. The use of a super-resolution algorithm only in azimuth reduces the computational complexity considerably and avoids the need to modify the normal SAR data processing stages.

Having identified the scattering centers in each range bin, we now use them to construct a pseudo-image of the target. For each range bin, k, we transform the list of extracted scattering centers into one row of the pseudo-image by low-pass filtering and sampling a signal in which each of the M_k scattering centers is represented by an impulse according to

$$y(l,k) = \sum_{m=1}^{M_k} |a_{m,k}|^2 \, h(l - \theta \omega_{m,k}) \qquad (7)$$

where the low-pass filter response is given by

$$h(l) = \frac{sin(\pi l)}{\pi l}. \qquad (8)$$

The cross-range index is denoted by l and the scaling constant θ given by $\theta = \lambda \left(4\pi \Delta \phi \Delta r\right)^{-1}$ where λ is the wavelength, $\Delta \phi$ is the azimuth increment of the HRR profiles and where Δr is chosen to match the cross-range resolution of the pseudo-image. The cross-range resolution is independent both of P and of the original azimuth aperture and may be conveniently chosen to match the range resolution.

Fig. 3 shows the reconstructed image both (a) with and (b) without the use of the adaptive mask. Two of the pixels that lie outside the mask are explicitly identified in Fig. 3(b) and it can be seen that the corresponding impulses are missing in Fig. 3(a). Also visible in this figure is the effect of the cross range low-pass filter impulse response (8). The smearing introduced by the filter is exaggerated by the log-intensity scale of the figure.

Unfortunately, when the target mask is applied, it is possible for some range bins to contain no valid scattering centers at all. These "black lines" can occur in any of the range bins, but most commonly arise in those range bins where there is no target. The pixels in that range bin will be identically zero whenever this happens and, because of the log operation in (9) below, these pixels will dominate the feature vector resulting in poor performance. To avoid this situation, we detect any range bin, k, that is empty of scattering centers and, if necessary, increase the sensitivity so that at least $M_k = 1$ (6) scattering center is found that lies within the cross-range limits of the target mask. The scattering centers that are introduced are always of very low magnitude and we found that this approach gave better recognition results than clipping the pixel intensities to a lower bound.

Fig. 3(a) shows the reconstructed target image before this black line elimination procedure has been applied and it can be seen that there are black lines visible at the top and bottom of the image. In Fig. 3(c) additional scattering centers have been allowed and each range bin now contains at least one scattering center. The added scattering centers are by definition of low intensity.

Following the image reconstruction process, the target feature vector is formed by taking the low frequency coefficients
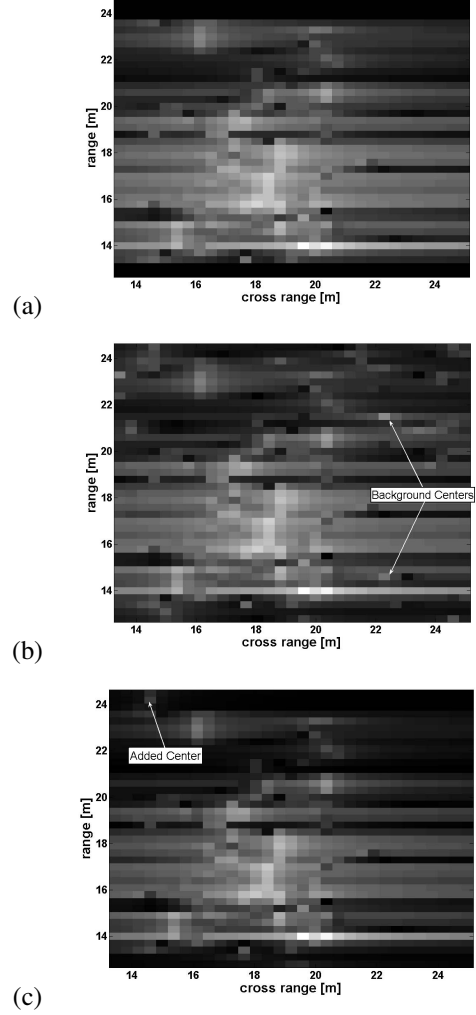

(a)

(b)

(c)

Fig. 3. Reconstructed target images (a) with and (b) without the use of a target mask. Two of the excluded scattering centers have been explicitly identified in (b). Image (c) shows the reconstructed target image with an added scattering centers in black lines.

of the 2-dimensional discrete cosine transform (DCT) of the log of the pseudo-image $y_n(l,k)$,

$$f_n(\eta, \zeta) = \qquad (9)$$
$$\sum_{k=0}^{K} \sum_{l=0}^{L} \alpha_k \beta_l \, \log\left(y_n(l,k)\right) \cos\left(\tfrac{\pi(2\eta+1)k}{2K}\right) \cos\left(\tfrac{\pi(2\zeta+1)l}{2L}\right),$$

where $\alpha_k = 1/\sqrt{2}$ and $\beta_l = 1/\sqrt{2}$ if $k = 0$ or $l = 0$, respectively, but $\alpha_k = \beta_l = 1$ otherwise. The low-order coefficients of $f_n(\eta, \xi)$ are retained and used as a feature vector $\mathbf{u}_M(n)$.

IV. RECOGNITION ARCHITECTURE

As noted in Sec. III-A, a sequence of N HRR profiles will result in $N - 2P$ feature vectors which represent the radar return from a sequence of viewpoints that necessarily changes smoothly. In order to impose this constraint in the recognizer, we use a Hidden Markov Model (HMM) whose states represent the target orientation. Each of the S states in the HMM corresponds to a contiguous range of target

orientations and hence a distinct target signature, since the radar signature varies with orientation. Two successive feature vectors may only correspond either to the same state or to adjacent states. The parameters of the HMM specify a statistical description of the target signature within each state and also a set of inter-state transition probabilities. The number of HMM states, S, must be chosen to be large enough to model adequately the variation of target signature with orientation but small enough to ensure that sufficient training observations are available for each state. The angular resolution that is required for the HMM may be estimated from the ratio of the radar range resolution to the maximum target dimension. If, for example, the range resolution is 1 m and the target extends to a maximum of 10 m from its center, we would require an angular resolution of about $6°$ to ensure that a scattering center occupies only one range bin in any state. For ground-based or sea-based targets whose rotation is restricted to a single axis, a resolution of $6°$ would thus require $S = 60$ states in the HMM.

To define the initial model parameters, we assume that all the HMM states correspond to equal aspect ranges of $360°S^{-1}$. For each HMM state we train a Gaussian mixture model (GMM) using all available training data from the corresponding range of aspects of a particular target. The transition probabilities are initialized to the ratio of the azimuth increment between successive feature vectors and the aspect range of the states. Using these initial values, the HMM is then trained using Baum-Welch training [10], [40] until the HMM parameters converge.

The azimuth interval represented by an individual state can change substantially during the training process according to how rapidly the feature vector varies in azimuth. This is demonstrated in Fig. 4(a) which shows the smoothed log likelihood of test feature vectors as a function of azimuth angle for each of three consecutive model states after training is complete. The three model states were initially trained with data from consecutive $6°$ azimuth intervals in the region of $60°$. We see from Fig. 4(a) that the log likelihood does indeed peak at a target orientation of $60°$ and that there is a second peak at around $250°$ due to target symmetry. An enlarged view of the primary peak is shown in Fig. 4(b) which shows that each state has retained clearly defined boundaries. State 10 is however much broader than its initial width of $6°$ and now covers about $10°$, state 11 covers a relatively narrow angular extent about $2°$ and state 12 becomes broader again. The mean target orientation corresponding to the center of a particular state can be determined, if required, by taking the centroid of the corresponding primary peak in Fig. 4(a).

V. EXPERIMENTAL EVALUATION

A. Database and Evaluation Methodology

The MSTAR data set was collected by the Sandia National Laboratory (SNL) SAR sensor platform [32]. The collection was jointly sponsored by US Defence Advanced Research Projects Agency DARPA and Air Force Research Laboratory as part of the Moving and Stationary Target Acquisition and Recognition (MSTAR) program. SNL used an X-band
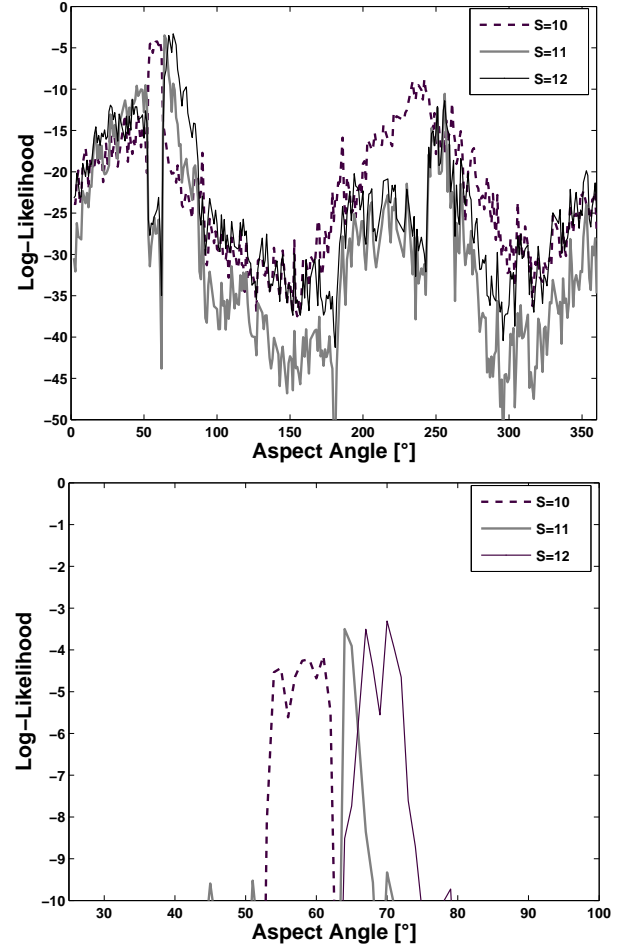

Fig. 4. The log-likelihood of a target for state $S = 10$, $S = 11$, and $S = 12$ shown (a) from $0°$ to $360°$ and (b) from $25°$ to $100°$.
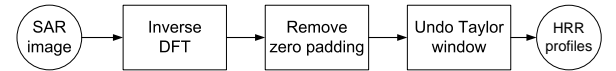

Fig. 5. Procedure to convert SAR images to HRR profiles. The operations are in cross-range.

SAR sensor in 0.3 m resolution spotlight mode. The MSTAR public-release dataset consists of 10 target classes whose identification codes are listed in Table 1. The dataset contains only a single example for most of the target classes but for two of the classes, T72 and BMP2, it includes multiple variants which differ in such details as antenna position or the presence of fuel drums [32]. These target variants are distinguished by suffices in Table III. Fig. 1 shows (a) a sequence of HRR profiles and (b) a SAR image of the tank T72 at a depression angle of $17°$ and azimuth angle of $91.7°$. The SAR image is a 128×128 pixel array representing a scene that is approximately 38×38 m^2 with image resolution of 0.3 m in both range (vertical) and cross range (horizontal) direction.

The final few stages of the image formation process must be inverted to obtain the HRR profiles that are needed to form the feature vector sequence. An image can be converted into a sequence of about 100 aspect-dependent HRR profiles with the procedure summarized in Fig. 5 [22]. The angular spacing

of the HRR profiles is approximately $0.03°$ so a single image covers $100 \times 0.03° = 3°$ in azimuth aperture.

Each image from the MSTAR database was converted into a sequence of HRR profiles. The training set includes a total of 2747 SAR images at $17°$ depression angle and is used to train ten target models, with approximately 270 sequences per target. For testing, we use a separate set of 2474 SAR images at $15°$ depression angle, with about 250 for each target. We do not apply any compensation for the slight mismatch in depression angle.

Each target feature vector contains 54 elements that are chosen from the low order coefficients of the DCT image that satisfy $0 < \eta + \zeta \leq 10$ in (9). For $P = 25$, each feature vector is formed from a consecutive sequence of 51 HRR profiles corresponding to an aperture of $1.5°$. For most test sequences, the number of HRR profiles is $N = 100$ which corresponds to an azimuth aperture of $3°$ and which results in $N - 2P = 50$ feature vectors. The maximum number of scattering centers, M, is fixed at 10 for each range bin. The HMM contains $S = 60$ states which, after training, correspond to different target orientation ranges. Within each state, feature vector distributions are represented by diagonal-covariance Gaussian mixtures. Model training and recognition are performed using the HTK recognition software [40].

We present the results of our evaluations for both closed-set and open-set identification experiments. For the closed-set identification experiments, we take each observation sequence, use the HMM to determine the model with the highest likelihood and characterize the performance in the form of test set misclassification rate (MCR). For the open-set identification experiments, we classify an observation sequence as an identified target if its log-likelihood with any target model exceeds a given threshold. By varying the threshold, we demonstrate the performance through the use of receiver operating characteristic (ROC) curves [33].

We note that our recognizer also implicitly determines the orientation of the target. The HMM state with highest probability can be obtained directly from the recognizer. The mapping between states and aspect angles is not uniform but can be obtained during training by determining, for each state, the centroid of the principal peak in the graph shown in Fig. 4(a).

B. Closed-set experiments

In the closed-set experiments, we compare the performance of different feature sets, demonstrate how varying the feature vector dimension, the HMM parameters and the feature vector aspect aperture affect the recognition performance and also explore how robust the approach is to target variants.

We assess the choice of the MUSIC algorithm for determining the cross range position of the scattering centers by comparing its identification performance with that obtained when using the discrete Fourier transform (DFT) ($\mathbf{u}_F$) or autoregressive modeling ($\mathbf{u}_A$) for spectral estimation in cross range after the adaptive mask operation. We denote the features derived using the MUSIC algorithm as $\mathbf{u}_M$. The derivation of $\mathbf{u}_F$ and $\mathbf{u}_A$ is explained below and illustrated in Fig. 6.
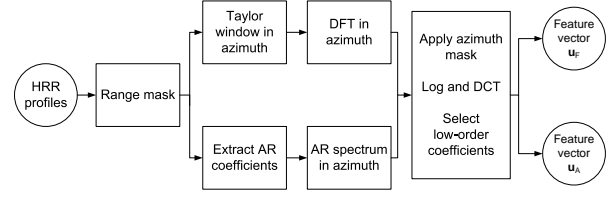

Fig. 6. Fourier transform and AR modelling-based feature extraction procedure.

TABLE I
MISCLASSIFICATION RATE (%) FOR CLOSED-SET TARGET IDENTIFICATION. THE TABLE COMPARES THE PERFORMANCE OF THREE TARGET FEATURE SETS.

Target	$\mathbf{u}_F$	$\mathbf{u}_A$	$\mathbf{u}_M$
BMP2	5.2	4.1	0.5
BRDM2	12.4	10.9	2.6
BTR60	4.2	3.6	0.5
BTR70	2.6	3.1	0.5
D7	1.5	1.5	5.8
T62	9.6	10.3	7.7
T72	0.0	0.5	1.5
ZIL131	3.7	1.8	4.4
ZSU234	3.7	2.9	1.1
2S1	18.7	15.7	1.5
MCR	6.5	5.8	2.8

To obtain $\mathbf{u}_F$, we window the HRR profile sequence in cross range using a Taylor window $w_T(p)$ [2] and take the DFT in each range bin lying within the target mask. The resultant image is therefore defined as

$$y_F(l,k) = \Big| \sum_{p=-P}^{P} w_T(p)x(p,k)e^{-j\frac{2\pi}{2P+1}pl} \Big|. \quad (10)$$

We replace $y(l,k)$ with $y_F(l,k)$ in (9) to obtain the low order coefficients $\mathbf{u}_F$.

To obtain $\mathbf{u}_A$, we apply autoregressive (AR) spectral estimation using the model

$$x(p,k) = -\sum_{m=1}^{M} a_m(k)x(p-m,k) + v(p,k) \quad (11)$$

where $a_m(r)$ are complex AR coefficients and $v(p,k)$ is a white noise input of zero mean and unit variance. The AR coefficients are estimated through the covariance method [18], because it does not need windowing and, like MUSIC, provides super-resolution spectral estimates. We form a pseudo-image by converting the AR coefficients $a_m(r)_{m=1}^{M}$ to a spectrum in the p direction

$$y_A(p,k) = \frac{1}{\Big| \sum_{m=1}^{M} a_m(k)e^{-j2\pi mp/M} \Big|}. \quad (12)$$

which convert to a feature vector, $\mathbf{u}_A$, using (9).

Table I shows the misclassification rate (%) for a closed-set identification tasks using 10 different targets for each of the three feature sets. The table presents the target misclassification rate for each target and the overall test set misclassification rate (MCR). The MCR differs slightly from the average target MCR because the number of test sequences is not the same

TABLE II
CONFUSION MATRIX FOR THE CLOSED-SET TARGET IDENTIFICATION USING THE MUSIC DERIVED TARGET FEATURE SET ($\mathbf{u_M}$). THE TABLE ALSO
SHOWS THE INDIVIDUAL TARGET MISCLASSIFICATION RATE (MCR), TARGET MODEL MISTRUST RATE (MTR) AND THE TEST-SET MISCLASSIFICATION
RATE OF 2.8%.

| | Recognized as | | | | | | | | | | MCR [%] |
Targets	BMP2	BRDM2	BTR60	BTR70	D7	T62	T72	ZIL131	ZSU234	2S1	
BMP2	193	0	1	0	0	0	0	0	0	0	0.5
BRDM2	0	189	1	2	2	0	0	0	0	0	2.6
BTR60	0	0	193	0	1	0	0	0	0	0	0.5
BTR70	0	1	0	191	0	0	0	0	0	0	0.5
D7	0	0	2	1	258	5	0	7	0	1	5.8
T62	1	4	3	7	0	253	0	1	4	1	7.7
T72	0	1	0	1	0	0	270	1	1	0	1.5
ZIL131	2	1	1	1	3	1	0	261	3	0	4.4
ZSU234	0	0	0	0	1	0	0	2	271	0	1.1
2S1	0	0	0	0	0	0	2	2	0	270	1.5
MTR [%]	1.5	3.6	4.0	5.9	2.6	2.3	0.7	4.7	2.9	0.7	2.8

for each target. We can see that the MUSIC derived feature set ($\mathbf{u_M}$) performs better than the other two methods ($\mathbf{u_F}$ and $\mathbf{u_A}$) with an MCR of 2.8%. It is also substantially lower than the 4.2% obtained from the target features without applying the adaptive mask [8].

Table II shows the confusion matrix for the target identification experiment using the MUSIC derived target feature set ($\mathbf{u_M}$). Each row shows how many test sequences from an individual targets were classified. The last column shows each target's misclassification rate and the last row shows the mistrust rate of each target model. For example, 1.5% of T72 test sequences were wrongly classified and of all test sequences classified as T72 only 0.7% were not that target.

The error rate of 2.8% can be compared with other results using the MSTAR database. Comparison can be made between the error rate of 4.1% presented in the work of Novak et. al. which used a super-resolution approach on the SAR images with a 10-target identification task [26]. Another super-resolution approach was presented by Cetin et. al. who achieved a 0.85% error rate on a 3-target identification task [4]. An average error rate of 8.6% was achieved by Nilubol and Mersereau using Radon transformed SAR images as inputs to HMMs for a 7-target identification task [24]. The best performance achieved when using HRR profiles directly were 17.6% [38] and 17.8% [22]. We note however that since these techniques rely on individual HRR profiles, their methods will be considerably less sensitive to target motion than our technique unless the value of P is made very small.

C. Influence of HMM structure

Increasing the number of HMM states (S) and Gaussian mixtures (w) results in a richer model that is better able to model the target signature. However, if insufficient training data is used, a richer model is likely to suffer from overtraining and will fail to generalize well to unseen test data. Fig. 7 shows the test set misclassification rate for the 10-target closed-set identification experiment using different number of HMM states, S, and Gaussian mixture components, w. When $w = 1$ the best performance is achieved with $S = 60$ and degrades for larger or smaller S due to overtraining or undermodeling respectively. For the case $w = 2$, the optimum
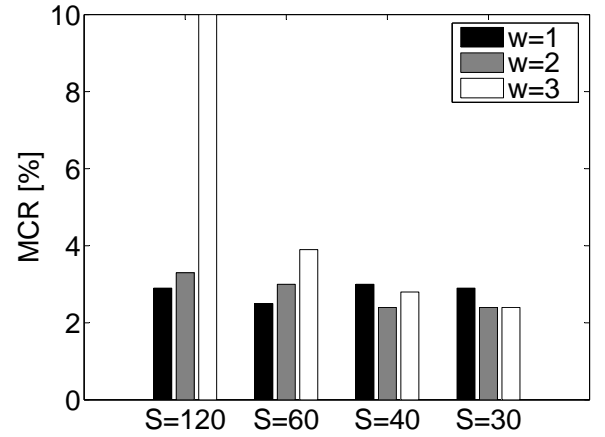

Fig. 7. MCR with S =120, 60, 40 and 30; w =1,2 and 3.

performance is slightly improved and occurs for $S = 40$. When $w = 3$, overtraining is much more evident and the optimum performance occurs for $S \leq 30$. The benefit of using additional mixtures with the training data available in MSTAR is very small and in the remaining experiments we retrained our initial configuration of $S = 60$ and $w = 1$.

D. Influence of the feature vector dimension

To investigate the influence of feature vector dimension, we constructed the feature vector by choosing the coefficients, $f_n(\eta, \zeta)$ from (9) with the highest between-class to within-class covariance matrix trace ratio. Here we define a "class" to mean a single Gaussian mixture component in a particular state of a particular target model [13, Chapter 3]. The between-class covariance matrix is the covariance matrix of the class mean vectors and the within-class covariance matrix is an average of the class covariances. A coefficient $f_n(\eta, \zeta)$ with a high ratio can therefore discriminate better between the classes.

Fig. 8 plots the test set misclassification rate of the 10-target closed-set identification experiment with the feature vector dimension varying from 20 to 100. It can be seen that feature sizes below 40 show an increased error rate because of undermodeling and that feature sizes above 90 show some signs of overtraining but that within these limits the size is not
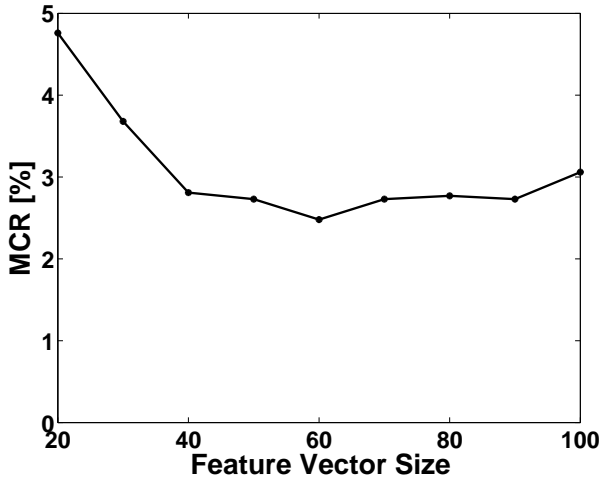
Fig. 8. MCR using different number of coefficients to form the feature vector.

	3° aperture		6° aperture		
Target	Strict	Class	Strict	Class	Class from [22]
BMP2-c21	16.0	1.5	9.8	0.0	6.1
BMP2-9563	15.5	1.6	11.4	0.5	7.0
BRDM2	7.3	-	1.5	-	8.6
BTR70	0.0	-	0.0	-	14.1
BTR60	2.6	-	1.0	-	13.7
D7	1.1	-	0.0	-	4.8
T62	5.1	-	1.1	-	9.0
T72-a04	11.7	1.1	4.7	0.4	2.9
T72-a10	10.3	0.4	4.1	0.0	3.4
T72-a62	12.8	1.1	5.8	0.0	3.9
T72-132	2.1	0.0	0.0	0.0	7.6
ZIL131	0.7	-	0.0	-	10.3
ZSU234	1.5	-	0.0	-	8.5
2S1	4.0	-	2.6	-	4.7
MCR	**6.5**	**2.0**	**3.0**	**0.5**	**7.0**
BMP2-9566	-	6.6	-	3.6	37.5
T72-a05	-	1.5	-	0.0	14.5
T72-a07	-	3.3	-	0.7	16.1
T72-a32	-	4.4	-	2.6	22.9
T72-a63	-	0.0	-	0.0	15.9
T72-a64	-	*12.8*	-	*7.7*	*31.7*
T72-s7	-	*13.8*	-	*10.6*	*17.1*
T72-812	-	*10.9*	-	*9.3*	*29.6*
MCR	-	**6.7**	-	**4.3**	**23.2**

critical. In the remaining experiments we therefore retained our original feature vector comprising a predefined set of 54 low-order DCT coefficients.

E. Robustness to target variants

The MSTAR public release dataset includes 11 variants of the T72 tank and 3 variants of the BMP2 vehicle, manifested by, for example, different configurations of fuel tanks and antennae [32], [22]. To evaluate the robustness of the recognizer to these variations, we train models on two BMP2 variants and four T72 variants as well as the eight other targets, and conduct recognition tests on all 22 targets in the database. If the recognizer identifies an incorrect variant of the correct tank model, it is counted as an error in the "Strict" column of Table III but as a correct identification in the "Class" column. The misclassification rate (MCR) is given separately for the training-set targets and the unseen targets in rows 15 and 24 respectively.

The average class error rate for targets included in the training set is 2.0% using test sequences covering 3° in azimuth, as shown in the first pair of columns of Table III. In most cases the system was able to identify the precise variant of a particular target correctly. For the unseen variants listed in the lower section of the table, the average class error rate is 6.7% with over 70% of the errors arising from T72-a64, T72-s7 and T72-812 (italic in table). The difference in performance between the T72 variants can be explained by their configurations as detailed in [37] which also defines which variants are seen and unseen. The variants a62, a63, a64 and s7 are similarly configured but a62 is the only example in the training set of its kind. This could explain the poor performance of s7. The 812 variant includes structures not found on any of the other vehicles [37] and this may explain why it is not recognized as well as the others.

To evaluate the benefits of extended target observation, we combine the results of two consecutive sequences covering an overall azimuth aperture of 6° by adding their likelihoods. We see in the second pair of columns that the classification

performance improves considerably. The average class error rates for seen and unseen targets decrease to 0.5% and 4.3% respectively.

The final column of Table III shows the "Class" error rates obtained by Liao et. al. [22] for the same 6° azimuth aperture task but basing recognition on individual HRR profiles with 0.1° azimuthal sampling. The increased error rates in this approach represent the price paid for insensitivity to target-motion. Another study by Novak et. al. [25] obtained an average class error rate of 4.7% using SAR images for a 6-target identification task against 10 models.

F. Influence of feature vector azimuth aperture

We now investigate the sensitivity of the recognition technique to the number of HRR profiles, $2P + 1$, used to derive each feature vector. In the experiments described above, P was set to 25 corresponding to an azimuth aperture of $1.5°$. We expect that smaller values of P will make the feature set less sensitive to target motion but will give worse performance for static targets due to increased coefficient noise. It is therefore of interest to see how performance varies with the choice of P. The maximum number of scattering centers, M, is chosen to be $\lfloor (2P + 1)/5 \rfloor$ for each range bin. The total number of HRR profiles, N, used for each test remains constant and so the number of extracted feature vectors, $N - 2P$ increases as P is reduced. Fig. 9 shows how varying P affects the test set misclassification rate of the closed-set 10-target identification performance. We see that reducing P to 15 has very little effect
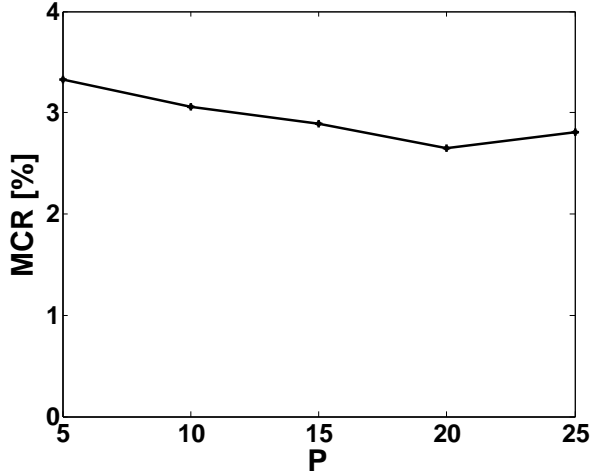
Fig. 9. MCR using different number of HRR profiles to derive the feature vectors.

on performance and that reducing it still further to 5 results in only a modest increase in error rate. We therefore expect that our recognizer will perform well for targets moving at moderate speeds.

G. Open-set experiments

The experiments represented above are based on closed-set identification where the observation is assumed to be of one of the known targets. It is also important that a recognizer should be able to reject a confuser. As in the closed-set experiments, a test sequence is compared against each of the 10 target models to give a list of likelihoods but the observed sequence will be rejected as a non-target if the highest likelihood falls below a predefined threshold. Otherwise, the observation is accepted as a genuine target, whose identity is that of the target model having the highest likelihood. This procedure gives rise to three performance measures that we present below: (1) probability of detection which is the number of targets detected out of number of targets tested; (2) probability of false alarm which is the number of confusers detected out of the number of confusers tested and (3) probability of correct classification which is the number of targets correctly classified out of the number of targets tested. The threshold can be set to achieve a given trade-off between these measures.

The test and training sets followed the recommendations of [33] with three "known" target types and two "confuser" targets chosen for their similarity to the known targets. The confuser vehicles were T62 and BTR60.

Fig. 10 shows the probability of detection as a function of the probability of false alarm controlled by a particular threshold for each of the two vehicle types. For both T72 and BMP2, probability of detection approaches 100% for a probability of false alarm of 50%. This result can be compared with results based on the same task which are reported in the literature. Using approaches based on SAR images and with a probability of false alarm of 50%, approximately 70% and 80% targets are detected in [32] and [42] respectively.

Fig. 11(a) and (b) plot the probability of correct classification as a function of the probability of false alarm for the two vehicle types. At a probability of false alarm of 50%, the overall probability of correct classification is around 75% with the BMP vehicles achieving slightly better performance. These values are higher than those in [32] and comparable with those in [42]. However, their technique is based on complete SAR images while our technique uses a small number of HRR profiles for each feature vector.

VI. SUMMARY AND CONCLUSIONS

This paper has presented a new feature extraction technique for automatic target recognition based on a sequence of HRR profiles, which achieves excellent performance on the MSTAR databases. The technique characterizes the target scattering centers in radar returns. First, a maximum likelihood adaptive mask is used to exclude the spurious scattering centers arising from the clutter adjacent to the target. Target scattering centers are then determined using the super-resolution MUSIC algorithm and the feature vector is constructed from the scattering centers by converting the continuous displacements to discrete values, taking the 2-dimensional DCT of the target image and retaining only low-order coefficients. Although in this work, we base recognition only on the extracted target scattering centers, we note that for isolated targets on level ground, the error rate can be reduced still further by the inclusion of features that characterize the shape of the radar shadow that is visible above the target in Fig. 1(a) [8].

The method was tested for various different automatic target recognition configurations. In the closed-set experiments, a test set misclassification rate of 2.8% is achieved on a 10-target identification task. We assessed how robust the method is to the choice of feature vector size and showed how the performance is affected by the number of HMM states and the number of Gaussian mixture components. The robustness to target variants was also tested. Unseen variants were misclassified 6.7% of the time where we found that most of the errors arose from three vehicles whose configuration was significantly different to the seen variants (e.g. extra fuel tank and antenna position). In the open-set target identification
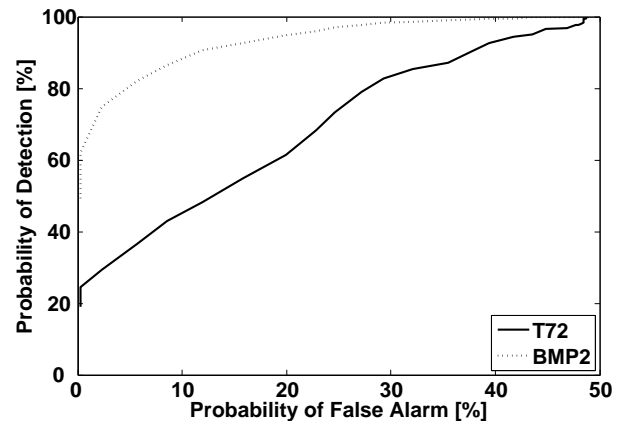

Fig. 10. The relationship between the probability of detection and the probability of false alarm for the T72 and BMP2 targets.
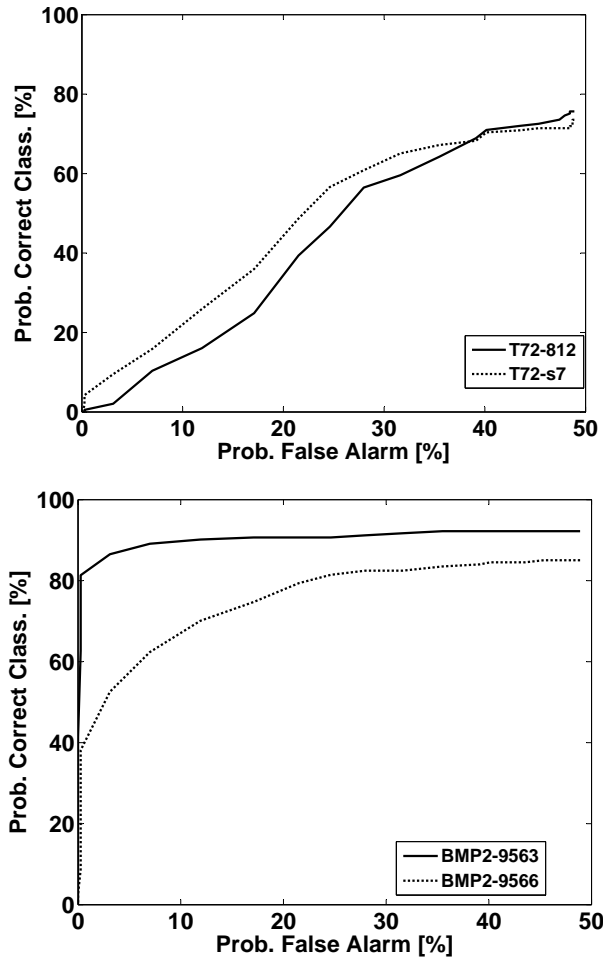
Fig. 11. The ROC curve for (a) the T72 target and (b) the BMP target

experiments using three targets and two confusers, all the targets are detected and approximately 75% are recognized correctly with a probability of false alarm of 50%. An important design decision is the choice of aspect aperture for each feature vector, P. The experiments showed that for stationary targets, the performance is only slightly decreased as the aspect aperture is reduced. These results are very encouraging for recognition of moving targets where using fewer HRR profiles is expected to improve performance.

ACKNOWLEDGEMENT

REFERENCES

[1] B. Bhanu and G. Jones, III. Object recognition results using MSTAR synthetic aperture radar data. In *Proc. IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pages 55–62, 2000.

[2] W. G. Carrara, R. S. Goodman, and R. M. Majewski. *Spotlight synthetic aperture radar, signal processing algorithms*. Artech House Inc., 1995.

[3] R. Carriere and R. L. Moses. High resolution radar target modeling using a modified Prony estimator. *IEEE Trans. on Antennas and Propagation*, 40(1):13–18, 1992.

[4] M. Cetin, W. C. Karl, and D. A. Castanon. Feature enhancement and ATR performance using nonquadratic optimization-based SAR imaging. *IEEE Trans. on Aerospace and Electronic Systems*, 39(4):1375–1395, 2003.

[5] G. Cheng, W. Zhao, J. Zhang, and S. Mao. A Practical Kernel Criterion for Feature Extraction and Recognition of MSTAR SAR Images. In *Intl Conf on Signal Processing*, volume 4, 2006.

[6] H. Chiang, R. L. Moses, and L. C. Potter. Model-based classification of radar images. *IEEE Trans. on Infromation Theory*, 46(5):1842–1854, 2000.

[7] J. Cui, J. Gudnason, and M. Brookes. Automatic Recognition of MSTAR Targets using Radar Shadow and Superresolution Features. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pages 589–592, March 2005.

[8] J. Cui, J. Gudnason, and M. Brookes. Radar Shadow and Superresolution Features for Automatic Recognition of MSTAR targets. In *Proc. IEEE Int. Radar Conference*, pages 534–539, May 2005.

[9] Stuart R. DeGraaf. SAR Imaging via Modern 2-D Spectral Estimation methods. *IEEE Trans Image Proc*, 7(5):729–761, May 1998.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, 39(1):1–38, 1977.

[11] M. D. DeVore, A.D. Lanterman, and J.A. O'Sullivan. ATR Performance of a Rician Model for SAR Images. *Proc. SPIE - The International Society for Optical Engineering*, 4050:34–45, 2000.

[12] M. D. DeVore and J. A. O'Sullivan. Quantitative statistical assessment of conditional models for synthetic aperture radar. *IEEE Trans. on Image Processing*, 13(2):113–125, 2004.

[13] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, 2 edition, 2001.

[14] D. F. Fuller, A. J. Terzuoli, P. J. Collins, and R. Williams. 1-D feature extraction using a dispersive scattering center parametric model. In *IEEE International Symposium Antennas and Propagation Society*, volume 2, pages 1296–1299, 1998.

[15] M. M. Giray and S. R. Mishra. A hybrid approach for 2-D RCS imaging. In *IEEE International Symposium Antennas and Propagation Society*, pages 1126–1129, 1992.

[16] M. P. Hurst and R. Mittra. Scattering centre analysis via Prony's method. *IEEE Trans. on Antennas and Propagation*, 35:986–988, 1987.

[17] G. Jones and B. Bhanu. Quasi-invariants for recognition of articulated and non-standard objects in SAR images. In *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pages 88–97, 1999.

[18] S Kay. *Modern spectral estimation*. Prentice hall, 1 edition, 1988.

[19] J. B. Keller. Geometrical theory of diffraction. *Opt. Soc. Amer.*, pages 116–130, 1962.

[20] D. P. Kottke, Fwu Jong-Kae, and K. Brown. Hidden Markov modeling for automatic target recognition. In *Proc. IEEE Thirty-First Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 859–863, 1997.

[21] J. Li and P. Stoica. Efficient mixed-spectrum estimation with applications to target feature extraction. *IEEE Trans. on Signal Processing*, 44(2):281–295, 1996.

[22] X. Liao, P. Runkle, and L. Carin. Identification of ground targets from sequential high-range-resolution radar signatures. *IEEE Trans. on Aerospace and Electronic Systems*, 38(4):1230–1242, October 2002.

[23] D. E. Nelson, J. A. Starzyk, and D. D. Ensley. Iterated wavelet transformation and signal discrimination for HRR radar target recognition. *IEEE Trans. on Systems, Man and Cybernetics*, 33(1):52–57, 2002.

[24] C. Nilubol and R. M. Mersereau. Improved hidden Markov model classifier for SAR images. *Proc. SPIE - The International Society for Optical Engineering*, 3720:113–122, 1999.

[25] L. M. Novak, G. J. Owirka, and W. S. Brower. Performance of 10- and 20-target MSE classifiers. *IEEE Trans. on Aerospace and Electronic Systems*, 36(4):1279–1289, 2000.

[26] L. M. Novak, G. J. Owirka, and A. L. Weaver. Automatic target recognition using enhanced resolution SAR data. *IEEE Trans. on Aerospace and Electronic Systems*, 35(1):157–175, 1999.

[27] Leslie M. Novak, Gregory J. Owirka, William S. Brower, and Alison L. Weaver. The Automatic Target-Recognition System in SAIP. *The Lincoln Lab Journal*, 10(2):187–202, 1997.

[28] J. W. Odendaal, E. Barnard, and C. W. I. Pistorius. Two-dimensional superresolution radar imaging using the MUSIC algorithm. *IEEE Trans. on Antennas and Propagation*, 42(10):1386–1391, 1994.

[29] J. A. O'Sullivan, M. D. DeVore, V. Kedia, and M. I. Miller. SAR ATR performance using a conditionally Gaussian model. *IEEE Trans. on Aerospace and Electronic Systems*, 37(1):91–108, 2001.

[30] R. P. Perry, R. C. DiPietro, and R. L. Fante. SAR imaging of moving targets. *IEEE Trans. on Aerospace and Electronic Systems*, 35(1):188–200, 1999.

[31] L. C. Potter, Chiang Da-Ming, R. Carriere, and M. J. Gerry. A GTD-based parametric model for radar scattering. *IEEE Trans. on Antennas and Propagation*, 43(10):1058–1067, 1995.

[32] T. D. Ross and J. C. Mossing. The MSTAR Evaluation Methodology. In *Proc SPIE: Algorithms for Synthetic Aperture Radar Imagery VI*, volume 3721, pages 705–713, August 1999.

[33] T. D. Ross, S. W. Worrell, V. J. Velten, J. C. Mossing, and M. L. Bryant. Standard SAR ATR evaluation experiments using the MSTAR public release data set. In *Proc SPIE: Algorithms for Synthetic Aperture Radar Imagery V*, volume 3370, pages 566–573, April 1998.

[34] P.R. Runkle, P.K. Bharadwaj, L. Couchman, and L. Carin. Hidden Markov models for multiaspect target classification. *IEEE Trans. on Signal Processing*, 47(7):2035–2040, July 1999.

[35] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. on Antennas and Propagation*, 34(3):276–280, 1986.

[36] P.B. Silverstein, O.S. Sands, and F.D. Garber. Radar target classification and interpretation by means of structural descriptions of backscatter signals. In *Proc IEEE Radar Conference*, pages 21–25, 12-13 March 1991.

[37] V. J. Velten, T. D. Ross, and J. C. Mossing S. W. Worrell M. L. Bryant. Standard SAR ATR Evaluation Experiments using the MSTAR Public Release Data Set. Technical Report ASC-98-0101, Air Force Research Lab, 1998. Includes Appendix.

[38] R. L. Williams. Analysis of a 1-D HRR moving target ATR. *Proc. SPIE - The International Society for Optical Engineering*, 1999.

[39] Yinan Yang, Yuxia Qiu, and Chao Lu. Automatic Target Classification &Experiments on the MSTAR SAR Images. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS International Workshop on Self-Assembling Wireless Networks. SNPD/SAWN 2005. Sixth International Conference on*, pages 2–7, 23-25 May 2005.

[40] S. Young, G Evermann, T. Hain, Kershaw D., G. More, J. O. D. Odell, D. V. V. Provey, and P. Woodland. *The HTK book*. Cambridge University Engineering Dept, 2 edition, 2002.

[41] Y. Zeng and J. A. Starzyk. Piecewise Linear Approach: a New Approach in Automatic Target Recognition. In *Proc. SPIE - The International Society for Optical Engineering*, volume 4050, pages 456–464, 2000.

[42] Q. Zhao and J. C. Principe. Support vector machines for SAR automatic target recognition. *IEEE Trans. on Aerospace and Electronic Systems*, 37(2):643–654, 2001.