Technical Report for AFRL and MRC

**Identification of Ground Targets From
Sequential High-Range-Resolution Radar Signatures**

Xuejun Liao, Paul Runkle and Lawrence Carin
Department of Electrical and Computer Engineering
Duke University
Durham, NC 27708-0291

Abstract - An approach to identifying ground targets from sequential high-range-resolution (HRR) radar signatures is presented. In particular, a hidden Markov model (HMM) is employed to characterize the sequential information contained in multi-aspect HRR target signatures. Features from each of the HRR waveforms are extracted via the RELAX algorithm. The statistical models used for the HMM states are formulated for application to RELAX features, and the expectation-maximization (EM) training algorithm is augmented appropriately. Example classification results are presented for the ten-target MSTAR data set.

I. INTRODUCTION

There are many applications for which one is interested in performing target detection and classification based on a sequence of high-range resolution (HRR) radar signals [1-7]. In such scenarios one implicitly views the target from a sequence of observations, due to target and/or sensor motion. One approach to processing such a sequence of waveforms is to form a SAR or ISAR image [5,6,8], with subsequent detection/classification performed in the image domain. In this paper we consider an alternative approach, in which one processes the sequence of HRR waveforms directly, without explicitly forming an image. In particular, the sequence of waveforms is processed using a hidden Markov model (HMM), HMMs having been utilized previously for processing data from several sequential-scattering problems [9-11]. The principal contributions of this paper involve development of techniques for characterization of the HMM states, for the case of HRR data. Moreover, algorithm performance is examined using the MSTAR data set [5]. Below we review the

HMM paradigm, as applied to multi-aspect scattering data, and discuss issues associated with its application to HRR data.

High-range resolution scattering from complex targets yields target signatures that are a strong function of the target-sensor orientation [1-7]. Nevertheless, there are generally contiguous sets of orientations (aspects) for which the scattering physics varies slowly, and for which the associated HRR waveforms can be approximated as stationary statistically. Each such set of angles is termed a “state” [9-11]. When one measures multiple HRR waveforms from a sequence of target-sensor orientations, one implicitly samples waveforms from a sequence of target states. Since the targets of interest are distant and possibly concealed, both the target identity and pose are unknown, and therefore the particular set of states that are sampled is deemed “hidden”. In many applications the sequence of sampled states can be characterized via a Markov process, the overall construct therefore yielding a *hidden* Markov model (HMM). This basic paradigm has been applied to several scattering problems [9-11], with characterization of the state-dependent statistics constituting the most problem-dependent HMM component. In this paper, we extend characterization of the state-dependent statistics to the case of HRR data.

There has been much previous work on characterizing HRR statistics for individual waveforms [1-4,7], with this extended here to the multi-aspect case. Some authors have developed models for the statistics of the HRR range bins, without performing feature extraction [4,12]. In this approach one typically must make assumptions concerning the range-bin-dependent statistics. For example, the constrained quadratic classifier [4,12] is based on an uncorrelated-Gaussian model for the range-bin amplitudes. Alternatively, one can perform feature extraction, with this often manifesting dimensionality reduction. Recent HRR studies have considered “peak-amplitude features” [4]. Other example feature extractors include the RELAX algorithm [13], from which waveform constituents are extracted, these characteristic of the principal target scattering centers. In this paper we consider RELAX feature extraction, and develop two distinct statistical models. The key novelty of these statistical models, *vis-à-vis* previous studies [1-4,7], is that the aforementioned HRR models are used as components in an HMM. In particular, a distinct statistical model is used for each state of the HMM, recalling that an HMM state is characterized by a set of target-sensor orientations over which the associated scattered waveforms are approximately stationary. The statistical relationships between

a *sequence* of HRR waveforms are accounted for via the HMM probabilities of transitioning from one state to the next, for a given angular sampling rate [9-11].

The remainder of the text is organized as follows. In Sec. II we briefly review the HMM construct, followed by a detailed discussion on how the state-dependent HRR statistics are characterized and employed in the context of the HMM states. The MSTAR data set is reviewed in Sec. III, along with a discussion of how such is used to generate a sequence of HRR waveforms. Several sets of example results are presented in Sec. IV, with conclusions summarized in Sec. V.

II. HMM PROCESSING OF SEQUENTIAL HRR DATA

A. HMM basics

Figure 1 represents a typical configuration for airborne interrogation of a ground target. The airborne radar periodically transmits coherent pulses of microwave energy, which impinge the ground target at depression angle ϕ . Each pulse is subsequently reflected from the target and received by the radar. After some pre-processing of these scattered radar echo pulses, a sequence of HRR signatures is obtained, each representative of the target as viewed from a distinct target-sensor orientation (with particular target-sensor orientations dependent on the sensor motion and target pose, with the latter typically unknown). For simplicity, in the studies presented here the depression angle is assumed constant, and therefore the variable target-sensor orientations are modeled as a change in the azimuthal orientation. The depression angle of the incident wave can be maintained as constant by controlling the sensor flight path, although in practice this constant-depression-angle assumption may have to be relaxed.

It is well known that HRR signatures exhibit significant variability as viewed from different orientations [1-4,7]. Nevertheless, there are typically contiguous angular sectors over which the scattered fields are approximately stationary statistically. As discussed in the Introduction, each such sector is termed a “state”, and the sequence of HRR measurements sample waveforms from a sequence of target states, with this sequence well characterized via a Markov model [9-11]. In practice the target

orientation is unknown (or “hidden”), in addition to the target identity, and therefore the sequence of scattered waveforms is characterized by a *hidden* Markov model (HMM).

In the classification algorithm an HMM is designed for each target of interest. A given set of sequential HRR data under test is submitted to all HMMs, and the data is associated with that target for which the respective HMM yields the largest likelihood. If the likelihoods from all HMM classifiers are below a prescribed threshold, then the sequential data can be declared not representative of any of the targets seen while training.

Assume a target is partitioned into N distinct states, denoted by the set $S=\{s_1, s_2, \dots, s_N\}$. As discussed above, here the states constitute an azimuthal partitioning. The HMM state-transition probabilities are denoted by the matrix $A=\{a_{ij}\}$, where a_{ij} is the probability of transiting on consecutive measurements from state i to state j . Further, the initial-state probabilities are denoted by the vector $\pi=\{\pi_i\}$, where π_i is the probability of sampling state i on the first measurement (of the sequence). For the sequence of HRR measurements, assume that $\delta\theta$ represents the change in the target-sensor orientation, upon consecutive measurements. Let θ_i represent the angular extent of state i . We assume $\delta\theta < \theta_i$, for all i , implying that on consecutive measurements one can stay in the same state or transition into an adjacent state. This yields a tri-diagonal state-transition matrix A . Based on the aforementioned assumptions with regard to $\delta\theta$ and θ_i , one can readily demonstrate the following estimations for the state-transition probabilities:

$$a_{i,i-1} = a_{i,i+1} = \delta\theta/(2\theta_i) \quad a_{i,i} = (\theta_i - \delta\theta)/\theta_i \quad (1)$$

Moreover, if the initial target pose is uniformly distributed azimuthally, we similarly have

$$\pi_i = \theta_i / \sum_{i=1}^N \theta_i \quad (2)$$

As discussed further below, (1) and (2) constitute initial estimates for A and π , with these refined via Baum-Welch training [14] (see Sec. IIC).

B. State-dependent statistics: RELAX

Consider the frequency-dependent, complex HRR waveform $\hat{x}(\omega)$, scattered from a given target at a particular (generally unknown) target-sensor orientation. The RELAX algorithm [13] is used to extract point-scattered wavefronts from $\hat{x}(\omega)$, from which the range (time) dependent HRR waveform is realized

$$x(r) = \sum_{k=1}^K A_k^x w(r - r_k^x) \quad (3)$$

where $x(r)$ is the Fourier transform of an approximation to $\hat{x}(\omega)$, with $\hat{x}(\omega)$ approximated from the K wavefronts extracted from RELAX. Each RELAX-extracted wavefront is representative of a point scatterer, with $A_k^x = \sigma_k^x \exp(j\phi_k^x)$ constituting the complex amplitude of the k th wavefront (magnitude σ_k^x), and $w(r)$ is the real window function used to represent the spatial support of the wavefronts (defined by system bandwidth). In the work presented here $w(r)$ is a Gaussian with variance consistent with sensor bandwidth. For each complex HRR waveform (3) we define the real and positive function

$$\tilde{x}(r) = \sum_{k=1}^K \sigma_k^x w(r - r_k^x) \quad (4)$$

Now consider two HRR signatures $\tilde{x}(r)$ and $\tilde{y}(r)$, for which we wish a distance measure. The distance is defined in terms of the correlation

$$\begin{aligned} C(x, y) &= \int dr \tilde{x}(r) \tilde{y}(r) = \sum_{k=1}^K \sum_{i=1}^K \sigma_k^x \sigma_i^y \int w(r - r_k^x) w(r - r_i^y) dr \\ &= \sum_{k=1}^K \sum_{i=1}^K \sigma_k^x \sigma_i^y g(r_k^x - r_i^y) \end{aligned} \quad (5)$$

Since $w(r)$ is Gaussian, so is $g(r)$, with twice the variance. Using matrix notation, (5) can be re-written as

$$C(x, y) = \sigma_x^T \mathbf{W}_{xy} \sigma_y \quad (6)$$

with $\mathbf{W}_{xy}$ a $K \times K$ matrix representing the functions $g(r_k^x - r_i^y)$. Superscript T represents transpose of a column vector. Note that the correlation in (5) can be computed directly from the RELAX outputs, without having to explicitly form the range profile (4).

We now define the distance

$$d^2(x, y) = \sigma_x^T \mathbf{W}_{xx} \sigma_x + \sigma_y^T \mathbf{W}_{yy} \sigma_y - 2\sigma_x^T \mathbf{W}_{xy} \sigma_y \quad (7)$$

which is referred to as the spatially weighted distance (SWD), in which each element in the weighting matrix is a function of spatial distance between scatterers. In (7) $\mathbf{W}_{xx}$ and $\mathbf{W}_{yy}$ are of the same form as $\mathbf{W}_{xy}$. Note that if the wavefronts are spaced uniformly, such that $\Delta = r_{k+1}^x - r_k^x = r_{i+1}^y - r_i^y$ and $g(\Delta) = 0$, then this distance measure reduces to the traditional Euclidean distance

$$d^2(x, y) = \sigma_x^T \sigma_x + \sigma_y^T \sigma_y - 2\sigma_x^T \sigma_y = (\sigma_x - \sigma_y)^T (\sigma_x - \sigma_y) \quad (8)$$

The distance measure in (7) plays a pivotal role in the HRR HMM, as currently designed. In particular, for the N states $S = \{s_1, s_2, \dots, s_N\}$ characteristic of a given target, we define a codebook of N codes $C = \{c_1, c_2, \dots, c_N\}$, with code c_n associated with state s_n . Recall that the HMM requires a state-dependent statistical measure. In the context of RELAX feature extraction, we have implemented this in two ways.

Consider a sequence of M HRR waveforms, representative of viewing the target from a sequence of M orientations, with each waveform parsed into a set of K point-scatter features via RELAX. In the context of a discrete HMM [11,12], each set of RELAX features is mapped into a distinct codebook element (a code from C), using a nearest-neighbor mapping implemented via the distance measure in (7). In this context each HRR waveform in the sequence of M is mapped to a distinct code (analogous to vector quantization [15]), and therefore after quantization the original M HRR waveforms are represented by a sequence of M codes (each a member of C). The HMM is used to define the target for which such a sequence of codes is most likely. The state-dependent probability of observing a particular code is characterized via a matrix $\mathbf{B} = \{b_{ij}\}$ where b_{ij} represents the probability of observing code i in state j .

While a discrete HMM is relatively simple, being characterized by the set of codes C , state-transition matrix $\mathbf{A}$, initial-state probabilities π , and the state-dependent probability matrix $\mathbf{B}$, such a model has well-known limitations. In particular, there is distortion [15] when one performs the aforementioned vector quantization, this undermining performance. Moreover, as the number of targets of interests increases, the number of codes also increases, making $\mathbf{B}$ a large matrix (this issue is discussed further below).

To mitigate these problems, we have also considered a continuous HMM [12]. We again define a codebook $C=\{c_1, c_2, \dots, c_N\}$, with code c_n associated with state s_n . For a given HRR signal x the likelihood that x is associated with state s_n is defined as

$$p(x|s_n) = \frac{1}{\sqrt{2\pi\eta_n}} \exp\left[-\frac{d^2(x, c_n)}{2\eta_n^2}\right] \quad (9)$$

The likelihood in (9) is clearly motivated by a Gaussian distribution, but here we explicitly utilize the distance measure d^2 , which is tied to the RELAX features extracted from the HRR waveform. Therefore, the continuous HMM is of the same form as its discrete counterpart, with the $\mathbf{B}$ matrix in the latter replaced in the former by $p(x|s_n)$.

It is worthwhile to emphasize an important distinction between the continuous and discrete HMMs. As implemented here both methods employ a codebook C , but they utilize such in distinct ways. Considering first the discrete HMM, the HMM matrix $\mathbf{B}$ associated with a given target must include the codes across *all* targets of interest. If, by contrast, the $\mathbf{B}$ matrix associated with a given target T_i were only to use codes associated with T_i , then the quantization procedure for the associated HMM would map *any* HRR profile into a code characteristic of T_i . Therefore, in $\mathbf{B}$ we utilize the complete set of codes for all targets of interest, accounting for target-dependent code diversity. Consequently the size of $\mathbf{B}$ grows with the number of targets of interest. On the other hand, the density function $p(x|s_n)$ utilized in the continuous HMM accounts for waveforms that are not matched to a given target-dependent codebook by yielding a low likelihood (ideally, if x is not associated with state s_n , $p(x|s_n)$ will be small). Hence, in defining $p(x|s_n)$ for a given continuous HMM, one only employs codes from the associated target, and therefore the continuous-HMM complexity does not increase with the number of targets.

Concerning the design of the state decomposition $S=\{s_1, s_2, \dots, s_N\}$ and the associated set of codes $C=\{c_1, c_2, \dots, c_N\}$, we initially divide the scattering data into N states, using a uniform angular decomposition (for N states, each has angular extent $360^\circ/N$). Using the training data associated with target T_i we perform RELAX on all HRR waveforms, and the code c_n is the average of the RELAX

features from state s_n . The above state decomposition and codes only constitute an initialization, with this optimized subsequently via the expectation-maximization (EM) algorithm discussed in Sec. IIC.

C. Expectation-maximization HMM optimization

The discussions in Sec. IIB employed state-dependent statistical models based on an assumed state decomposition. Moreover, various averaging procedures have been employed to compute such HMM components as codebooks (discrete HMM) and density-function parameters (continuous HMM), once the state decomposition is defined. Clearly the quality of the associated HMM is dictated by the quality of such model-parameter estimations. We have therefore employed optimization algorithms to refine the HMM parameters, and therefore the model-parameter computations discussed in Sec. IIB are simply initializations for the model parameters, followed by the optimization procedure discussed below.

The expectation-maximization (EM) HMM training algorithm starts by defining the forward variable

$$\alpha_m(n) = p(o_1 o_2 \cdots o_m, q_m = s_n / \lambda) \quad (10)$$

and backward variable

$$\beta_m(n) = p(o_{m+1} o_{m+2} \cdots o_M / q_m = s_n, \lambda) \quad (11)$$

where $O = \{o_1, o_2, \dots, o_M\}$ is representative of the sequence of M observed HRR waveforms (more specifically, the components of the M HRR profiles used in the HMM model), q_m is the state sampled by the m th measurement (this “hidden”), s_n represents the n th state of a given target (with $n=1, 2, \dots, N$), and λ represents the associated HMM parameters. We define two intermediate variables

$$\begin{aligned} \xi_m(i, j) &= p(q_m = s_i, q_{m+1} = s_j | O, \lambda) \\ &= \frac{p(O, q_m = s_i, q_{m+1} = s_j / \lambda)}{p(O / \lambda)} \\ &= \frac{\alpha_m(i) a_{ij} b_j(o_{m+1}) \beta_{m+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_m(i) a_{ij} b_j(o_{m+1}) \beta_{m+1}(j)} \end{aligned} \quad (12)$$

and

$$\begin{aligned}
\gamma_m(i) &= p(q_m = s_i | O, \lambda) \\
&= \frac{p(O, q_m = s_i | \lambda)}{p(O | \lambda)} \\
&= \frac{\alpha_m(i)\beta_m(i)}{\sum_{i=1}^N \alpha_m(i)\beta_m(i)} = \sum_{j=1}^N \xi_m(i, j)
\end{aligned} \tag{13}$$

In (12) the expression $b_j(o_{m+1})$ represents the likelihood of observing o_{m+1} in state j . For the discrete HMM, o_{m+1} corresponds to a discrete code (after quantization), while for a continuous HMM it corresponds to RELAX features (Sec. IIB). In particular, for the continuous HMM employing RELAX features $b_j(o_{m+1})$ is represented by (9).

Given the observation sequence $O = \{o_1, o_2, \dots, o_M\}$ and the current HMM parameters λ , the Baum-Welch algorithm re-estimates the HMM parameters as

$$\hat{\pi}_i = \gamma_1(i) \tag{14}$$

$$\hat{a}_{ij} = \frac{\sum_{m=1}^{M-1} \xi_m(i, j)}{\sum_{m=1}^{M-1} \sum_{j=1}^N \xi_m(i, j)} \tag{15}$$

and the code associated with state n is approximated as

$$\hat{c}_n = \frac{\sum_{m=1}^M \gamma_m(n) \cdot o_m}{\sum_{m=1}^M \gamma_m(n)} \tag{16}$$

For the case of the continuous HMM, based on RELAX features, the EM update procedure is more involved. The associated details are summarized in Appendix A.

The EM algorithm iterates by using the new HMM parameters λ reflected in (14)-(16) to recompute the parameters in (10)-(13), with this process repeated until λ converges. Upon convergence the match between the training data (represented by the training observation sequences O) and HMM models is improved, generally increasing the likelihood of the training data when applied to the final HMM (relative to the HMM using the original model parameters). However, we note that the

EM algorithm does not in general yield the global optimal ML solution, and therefore a good initialization is important. We here utilize the initialization discussed in Secs. IIA and IIB.

A final point is in order concerning the HMM training. The EM algorithm is designed to maximize the likelihood of a given set of training data, as applied to the associated HMM statistical model. When training a given HMM, we therefore utilize no additional information concerning other targets that may be encountered. If desired, one could design the HMM for target T_i such that the likelihood of the training data is maximized when submitted to the associated model, while simultaneously minimizing the likelihood for training data associated with all other known targets $T_j \neq T_i$. This approach may yield improved classification performance, but it assumes knowledge of other targets that may be encountered. We have not utilized this HMM-design approach, although in some applications it may be appropriate.

III. DATA SET UNDER CONSIDERATION

We consider the MSTAR data set for testing the algorithms discussed in Sec. II. As elucidated in other papers [5,6], the MSTAR data consists of X-band scattering from ten targets. The data typically consists of image chips. These images have been converted into a sequence of HRR waveforms, through several filtering operations. The HRR data was provided to the investigators by the US Defense Advanced Research Projects Agency (DARPA), with the conversion of MSTAR images to HRR waveforms performed under the DARPA TRUMPETS program. All ground-based targets are of very similar shape (most are military vehicles, although there are a few civilian vehicles [5,6]).

We provide a brief summary on how MSTAR images are converted to aspect-dependent HRR signatures. Consider an MSTAR image $I(z,y)$, where z reflects the downrange dimension and y the cross range. The image consists of bipolar real pixel values. A fast Fourier transform (FFT) is performed in the y dimension, for all z . After deconvolving the cross-range window function, each z -dependent waveform, for a fixed y , corresponds to an HRR profile. As we view the image $I(z,y)$ for different rotational positions of the target, we extract HRR profiles from numerous target-sensor

orientations. The multiple aspect-dependent HRR profiles are then segmented into contiguous 1° azimuthal bins. Within each such bin, 75% of the HRR profiles are randomly selected as training data, with the remaining 25% used for testing. After this process is completed, the profiles are distributed such that one can achieve a 0.1° angular sampling within the context of the M -dimensional HRR sequence ($\delta\theta = 0.1^\circ$). At this azimuthal sampling each target is characterized by 3,601 HRR waveforms over the full 360° aspect range, resulting in a cumulative total of 36,010 training signatures and 36,010 testing signatures for the ten targets (the testing and training data is distinct, as defined by the aforementioned random partitioning). The HRR signatures have a bandwidth of 600 MHz centered at 10 GHz, and have a range resolution of approximately 0.3 meters. The depression angle (relative to grazing) is 15° for all examples considered. The above segmentation of testing and training data was specified by DARPA under the TRUMPETS program.

IV. EXAMPLE RESULTS

A. Preliminaries

In the first set of example results, we consider a 3° angular aperture for the sequential HRR data under test, which at the angular sampling rate of $\delta\theta = 0.1^\circ$ corresponds to a sequence of $M=31$ scattered waveforms. Each of the ten MSTAR targets under consideration is partitioned into 120 states, over a 360° azimuthal angular range, corresponding to an initialization of 3° state sizes (this is reapportioned in the HMM training, so that the state sizes subsequently adjust to account for the scattering physics). The large number of states is reflective of the target complexity (as described further below) and the small operating wavelength (3 cm) relative to characteristic target dimensions.

In Sec. IIIA we discussed a discrete HMM, in which the features associated with each HRR waveform are mapped to a code, and therefore the sequence of M HRR profiles is transformed to a sequence of M codes. We have implemented such an HMM in the context of RELAX feature parsing (the codes represent a discretization of the RELAX feature space). We have found that, for the MSTAR data set, the discrete HMM yields performance worse than that of a continuous HMM, and

therefore all results reported here are for continuous HMMs. The discrete HMM is also undesirable due to the very large $\mathbf{B}$ matrix characteristic of the ten targets under test (see Sec. IIB).

One of the significant complicating issues in multi-aspect HRR scattering is the scintillation caused by relatively small scattering centers that come in and out of view with relatively small changes in aspect (target-sensor orientation). By limiting RELAX wavefront extraction to the relatively large scatterers, we mitigate scintillation to some extent. In the course of the work reported here, we have extracted the $K=15$ principal wavefronts (scattering centers) from each of the $M=31$ scattered waveforms in the 3° angular sequence. For these studies the parameter K has not been optimized. Such optimization is a subject of significant interest, in that K should be small enough to avoid relatively small scattering centers that may lead to scintillation, while being large enough to robustly extract the principal scattering centers and retain sufficient information to aid classification performance.

B. HMM-parameter convergence

Prior to showing classification performance on the MSTAR data, it is of interest to examine convergence of the HMM parameters, in the course of the EM-algorithm training [16]. We consider a sequence of 200 HRR profiles, over a 20° angular extent (recall the 0.1° degree sampling). We plot in Fig. 2 the likelihood that each of the 200 profiles is associated with a given HMM state, with the states considered encompassing the range of angles under consideration (here we consider five states). To be specific, for each iteration of the EM algorithm, we compute the likelihood that the m th measurement in the sequence (the m th HRR waveform) is associated with state s_i . This is quantified using $\gamma_m(i)$ from equation (13). Note that initially the states are defined by a uniform decomposition in angle, and each is 3° in extent (the initialization described in Sec. II). As the EM algorithm iterates, the state partitioning is adjusted, such that the states are no longer necessarily of the same angular extent. Moreover, upon convergence, the state decomposition and the associated parameters are such that each HRR profile is more clearly associated with a particular state. We see from Fig. 5 that the EM algorithm converges after five iterations and that the final state decomposition is different from the initialization. We reiterate that the state decomposition is dictated by the scattering physics, with each state representative of approximately stationary scattering (as a function of target-sensor orientation).

C. HMM classification

In Table 1 we present confusion-matrix results for the ten-target problem, considering all 3° sequences for all targets. Also shown in Table 1 are photographs of the ten (similar) targets under consideration. For the 3° angular data considered, we find that the RELAX-based HMM yields an average classification rate of 82%. In the RELAX algorithm there is no waveform averaging, with scintillation addressed by limiting the number of scattering centers extracted (here $K=15$ in (3)). We have also tried using other values of K , but for the data considered the results are similar. In fact, when K increases to $K=20$, performance actually declines slightly, presumably due to enhanced scintillation (inclusion of five additional scattering centers, the characteristics of which change quickly with variable target-sensor orientation).

Note that with 120 states per target, each state initialized at 3° in extent (with this adjusted during training, see Sec. IVB), a test sequence of 3° extent is likely to only realize a single state transition. Multiple targets may have similar states, but the character of multiple adjacent states is likely to be more target dependent. Therefore, by increasing the angular extent of the test sequence, the additional state transitions are likely to improve classification. In Table 2 we reconsider the RELAX-based HMM, but now consider sequences of 6° angular extent ($M=61$, with 0.1° angular sampling). We see in Fig. 5 that classification performance now improves considerably, to an average of 92%.

Before closing, we note that other authors have considered processing the MSTAR data, although most of these studies have been performed in the image domain [5,6]. The results reported here, in which we process a sequence of HRR profiles rather than forming an image, are comparable to the results in these other studies.

V. CONCLUSIONS

We have addressed multi-aspect target classification using a sequence of HRR waveforms. Rather than forming an image, and performing classification in the image domain, classification has been performed with the HRR waveforms directly. The sequence of waveforms are processed via a hidden Markov model (HMM), the states of which reflect a generally contiguous set of target-sensor

orientations for which the wave scattering is relatively stationary. The sequence of HRR profiles, from a given target, implicitly sample HMM states. When performing classification, the target identity and orientation (pose) are unknown, and therefore the sequence of states being sampled is “hidden”. We have characterized the statistics of the HMM states in the context of feature parsing, with such performed via RELAX [13]. A state-dependent statistical model has been devised for the variation of the RELAX features.

The multi-aspect HRR problem is highly complicated by the size of the radar wavelength relative to characteristic target dimensions. The high degree of resolution offers potential for classification, but it also leads to scattering that changes very quickly with aspect. This has led to the large number of HMM states used in the classifiers. In the results presented here there are 120 states for each target, over a 360° azimuthal extent. One can use fewer states, but classification performance deteriorates. In future research the model must be extended to perform the state decomposition over 4π steradians, with states defined by solid angles. This will lead to a very large number of states, reflective of target complexity. However, in practice one may limit the number of states under test. For example, for HRR sensing of airborne targets, one can use Doppler processing to estimate the target flight path. Such information can significantly reduce the number of states that must be tested when testing with the HMM.

APPENDIX A

The maximum-likelihood (ML) estimate of the HMM parameters λ is formulated as

$$\lambda_{ML} = \arg \max [\log P(O | \lambda)] \quad (\text{A1})$$

where recall that $O = \{o_1, o_2, \dots, o_M\}$ is the sequence of M observations. The motivation of the EM algorithm is that associated with the observation data O there can be some unobservable or hidden data such that if the hidden data were made known it would be easier to fit the observed data to the model [16]. In our case the states are hidden. The key to the EM algorithm is to devise an auxiliary function [16]

$$Q(\bar{\lambda} | \lambda) = E_{Q|O, \lambda} \log P(O, Q | \bar{\lambda}) \quad (\text{A2})$$

where Q is the hidden data, O is the observed data, λ and $\bar{\lambda}$ are the present and new parameters, respectively.

The EM algorithm consists of the following two steps [16]: step 1 (expectation) in which we compute $Q(\bar{\lambda} | \lambda^k) = E_{Q|O, \lambda^k} \log P(O, Q | \bar{\lambda})$, and step 2 (maximization) in which we find

$\lambda^{k+1} = \arg \max_{\bar{\lambda}} Q(\bar{\lambda} | \lambda^k)$. It can be shown that λ^k will increase monotonically to a maxima [16], although not necessarily to a global maxima.

We now apply the EM algorithm to the HRR-HMM and derive the extended Buam-Welch algorithm. For training $O = \{o_1, o_2, \dots, o_M\}$ there exist a set of possible states paths, with invalid paths accounted for by the paths with one or more zero probability transitions. Denote a state path as $Q = \{q_1, q_2, \dots, q_M\}$, then

$$\begin{aligned} \log P(O, Q | \bar{\lambda}) &= \log [\bar{\pi}_{q_1} P(o_1 | q_1, \bar{\lambda}) \prod_{m=1}^{M-1} \bar{a}_{q_m q_{m+1}} P(o_{m+1} | q_{m+1}, \bar{\lambda})] \\ &= \log \bar{\pi}_{q_1} + \log P(o_1 | q_1, \bar{\lambda}) + \sum_{m=1}^{M-1} [\log \bar{a}_{q_m q_{m+1}} + \log P(o_{m+1} | q_{m+1}, \bar{\lambda})] \end{aligned} \quad (A3)$$

Taking conditional expectation $E_{Q|O, \lambda}$ of (A3) and using the Markov property of Q , we have

$$\begin{aligned} Q(\bar{\lambda} | \lambda) &= E_{Q|O, \lambda} [\log \bar{\pi}_{q_1} + \log P(o_1 | q_1, \bar{\lambda})] + \sum_{m=1}^{M-1} E_{Q|O, \lambda} [\log \bar{a}_{q_m q_{m+1}} + \log P(o_{m+1} | q_{m+1}, \bar{\lambda})] \\ &= \sum_{n=1}^N P(q_1 = s_n | O, \lambda) [\log \bar{\pi}_n + \log P(o_1 | q_1 = s_n, \bar{\lambda})] \\ &\quad + \sum_{m=1}^{M-1} [\sum_{i=1}^N \sum_{j=1}^N P(q_m = s_i, q_{m+1} = s_j | O, \lambda) \log \bar{a}_{ij} + \sum_{n=1}^N P(q_{m+1} = s_n | O, \lambda) \log P(o_{m+1} | q_{m+1}, \bar{\lambda})] \end{aligned} \quad (A4)$$

which is the expectation step. Substituting (12)-(13) into (A4), we have

$$\begin{aligned} Q(\bar{\lambda} | \lambda) &= \sum_{n=1}^N \gamma_1(n) P(q_1 = s_n | O, \lambda) [\log \bar{\pi}_n + \log P(o_1 | q_1 = s_n, \bar{\lambda})] \\ &\quad + \sum_{m=1}^{M-1} [\sum_{i=1}^N \sum_{j=1}^N \xi_m(i, j) \log \bar{a}_{ij} + \sum_{n=1}^N \gamma_{m+1}(n) \log P(o_{m+1} | q_{m+1}, \bar{\lambda})] \end{aligned} \quad (A5)$$

To accomplish the maximization step, we differentiate $Q(\lambda | \bar{\lambda})$ with respect to each parameter and equate the result to zero.

For the state transition matrix $\mathbf{A}$ and initial state distribution π , we must preserve the constraints $\sum_{j=1}^N a_{ij} = 1$ and $\sum_{j=1}^N \pi_j = 1$. These constraints are implemented via Lagrange multipliers, as in a traditional HMM [14], the results from which are given in (14) and (15).

We now focus on the HMM parameters required for the RELAX-based density function in (9), this component unique to the methodology presented here. For the parameters in (9), i.e., $\lambda_G = (\sigma_{c_1}, \sigma_{c_2}, \dots, \sigma_{c_N}; \mathbf{r}_{c_1}, \mathbf{r}_{c_2}, \dots, \mathbf{r}_{c_N}; \eta_1^2, \eta_2^2, \dots, \eta_N^2) \in \bar{\lambda}$, with $(\sigma_{c_n}, \mathbf{r}_{c_n})$ the RELAX features for the code associated with state n , and η_n^2 is the associated variance, we have

$$\frac{\partial Q(\bar{\lambda} | \lambda)}{\partial \bar{\lambda}_G} = \sum_{m=1}^M \sum_{n=1}^N [\gamma_m(n) \frac{\partial \log P(o_m | q_m = s_n, \bar{\lambda})}{\partial \bar{\lambda}_G}] \quad (\text{A6})$$

From (9) we have

$$P(o_m | q_m = s_n, \bar{\lambda}) = \frac{1}{\sqrt{2\pi\eta_n^2}} \exp\left[-\frac{\sigma_{o_m}^T \mathbf{W}_{o_m o_m} \sigma_{o_m} + \bar{\sigma}_{c_n}^T \bar{\mathbf{W}}_{c_n c_n} \bar{\sigma}_{c_n} - 2\sigma_{o_m}^T \bar{\mathbf{W}}_{o_m c_n} \bar{\sigma}_{c_n}}{2\eta_n^2}\right] \quad (\text{A7})$$

Substituting (A7) into (A6) yields for $\bar{\sigma}_{c_n}$

$$\frac{\partial Q(\bar{\lambda} | \lambda)}{\partial \bar{\sigma}_{c_n}} = \frac{-1}{2\eta_n^2} \sum_{m=1}^M [\gamma_m(n) (2\bar{\mathbf{W}}_{c_n c_n} \bar{\sigma}_{c_n} - 2\bar{\mathbf{W}}_{o_m c_n}^T \sigma_{o_m})] \quad (\text{A8})$$

Equating (A8) to zero gives the re-estimation formula

$$\bar{\sigma}_{c_n} = \frac{\bar{\mathbf{W}}_{c_n c_n}^{-1} \sum_{m=1}^M [\gamma_m(n) \bar{\mathbf{W}}_{o_m c_n}^T \sigma_{o_m}]}{\sum_{m=1}^M \gamma_m(n)} \quad (\text{A9})$$

Similarly, we obtain for $\bar{\mathbf{r}}_{c_i}$

$$\begin{aligned} \frac{\partial Q(\bar{\lambda} | \lambda)}{\partial \bar{\mathbf{r}}_{c_n}} &= \sum_{m=1}^M \gamma_m(i) \bar{\sigma}_{c_n} \circ [\bar{\mathbf{W}}_{o_m c_n} (\hat{\mathbf{o}}_{o_m} \circ \bar{\mathbf{r}}_{o_m}) - \bar{\mathbf{W}}_{c_n c_n}^{diag \rightarrow 0} (\bar{\sigma}_{c_n} \circ \bar{\mathbf{r}}_{c_n})] \\ &\quad - \bar{\mathbf{r}}_{c_n} \circ [\sum_{m=1}^M \gamma_m^{(k)}(n) \bar{\sigma}_{c_n} \circ (\bar{\mathbf{W}}_{o_m c_n} \hat{\mathbf{o}}_{o_m} - \bar{\mathbf{W}}_{c_n c_n}^{diag \rightarrow 0} \hat{\mathbf{o}}_{c_n})] \end{aligned} \quad (\text{A10})$$

where $\circ$ denotes the element-wise multiplication of two vectors of the same dimension, and $\bar{\mathbf{W}}_{c_n c_n}^{diag \rightarrow 0}$ is $\bar{\mathbf{W}}_{c_n c_n}$ with the diagonal elements set to zero. Because of the nonlinear functional form of (A10) with respect to $\bar{\mathbf{r}}_{c_n}$ (note $\bar{\mathbf{r}}_{c_n}$ appears in $\bar{\mathbf{W}}_{O_m c_n}$ and $\bar{\mathbf{W}}_{c_n c_n}^{diag \rightarrow 0}$), it is not possible to derive a closed-form re-estimation formula for $\bar{\mathbf{r}}_{c_n}$ by setting (A10) to zero. We therefore use the gradient method to accomplish the maximization of $Q(\lambda | \bar{\lambda})$ with respect to $\bar{\mathbf{r}}_{c_n}$. The re-estimation of $\bar{\mathbf{r}}_{c_n}$ is accordingly performed via the following iteration: step 1, let $\bar{\mathbf{r}}_{c_n}^{(0)} = \mathbf{r}_{c_n}$ and $k = 0$; step 2,

$$\bar{\mathbf{r}}_{c_n}^{(k+1)} = \bar{\mathbf{r}}_{c_n}^{(k)} + \rho_k \left. \frac{\partial Q(\bar{\lambda} | \lambda)}{\partial \bar{\mathbf{r}}_{c_n}} \right|_{\bar{\mathbf{r}}_{c_n} = \bar{\mathbf{r}}_{c_n}^{(k)}}, \text{ step 3, if } \frac{\|\bar{\mathbf{r}}_{c_n}^{(k+1)} - \bar{\mathbf{r}}_{c_n}^{(k)}\|}{\|\bar{\mathbf{r}}_{c_n}^{(k)}\|} \text{ is small enough, stop. Otherwise, } k = k + 1,$$

and go to step 2. Note that ρ_k in Step 2 is the learning coefficient that controls the convergence speed, and ρ_k is usually chosen as an exponentially decreasing sequence.

To obtain $\bar{\eta}_n^2$, we have

$$\begin{aligned} \frac{\partial Q(\bar{\lambda} | \lambda)}{\partial \eta_n^2} &= \sum_{m=1}^M \sum_{n=1}^N \gamma_m(n) \frac{\partial}{\partial \eta_n^2} \left[\log \frac{1}{\sqrt{2\pi\mu_n^2}} - \frac{d^2(O_m, c_n)}{2\bar{\eta}_n^2} \right] \\ &= \sum_{m=1}^M \gamma_m(n) \left[-\frac{1}{2\mu_n^2} + \frac{d^2(O_m, c_n)}{2\bar{\eta}_n^4} \right] \end{aligned} \quad (\text{A11})$$

where d is the spatially weighted distance as defined in (7). Equating (A11) to zero, we obtain the re-estimation formula

$$\bar{\eta}_n^2 = \frac{\sum_{m=1}^M \gamma_m(n) d^2(O_m, c_n)}{\sum_{m=1}^M \gamma_m(n)} \quad (\text{A12})$$

REFERENCES

- [1] H. J. Li and S. H. Yang, "Using range profiles as feature vectors to identify aerospace objects", *IEEE Trans. Antennas and Propagation*, Vol.41, No.3, pp.261-268, 1993.
- [2] S. Hudson and D. Psaltis, "Correlation filters for aircraft identification from radar range profiles", *IEEE Trans. Aero. and Electronics System*, Vol. 29, No.3, pp. 741-748, 1993.
- [3] A. Zyweck and R. E. Bogner, "Radar target classification of commercial aircraft", *IEEE Trans. Aerospace and Electronics System*, Vol.32, No.2, pp.598-606, 1996.
- [4] R.A. Mitchell and J.J. Westerkamp, "Robust statistical feature based aircraft identification", *IEEE Trans. Aerospace and Electronic Systems*, Vol. 35, No.3, pp.1077-1094, 1999.
- [5] H.-C. Chiang, R. L. Moses, and L.C. Potter, "Model-based classification of radar images", *IEEE Trans. Information Theory*, Vol.46, No.5, pp. 1842-1854, 2000.
- [6] L.M Novak, "State-of-the-art of SAR automatic target recognition", The Record of the IEEE 2000 International Radar Conference, 2000, pp. 836 –843, 2000.
- [7] X. Liao, Z. Bao, and M. Xing, "On the aspect sensitivity of high resolution range profiles and its reduction methods", The Record of the IEEE 2000 International Radar Conference, 2000, pp. 310–315, 2000.
- [8] C.V. Jakowatz *et al.*, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Kluwer Academic Press.
- [9] P. Runkle, P. Bharadwaj, and L. Carin, "Hidden Markov model multi-aspect target classification," *IEEE Trans. Signal Proc.*, vol. 47, pp. 2035-2040, July 1999.
- [10] P. Runkle, L. Carin, L. Couchman, T.J. Yoder, and J.A. Bucaro, "Multi-aspect target identification with wave-based matching pursuits and continuous hidden Markov models *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, pp. 1371-1378, Dec. 1999.
- [11] P. Runkle, L. Nguyen, J. McClellan and L. Carin, "Multi-aspect target detection for SAR imagery using hidden Markov models," *IEEE Trans. Geoscience and Remote Sensing*, vol. 39, pp. 46-55, Jan. 2001.
- [12] R.A. Mitchell and R. DeWall, "Overview of high range resolution radar target identification," *Proc. Automatic Target Recognition Working Group*, Monterey, CA, 1994.
- [13] J. Li, P. Stoica, "Efficient mixed-spectrum estimation with application to target feature extraction", *IEEE Trans. Signal Processing*, Vol.44, No.2, pp.281-295, 1996.

- [14] L.R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition”, *Proc. IEEE*, Vol. 77, No. 2, pp. 257 –286, 1989 .
- [15] Y. Linde, A. Buzo and R.M. Gray, “An algorithm for vector quantizer design,” *IEEE Trans. Comm.*, vol. 28, pp. 84-95, Jan. 1980.
- [16] T.K. Moon, “The expectation-maximization algorithm,” *IEEE Sig. Proc. Mag.*, vol. 13, pp. 47-60, 1996.

Figure 1. Schematization of how the hidden Markov model (HMM) is used to process a sequence of HRR waveforms. In (a) a target is viewed from a constant depression angle and multiple azimuthal positions (b). The azimuthal variation of the scattered fields is characterized by a set of states (here three shown), over which the scattering physics is approximately stationary.

(a)

(b)

(c)

Figure 2. Convergence of the state decomposition as a function of the EM-algorithm iteration number. Results are shown for 20° degrees of sequential data, 0.1° angular sampling. The data corresponds to the ZSU target (see Table 1). (a) iteration one, (b) iteration three, (c) iteration five

		HMM models of different targets									
		T72	BTR70	BMP2	2S1	ZSU234	BTR60	BRDM2	D7	T62	ZIL131
Test data of different targets	T72	84.122	0.336	1.484	1.596	2.072	0.728	2.268	4.509	0.196	2.688
	BTR70	0.784	77.289	3.136	3.752	2.212	5.881	1.988	1.232	2.604	1.120
	BMP2	3.668	1.596	79.782	3.976	4.705	0.952	0.336	1.792	0.392	2.800
	2S1	2.100	2.436	3.360	81.994	2.632	1.736	1.456	1.232	1.680	1.372
	ZSU234	1.120	0.840	3.080	1.624	84.262	0.672	0.924	2.492	3.472	1.512
	BTR60	0.644	1.456	0.952	1.932	7.449	78.885	1.988	1.960	3.164	1.568
	BRDM2	0.196	5.013	3.752	1.596	1.624	2.408	82.890	1.064	0.000	1.456
	D7	1.932	1.568	0.924	1.008	1.512	1.232	1.260	88.715	1.428	0.420
	T62	1.484	0.756	0.364	0.952	4.088	1.288	0.168	3.640	84.290	2.968
	ZIL131	1.064	0.756	0.756	1.652	5.853	1.904	0.644	1.456	1.120	80.062
Average rate		82.229									

Table 1. Confusion matrix for the ten MSTAR targets. The classification rate (in percent) is shown using a 3° aperture and 0.1° azimuthal sampling (a sequence of 31 HRR waveforms). Photographs are shown of each of the targets.

		HMM models of different targets									
		T72	BTR70	BMP2	2S1	ZSU234	BTR60	BRDM2	D7	T62	ZIL131
Test data of different targets	T72	91.895	0.480	0.085	0.875	2.881	0.706	1.243	1.186	0.000	0.650
	BTR70	0.706	87.800	2.231	1.327	1.214	3.050	1.327	0.000	2.146	0.198
	BMP2	2.739	0.593	91.782	1.243	0.988	0.085	0.000	0.875	0.198	1.497
	2S1	0.593	1.666	0.508	93.279	1.186	0.056	1.497	0.000	0.339	0.875
	ZSU234	0.480	0.254	2.203	1.073	92.516	0.000	0.282	0.960	1.327	0.904
	BTR60	0.508	0.395	0.141	1.864	4.886	86.416	0.706	0.904	3.022	1.158
	BRDM2	0.113	2.400	1.469	1.186	0.085	0.311	93.646	0.000	0.000	0.791
	D7	0.000	0.226	0.113	0.000	0.169	0.056	1.045	97.628	0.621	0.141
	T62	0.000	0.000	0.847	0.480	2.287	0.000	0.000	0.480	95.369	0.537
	ZIL131	0.960	0.000	2.429	0.311	2.175	1.892	0.904	0.085	0.000	91.245
Average rate		92.158									

Table 2. Confusion matrix for the ten MSTAR targets. The classification rate (in percent) is shown using a 6° aperture and 0.1° azimuthal sampling (a sequence of 61 HRR waveforms).