Hidden Markov models for multi-perspective radar target recognition

Jingjing Cui, Jon Gudnason *Member, IEEE*, Mike Brookes *Member, IEEE*
Department of Electrical and Electronic Engineering
Imperial College London
{jingjing.cui02,jon.gudnason,mike.brookes}@imperial.ac.uk

Abstract—This paper presents a novel fusion technique for automatic target recognition from high range resolution RADAR profiles when observations from multiple viewpoints are available. The fusion technique entails only a straightforward modification of the transition probabilities of a single-viewpoint target model in which a Hidden Markov Model is used to represent the unknown target orientation. Evaluations using the MSTAR database indicate that the new technique can reduce classification errors by about two orders of magnitude when compared to single viewpoint observations and, in a 10-target classification experiment, gave almost perfect recognition.

Index Terms—Hidden Markov Models, Synthetic Aperture Radar, Multi-Perspective Classification.

I. INTRODUCTION

This paper develops a hidden Markov model architecture (HMM) for automatic target recognition using synthetic aperture radar (SAR) to include multi-perspective radar recordings. The architecture design takes the relationship between the radar profiles of the sensors into account and it is shown to achieve very high recognition performance using only two sensors.

Synthetic aperture radar is an important radar mode for both military, security and commercial applications. A network of radar sensors can collect backscattered radiation of targets from multiple perspectives and can potentially surpass the performance of a single-sensor radar target classifier. In this context three classification methods have investigated in [1], [2]; the Bayesian classifier, the K-nearest neighbour algorithm and the artificial neural network (ANN) were used to combine data collected from multiple perspectives and demonstrated an improvement in recognition performance. However none of these methods uses the information on the relationship among multiple perspectives.

In this paper, we show that target recognition performance can be substantially improved by fusing observations from multiple viewpoints. We suppose that SAR observations are made of a single target from multiple viewpoints either simultaneously or at different times as illustrated in Fig. 1. Our aim is to fuse these observation sequences in order to obtain improved target recognition performance. For simplicity, we assume below that there are only two viewpoints although the technique can straightforwardly be extended to an arbitrary number.
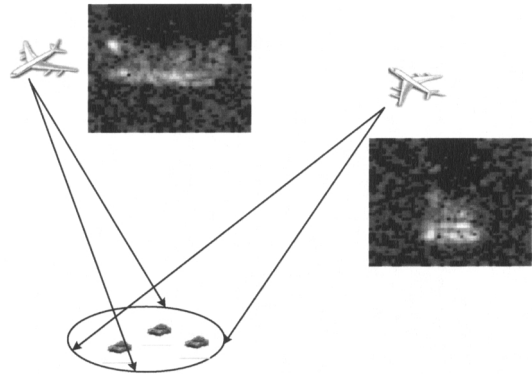
Fig. 1. An example of multisensor surveillance.

The work described in this paper is a development of the single-viewpoint target recognition system [3], [4] in which a super resolution technique was used to estimate the positions and amplitudes of target scattering centers and a Hidden Markov Model (HMM) was used to represent the unknown target orientation. The proposed fusion technique entails only a straightforward modification of the transition probabilities of the HMM.

We begin by giving a brief description of the recognition feature set in Section II and the single viewpoint system in Section III. In Section IV, we explain in details how data from multiple viewpoints are fused and in Section V we present evaluation results for the fusion system. We conclude the paper in Section VI with a brief summary.

II. RECOGNITION FEATURE SET

The features that we use for target recognition are derived from the sequence of complex-valued high range resolution (HRR) profiles, $x(n,k)$, which may be obtained from a complex SAR image chip by the procedure described in [5]. The profile index is denoted by $n \in \{1, \ldots, N\}$ and the range-bin index by $k \in \{1, \ldots, K\}$. Fig. 2(a) shows a typical plot of $|x(n,k)|$ and Fig. 2(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 windows of $2P+1$ profiles and determine a feature vector from

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

$$y(l,k) = \sum_{m=1}^{M_k} |a_{m,k}|^2 \, h(l - \theta \omega_{m,k}) \qquad (3)$$

where the low-pass filter response is given by

$$h(l) = \frac{sin(\pi l)}{\pi l}. \qquad (4)$$

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.

Following the image reconstruction process, the target feature vector is formed by taking the low frequency coefficients of the 2-dimensional discrete cosine transform (DCT) of the log of the pseudo-image $y_n(l,k)$,

$$f_n(\eta,\zeta) = \qquad (5)$$
$$\sum_{k=0}^{K}\sum_{l=0}^{L} \alpha_k \beta_l \, \log\left(y(l,k)\right) \cos\left(\tfrac{\pi(2\eta+1)k}{2K}\right) \cos\left(\tfrac{\pi(2\zeta+1)l}{2L}\right),$$

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 $\mathbf{u}_M(n)$.

III. SINGLE VIEWPOINT TARGET RECOGNITION

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, since the radar signature varies with orientation, a distinct target signature. 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
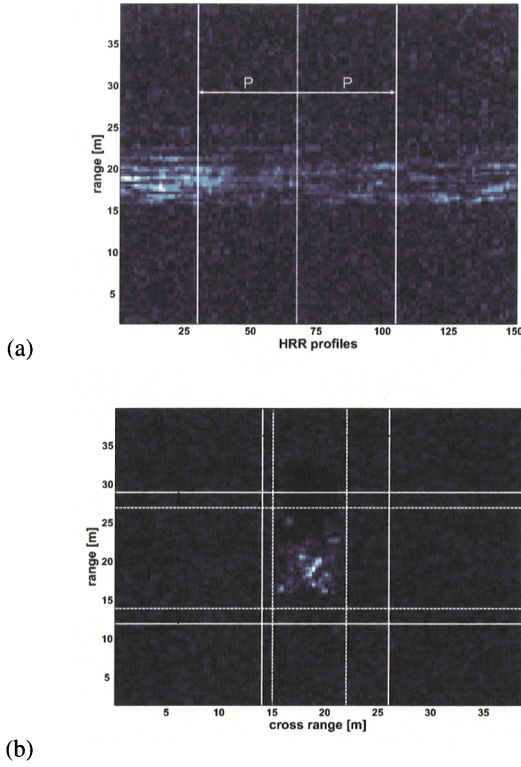

(a)

(b)

Fig. 2. (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.

each window. The feature vector centered on profile n is thus based on the profile subset, $x_n(p,k)$, defined by

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

In Fig. 2(a), the vertical lines indicate the range of profiles used to derive a single feature vector when $P = 37$. We also isolate the target in down- and cross-range using an adaptive mask described in [9]. This is shown by the broken lines in Fig. 2(b) where the mask has been snugly fitted to the target.

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 [6] based on the following data model

$$x(p,k) = \sum_{m=1}^{M_k} a_{m,k} e^{j\omega_{m,k}p} + v(p,k) \qquad (2)$$

where k is the range-bin index, M_k is the number of extracted scattering centers in the range bin, $p \in \{-P, \ldots, P\}$ is the HRR profile index, $a_{m,k}$ is the radar cross section of the mth scattering center of the kth range bin, $\omega_{m,k}$ corresponds to its azimuth location and $v(p,k)$ is assumed to be white noise. The use of a super-resolution algorithm only in azimuth reduces the

target dimension 20 m, 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.
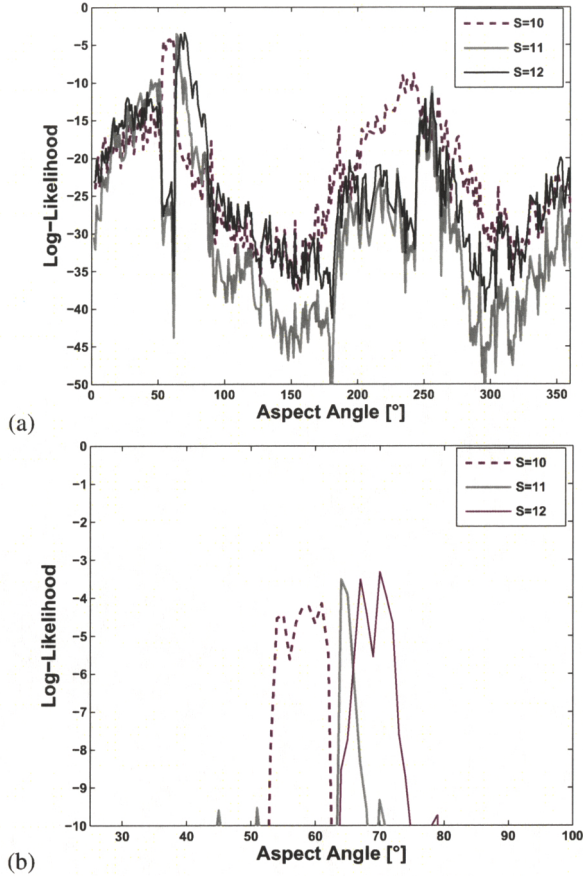

(a)

(b)

Fig. 3. 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°.

To define the initial model parameters, we assume that all states correspond to equal aspect ranges of $360° S^{-1}$. For each state we train a Gaussian mixture model (GMM) [7] 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, we then iteratively re-estimate the GMM parameters and the HMM transition probabilities using embedded Baum-Welch training [8] until their values converge.

The azimuth interval represented by a state can change substantially during the training process as is illustrated in Fig. 3(a). This figure shows the log likelihood of test feature vectors as a function of azimuth angle for each of three consecutive model states. The three model states were initially trained with data from consecutive 6° azimuth intervals in the region of 60°. We see from Fig. 3(a) that the log likelihood does indeed peak at a target orientation of 60° and that there is a second, smaller, peak at around 250° due to target symmetry. An enlarged view of the primary peak is shown in Fig. 3(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. 3(a).

IV. FUSION OF DATA FROM MULTIPLE VIEWPOINTS

We now suppose that target observation sequences are available from two different viewpoints, A and B, having a known angular separation of Φ. We further suppose that the two observations have the same sample rate in both range and cross-range. We would like to combine these observations to improve the performance of the recognizer without requiring any additional model training. The most straightforward way of combining the observation sequences is to apply them separately to the recognizer and, for each test model, to add the log likelihoods of the two sequences together. As we shall see below, this gives a significant improvement in performance but, by making use of the known value of Φ, it has a considerable advantage when single sequences are recognized with wrong states.

A. Joint Model

When the target orientation seen by sensor A lies in the range (u_i, v_i), that seen by sensor B will lie in the range $(u_i + \Phi, v_i + \Phi)$. If all states had the same width, $u_i - v_i$, and Φ were an integer multiple, K, of this width then the range $(u_i + \Phi, v_i + \Phi)$ would correspond exactly to the state $K + i$. We could then fuse the two observation sequences by forming a double-length feature vector from the two observations and performing recognition with a composite model in which the output distributions from states i and $i + K$ are combined. An equivalent architecture which leaves the feature vector length unchanged is shown in Fig. 4. All paths through this
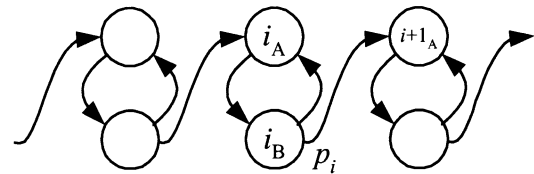

Fig. 4. Architecture of joint model.

model alternate between the upper and lower rows of states which correspond to sensors A and B respectively and the only allowed transition from state i_A is to state i_B as shown. The output distribution in state i_B is identical to that in state $(i + K)_A$. To perform recognition, we interleave the feature

vectors obtained from the two viewpoints to form a single sequence containing twice as many feature vectors and apply this to the model.

B. Unidirectional Model

Unfortunately, we cannot use the joint model described above because it is not the case that all states have the same width after training. Because of this, the azimuth range $(u_i + \Phi, v_i + \Phi)$ will in general overlap with more than one state. Suppose, for example, that $(u_i + \Phi, v_i + \Phi)$ has a non-empty intersection with each of the three states $j + 1$, $j + 2$ and $j + 3$. This means that when the target orientation seen by sensor A corresponds to state i, that seen by sensor B could correspond to any of states $j + 1$, $j + 2$ and $j + 3$. Accordingly, we use the HMM model architecture shown in Fig. 5 in which state i_A can be followed by any of i_{B1}, i_{B2} or i_{B3}; these three states have identical output distributions to states $(j + 1)_A$, $(j + 2)_A$ and $(j + 3)_A$ respectively. The transition probabilities from i_A to i_{B1}, i_{B2} or i_{B3}; model are proportional to the overlap widths between $(u_i + \Phi, v_i + \Phi)$ and (u_{j+1}, v_{j+1}), (u_{j+2}, v_{j+2}) and (u_{j+3}, v_{j+3}) respectively. Only the last of these states, i_{B3} has an allowed transition to state $(i + 1)_A$; the others always return to state i_A.
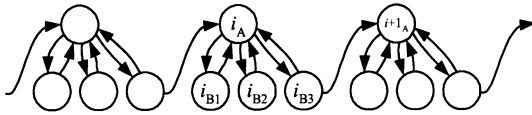

Fig. 5.   Architecture of unidirectional model.

C. Bidirectional Model

A disadvantage of the unidirectional model described above is that, although it retains the same feature length as the original trained model, it requires over three times as many states. We can simplify the model greatly by merging any states with identical output distributions. The resultant model architecture is shown in Fig. 6 in which, for clarity, only the transitions associated with state i have been included. We see that there are transitions in both directions between state i and each of states $j+1$, $j+2$ and $j+3$ and an additional transition from $j + 3$ to state $i + 1$. Thus, when all transitions are included, each state has multiple transitions in both the forward and reverse directions that correspond to azimuth changes of approximately $\pm\Phi$. The advantage of this model is that it contains exactly the same number of states as the original single-viewpoint trained model and that each state has an unchanged output distribution. The only modification needed is a recalculation of the transition probabilities as outlined for the unidirectional model above. A disadvantage of the model is that it no longer imposes an alternation between the forward jumps of $+\Phi$ from sensor A to sensor B and the reverse jumps of $-\Phi$ from sensor B to sensor A. There are now allowed paths through the model that consist entirely of jumps in one direction. We expect, however that these parasitic paths will have low probability and so will not arise in practice.
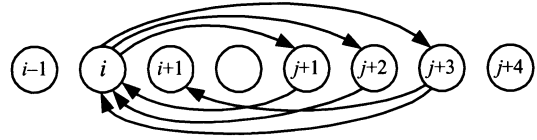

Fig. 6.   Architecture of bidirectional model.

V. EXPERIMENTAL RESULTS

The experimental evaluations are based on the Moving and Stationary Target Acquisition and Recognition (MSTAR) database collected by the Sandia National Laboratory. For the experiments, the SAR image chips were converted into a sequence of HRR profiles as described in [3], [4], [9], [5]. An HMM containing $S = 60$ states was trained for each target giving an average azimuth interval of $6°$ per state. For these closed-set classification tests, the HRR profile sequence from each viewpoint covers a total azimuth aperture of $3°$. Two sensors are located with an azimuth separation, Φ , that is varied from $10°$ to $180°$. We have evaluated the two unidirectional and bidirectional fusion models described above (labelled "unidir" and "bidir") as well as just summing the log likelihoods of independent recognition tests using the single-viewpoint model (labelled "sum"). In addition, we have shown the result of adding the two single-viewpoint log likelihoods onto those obtained from the fusion models (labelled "unidir+indep" and "bidir+indep" respectively). We also show the results for single viewpoint target recognition experiment using both a sequence that covers $3°$ and a $6°$ azimuth matching the total azimuth look angle of the multi-perspective recognizers for fair comparison.

Fig. 7 shows the recognition error rate for a closed set identification task using 10 different targets. A total of 2500 $15°$ depression angle test sequences was used and the graphs show the percentage test set misclassification rate as a function of azimuth separation, Φ , averaged over all target orientations. The inset caption includes the average misclassification rate over all azimuth separation angles. We see that the sum of log-likelihoods, the unidirectional and the bidirectional models all give a similar error rate for viewpoint separation greater than $60°$ except for the $\Phi = 120°$ separation where summing the log-likelihoods works much worse than the other two. In all cases the error rate decreases slightly with increasing Φ.

Although the error rates are similar for all three approaches described above, different classification errors are made in each case. The classification performance can therefore be improved by combining the approaches. When the sum of the log-likelihoods of each sequence is added to the likelihoods obtained from using the unidirectional and bidirectional models the error rates becomes extremely low, as shown in Fig. 8. An average of around 0.039% is achieved in each case, corresponding to only 1 misclassifications from the 2500 test sequences. Encouragingly, there is no significant difference between the unidirectional model and the much
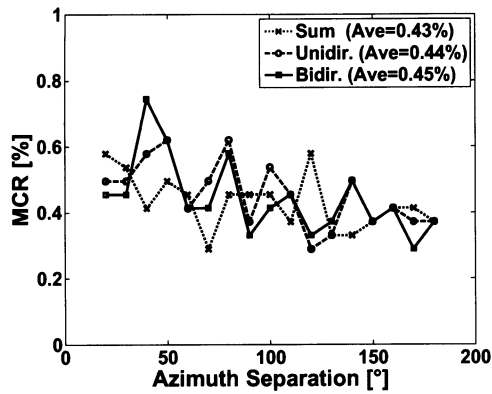
Fig. 7. Test set misclassification rate with azimuth separation changing from 10° to 180° for (a) summing the log-likelihoods, (b) the bidirectional model, and (c) the unidirectional model.
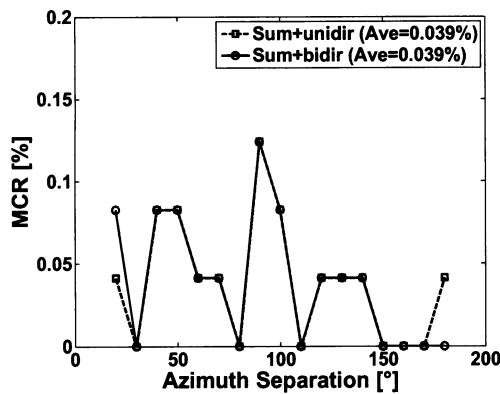

Fig. 8. Test set misclassification rate with azimuth separation changing from 10° to 180° for fusing the summing of the log-likelihoods with (a) the unidirectional and (b) the bidirectional model.

simpler bidirectional version and so the latter is the preferred choice.

A summary of the results is given in Table I. The single perspective target recognition using the longer test sequences achieves an average error rate of 0.5%, which indicates that the multi perspective recognition approach performs significantly better than the single perspective system.

VI. CONCLUSIONS

This paper presents a novel fusion technique based on HMMs for automatic target recognition from HRR profiles when observations from multiple viewpoints are available. The fusion technique only requires modification of the HMM transition probabilities and does not entail any retraining of the target models. Evaluations using the MSTAR database indicate that for viewpoint separations greater than 90° the new technique can reduce classification errors by about two orders of magnitude when compared to single viewpoint observations.

TABLE I
TEST SET MISCLASSIFICATION RATE (MCR) FOR SINGLE AND TWO-PERSPECTIVE RECOGNITION. THE SINGLE VIEW 6° AND THE TWO-PERSPECTIVE EXPERIMENTS ARE BASED ON THE SAME AMOUNT OF DATA AND THEREFORE COMPARABLE.

Method	(Average) MCR [%]
Single View, 3°	2.8
Single View, 6°	0.5
Sum of log-lik.	0.43
Unidirectional	0.44
Bidirectional	0.45
Sum + Unidir	0.039
Sum + Bidir	0.039

REFERENCES

[1] M. Vespe, C.J. Baker, and H.D. Griffiths, "Multi-perspective target classification," in *Proc IEEE Intl Radar Conf*, May 2005, pp. 877–882.

[2] M. Vespe, C.J. Baker, and H.D. Griffiths, "Aspect dependent drivers for multi-perspective target classification," in *Proc IEEE Conf on Radar*, Apr. 2006.

[3] 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*, March 2005, pp. 589–592.

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

[5] 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*, vol. 38, no. 4, pp. 1230–1242, Oct. 2002.

[6] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[7] 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*, vol. 39, no. 1, pp. 1–38, 1977.

[8] 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.

[9] J. Cui, J. Gudnason, and M. Brookes, "Maximum A-posteriori adaptive masking for clutter suppression in automatic radar target recognition," in *IEEE Conf. on Radar*, 2006.