

HIDDEN MARKOV MODELLING FOR SAR AUTOMATIC TARGET RECOGNITION

Chanin Nilubol, Quoc H. Pham, Russell M. Mersereau, Mark J. T. Smith, and Mark A. Clements

Center for Signal and Image Processing
School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, GA 30332, USA.

(Tel: 404-894-6291. Fax: 404-894-8363. Email: chanin,qpham,rmm,mjts,clements@ee.gatech.edu)

ABSTRACT

This paper discusses the application of Hidden Markov Models (HMMs) to solve the Translational and Rotational Invariant Automatic Target Recognition (TRIATR) problem associated with SAR imagery. This approach is based on a cascade of these stages: preprocessing, feature extraction and selection, and classification. Preprocessing and feature extraction and selection involve successive applications of extraction operations from measurements of the Radon transform of target chips. The features which are invariant to changes in rotation, position and shifts, although not to changes in scale are optimized through the use of feature selection techniques. The classification stage successively takes as its inputs the multidimensional multiple observation sequences, parameterizes them statistically using continuous density models to capture target and background appearance variability, and thus results in the TRIATR-HMMs. Experimental results have demonstrated that the recognition rate is as high as 99% over both the training set and the testing set.

Keywords: HMMs, ATR, SAR, invariant, Radon Transform, and rotation.

1. INTRODUCTION

The problem of Automatic Target Recognition (ATR) is a difficult one. The appearance of a target can often assume a wide variety of rotational and shifted positions and may be altered by near-by and occluding clutter. Effective end-to-end recognition systems that performs well in unpredictable environments remains an unsolved problem. Details about some complex approaches that have been used to overcome these difficulties by deploying statistical pattern recognition, model vision based methods, or neural networks are reviewed in [1] and [2]. Hidden Markov Models (HMMs) offer potential to address ATR in a unique way. Historically, they have been used for various applications of speech recognition, optical character recognition, and handwriting verification. The power of HMM algorithms may be attributed to the fact that the observed data can be stretched or shrunk to fit the models, and that they

This work was supported by the United States Army Research Laboratory under cooperative agreement of DAAL01-96-2-0001. Authors would like to thank Mr. Anson Dickson at Wright Laboratory for providing us a public MSTAR dataset.

are computationally efficient. In the case of speech recognition, for instance, speaker identifiers have been developed that are insensitive to different dialects. Likewise, the design of target recognizer that can tolerate a certain degree of variability in appearance should be possible. In this paper, we consider ATR of SAR images. The approach taken is to parameterize the Translational and Rotational Invariant Automatic Target Recognition Hidden Markov Models (TRIATR-HMMs) with features that are independent of rotational effects. The Radon Transform and Discrete Fourier Transform (DFT) are two attractive representations that can be used because. The Radon transform displays image information in polar form that is independent of the cyclical ordering. The Discrete Fourier transform (DFT) magnitude remains unchanged with respect to such orderings. If target reflection is ignored, rotational and translational invariant features and precise HMM parameter estimates can be achieved using these two transformations and the target variation problem can be mitigated.

In Section 2, an overview of HMMs is introduced. In Section 3, the system configuration and implementation are summarized. Experimental results illustrating the performance of the system is shown in Section 4, followed by a summary in Section 5.

2. OVERVIEW OF HIDDEN MARKOV MODELS

2.1. Introduction to HMMs

A discrete Markov process is characterized by a finite number of states. The system is in one of these states at any time and changes between the states take place at equally spaced discrete times according to the state transition probabilities associated with each state. In the case of a first-order Markov chain, the state transition probabilities do not depend on the history of the process, but only on the current state. If there is a unique observation symbol joined with each state, the process is characterized by the above parameters. For many problems of interest, different observation symbols may occur in each state and given an observation, it is not possible to decide which state the model is in. In this case the observation is a probabilistic function of the state. The underlying stochastic process is not observable, it is hidden, and therefore these models are called hidden Markov models. Indeed, an HMM is nothing more than a probabilistic function of a Markov process.

2.2. Necessity of HMMs

HMMs are useful when one can think of underlying events probabilistically generating surface events. One widespread use of this is tagging—assigning parts of speech (or other classifiers) to the words in a text. We think of there being an underlying Markov chain of parts of speech from which the actual words of the text are generated.

When this general model is suitable, the further reason that HMMs are very useful is that they are one of a class of models for which there exist efficient methods of training through use of the Expectation Maximization (EM) algorithm. Given sufficient data that is assumed to have been generated by (some) HMM—where the model architecture is fixed but not the transition probabilities—this algorithm allows us to learn the model parameters automatically that best account for the observed data. This is in contrast to the traditional Artificial Neural Network (ANN) architectures where the conditions are interrelated, the parameter vector (weights matrix) is random and it is estimated in such a manner so that condition discrepancy is maximized. However, HMMs do have important limitations. The biggest is that HMMs do not capture any higher-order correlations. Furthermore, the assumption that the distributions of individual observation parameters can be well represented as a mixture of Gaussian or autoregressive densities [3] may be inappropriate for underlying distribution in many cases. However, it is not unreasonable to make this assumption in this experiment due to sufficient training data.

3. MULTISTAGE TARGET CLASSIFICATION ALGORITHM

3.1. Data Preparation

The MSTAR database consisting of training and testing datasets is investigated as part of the evaluation effort. This computer generated database spans 360 degrees of target azimuth pose angle variation in 1-2 degree increments with 17 and 15 degrees of depression angle for the training and testing set respectively, i.e., there is a deliberate mismatch between the training and testing images. The public portion of the database is composed of two trucks and one tank: BMP, BTR and T72, respectively. In addition, the BMP and T72 directories provide images at different levels of noise. Figure 1 depicts original targets at different orientations.

3.2. System Configuration

The system configuration in Figure 2 depicts the training phase as well as the testing phase. In each of the following subsections, details about each block are discussed.

3.2.1. Training Phase

Referring to Figure 2 (a), the target chip is first preprocessed by a normalizing filter. The filter consists of a rectangular annular window that border the target chip like a picture frame. The thickness of the frame can be varied depending on how much of the background information is to be

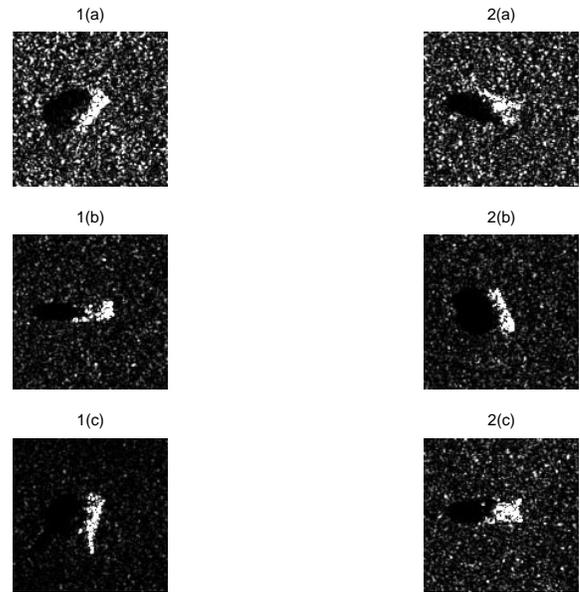


Figure 1: High resolution (1 ft by ft) synthetic-aperture radar (SAR) images of three signatures. Column 1 and 2 depict images from the training and testing dataset, respectively. 1(a), 2(a):BMP, 1(b), 2(b): BTR, 1(c), 2(c): T72.

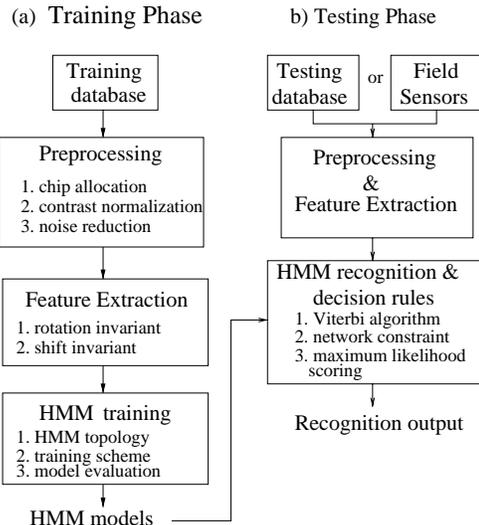


Figure 2: A diagram for training and testing with HMM (a) Training Phase and (b) Testing Phase

(a) Training				(b) Testing			
(10,1)				(10,1)			
Percent Classified as				Percent Classified as			
	BMP	T72	BTR		BMP	T72	BTR
BMP	37.8	30.0	32.2	BMP	24.1	57.0	18.9
T72	14.8	73.0	12.2	T72	12.1	74.4	13.5
BTR	31.5	12.9	55.6	BTR	28.2	31.3	40.5
total percent recognition = 55.4				total percent recognition = 47.9			
(10,2)				(10,2)			
BMP	47.3	25.9	26.8	BMP	30.5	55.1	14.4
T72	19.2	74.1	6.7	T72	17.6	75.1	7.3
BTR	19.0	2.6	78.4	BTR	26.2	21.5	52.3
total percent recognition = 63.2				total percent recognition = 52.6			
(10,3)				(10,3)			
BMP	91.2	0.6	8.2	BMP	89.2	5.0	5.8
T72	0.3	99.7	0	T72	0	99.8	0.2
BTR	2.6	0	97.4	BTR	16.9	1.5	81.6
total percent recognition = 95.7				total percent recognition = 92.6			
(10,4)				(10,4)			
BMP	88.6	2.6	8.8	BMP	90.8	5.3	3.9
T72	0.4	99.3	0.3	T72	0.3	99.3	0.4
BTR	11.6	0	88.4	BTR	36.4	2.6	61.0
total percent recognition = 93.1				total percent recognition = 90.1			
(10,5)				(10,5)			
BMP	90.9	1.9	7.2	BMP	92.8	2.6	4.6
T72	1.5	97.9	0.6	T72	1.7	98.3	0
BTR	17.2	0	82.8	BTR	39.5	1.0	59.5
total percent recognition = 92.8				total percent recognition = 90.3			
(10,6)				(10,6)			
BMP	92.4	1.4	6.2	BMP	90.9	2.9	6.2
T72	1.0	99.0	0	T72	1.2	98.6	0.2
BTR	19.0	0	81.0	BTR	39.5	1.5	59.0
total percent recognition = 93.6				total percent recognition = 89.6			
(10,7)				(10,7)			
BMP	86.9	4.5	8.6	BMP	84.1	7.4	8.5
T72	3.0	96.7	0.3	T72	2.2	96.9	0.9
BTR	9.5	0.9	89.7	BTR	27.7	2.6	69.7
total percent recognition = 91.5				total percent recognition = 87.5			
(10,8)				(10,8)			
BMP	77.1	1.4	21.5	BMP	79.1	3.6	17.3
T72	2.2	97.8	0	T72	1.9	97.9	0.2
BTR	6.1	0	93.9	BTR	21.5	1.5	77.0
total percent recognition = 88.4				total percent recognition = 86.8			

Table 1: An effect of the number of best features on the classification performance evaluated on two datasets: (a) training and (b) testing, where (m,n) is a model with m states and n best feature(s). The empirical results are based on a suboptimal Sequential Forward Search (SFS) search strategy with Euclidean inter-class distance selection criterion.

included in the filter calculations. The background statistics are modeled by calculating the mean, u , and standard deviation, σ , of the pixels contained within the outer and inner borders of the frame. The mean is subtracted from the pixel being normalized, x , and the result is divided by σ as shown,

$$y_{ij} = \begin{cases} (x_{ij} - u_{ij})/\sigma_{ij} & \text{if } y_{ij} \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

where y is the normalized output. The purpose of this filter is to highlight regions of the image that contrast substantially with the local background characteristics.

The next stage of the feature extraction is to take the Radon transform to incorporate target signatures at as many orientations as possible. To reduce the computation, the target chip is down-sampled by a factor of 2 before the Radon transform is taken. The 64 projections of a target chip are taken. The mean, standard deviation, and width of each projection is then extracted. The width is calculated by counting the number of pixels with value greater than the mean. They then feed into the Fourier transform, sort, and shift invariance techniques. The shift-invariance method is achieved by shifting maximum value to the center of the vector and then moving the centroid of the new

vector to the center. By doing that, we have a shift and a rotation invariant feature matrix of 64×9 .

Due to the fact that the magnitude of the Fourier transform can be very large at some points, the original feature vector is replaced by its mean/gain normalization (the ratio of the vector and the mean vector are computed across all images within the same signature type). This is done to reduce the dynamic range of the feature values and thus allow more precise estimation of the Gaussian parameters of the HMM and, eventually, better recognition performance.

In the HMM approach, the problem can be stated as follows. The HMMs for each target type should be carried out with the ability to discriminate between target types. The topology of the HMM includes a user-specified number of emitting states, a prespecified left-to-right model with a skip transition allowed, a diagonal covariance matrix, single mixture, continuous density and a multi-input-multi-output system.

A choice of size and type (topology) of HMMs have been studied for ATR systems. In this case, small structures with 10 states or fewer are suitable not only for a level of complexity of MSTAR dataset but also for statistically clustered data (no hand registration). Of particular relevance to this work is the challenge to distinguish between the two trucks which look very similar visually.

The observation sequence for each training token is segmented into state clusters by determining the optimum alignment of the current model with each training token. The segmentation is achieved using the Viterbi algorithm. The initial estimate of the model parameters can be assigned on the basis of any available model that is appropriate to the data and probabilistic requirements. The model parameters are then reestimated according to histograms of the results of the segmentation. The final step in the training loop is the test for convergence, the terminating condition for the training procedure. The resulting model is compared to the previous model by examining the accumulated likelihood (Viterbi) scores between training tokens and each model. If the difference falls below a threshold, then model convergence is assumed. Otherwise, the overall training loop is repeated.

3.2.2. Testing Phase

The preprocessing and feature extraction are repeated in the testing phase. The difference from the training begins at the Viterbi decoder, which can also be viewed as a transformer that maps image features into a sequence of coded states. The decoder finds the single best state sequence by optimally mapping the observation sequence of the data onto each target model. The mapping is computed from the observation probabilities, state transition probabilities and other initial model parameters according to a maximum likelihood criterion.

The recognition procedure is accomplished by evaluating two Bayesian distortion measures of an observation vector with respect to the all target models. For each unknown input sequence which is to be recognized, the processing of Figure 2 (b) must be carried out, namely a calculation of model likelihoods for all possible models, and followed by selection of the target model whose likelihood is high-

est. This problem is similar to isolated word recognition in speech processing [3].

4. EXPERIMENTAL RESULTS

Tables 1 and 2 show the classification performance of the baseline classifier as a confusion matrix that tabulates the correct and incorrect classifications. Several combinations of a number of state running from 2 to 11 and a number of channels (best features) ranging from 1 to 9 were tested. An HMM with 10 states and 3 best features provided the best overall performance among 90 different reference models that were tested. Recall that the classifier used HMM parameters constructed purely from training targets; for both tables, the classification results shown in the left column (a) belong to training dataset whereas the right column (b) shows the detection rate for the testing dataset that passed the same preprocessing and feature extraction stages. Table 1 shows the classification performance of both the training and testing data set with the number of state, m, fixed at 5. Table 2 shows the percent of correct classification with the number of features fixed at 7. From table 1, when the number of best features is allowed to increase, the performance is better when 3 to 6 best features are used. Regardless of the similarity between BMP and BTR and the mismatch conditions, the ability to correctly classify between targets is retained at an acceptable level.

There are a few curious properties to observe about these two tables. First of all, from observing Table 1, it can be seen that distinguishing T72 from other classes is achievable with a smaller number of best features compared to the other two targets. Second, to get an idea of how large the number of state should be, the number of state is raised from 7 to 11 with the number of best features (obtained from an optimal search branch-and-bound (BB)) fixed at 7, from table 2, the best performance is attained by a HMM with 9 or 10 states. The reason for this is that at first the training sequences containing 695, 688, and 232 images for BMP, T72, and BTR respectively, when trained, yield poor BTR statistical parameters due to its insufficient number of BTR training images thus, as the number of state increases from 10 to 11, recognizing BTR causes the problem in both training and testing cases although BMP and T72 are more correctly classified. It is clear from these two tables that BMP and T72 benefit from having more states than the BTR.

5. SUMMARY

HMMs implemented with the proper choice of features were found useful in solving ATR problem. The performance ranges according to the number of states, the topology and the complexity of training data used. In the experiment, it was found empirically that an HMM with 10 states and 3 best features could effectively distinguish all 3 military targets: BMP, BTR and T72. The entire public MSTAR database was used as part of the investigation.

(a) Training				(b) Testing			
(7,7)				(7,7)			
Percent Classified as				Percent Classified as			
	BMP	T72	BTR		BMP	T72	BTR
BMP	84.7	4.6	10.7	BMP	85.4	5.8	8.8
T72	14.4	85.5	0.1	T72	12.3	87.7	0
BTR	24.1	1.3	74.6	BTR	37.4	3.6	59.0
total percent recognition = 83.6				total percent recognition = 82.6			
(9,7)				(9,7)			
Percent Classified as				Percent Classified as			
	BMP	T72	BTR		BMP	T72	BTR
BMP	90.9	1.1	8.0	BMP	87.1	2.4	10.5
T72	1.7	98.2	0.1	T72	1.0	98.8	0.2
BTR	22.0	0.4	77.6	BTR	29.2	0.5	70.3
total percent recognition = 92.1				total percent recognition = 89.7			
(10,7)				(10,7)			
Percent Classified as				Percent Classified as			
	BMP	T72	BTR		BMP	T72	BTR
BMP	87.9	1.7	10.4	BMP	88.3	2.7	9.0
T72	0.7	98.8	0.5	T72	1.0	97.9	1.1
BTR	12.0	1.3	86.7	BTR	25.1	2.0	72.9
total percent recognition = 92.4				total percent recognition = 90.2			
(11,7)				(11,7)			
Percent Classified as				Percent Classified as			
	BMP	T72	BTR		BMP	T72	BTR
BMP	94.5	1.3	4.2	BMP	94.5	2.4	3.1
T72	1.3	98.7	0	T72	1.2	98.8	0
BTR	18.5	0	81.5	BTR	41.5	1.0	57.5
total percent recognition = 94.2				total percent recognition = 91.0			

Table 2: An effect of increasing the number of state from 7 to 11 on the classification performance evaluated on two datasets: (a) training and (b) testing , where (m,n) is a model with m states and n best feature(s). The empirical results are based on an optimal branch-and-bound (BB) search strategy with Euclidean inter-class distance selection criterion.

6. REFERENCES

- [1] B. Bhanu, D.E. Dudgeon, E.G. Zelnio, A. Rosenfeld, D. Casasent and I.S. Reed, "Special issue on Automatic Target Recognition," IEEE Trans. IP, vol. 6, 1997.
- [2] D.E. Dudgeon and R.T. Lacoss, "Special Issue on Automatic Target Recognition," The Lincoln Laboratory Journal, vol. 6, no. 1, Nov. 1993.
- [3] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proc. IEEE, vol 77, pp. 257-286, Feb. 1989.
- [4] O.K. Al-Shaykh and J.F. Doherty, "Invariant image analysis based on Radon transform and SVD," IEEE Trans. on Circuits and Systems II, vol. 43, pp.123-133, Feb. 1996.
- [5] D.E. Kreithen, S.D. Halversen and G.J. Owirka, "Discriminating Targets from Clutter," The Lincoln Laboratory Journal, vol. 6, pp.28-29, Nov. 1993.