# HIDDEN MARKOV MODELING FOR AUTOMATIC TARGET RECOGNITION *

*Dane P. Kottke, Jong-Kae Fwu, Kathy Brown*
Signal Processing Center
Sanders, A Lockheed Martin Company, Nashua, NH 03061
dkottke@sanders.com, fwu@sanders.com, klbrown@sanders.com

## ABSTRACT

A novel approach for applying Hidden Markov Models (HMM) to Automatic Target Recognition (ATR) is proposed. The HMM-ATR captures target and background appearance variability by exploiting flexible statistical models. The method utilizes an unsupervised training procedure to estimate the statistical model parameters. Experiments upon a Synthetic Aperture Radar (SAR) database were performed to test robustness over range of target pose, variation in target to background contrast, and mismatches in training and testing conditions. The results are compared against a template matching approach. The HMM captures target appearance variability well and significantly outperforms template matching in both robustness and flexibility.

## 1. INTRODUCTION

Automatic target recognition (ATR) is the processing of sensor data to detect, localize, and recognize targets. ATR is difficult due to variations target appearance and environmental conditions. A target's appearance will be affected by changes in its pose, location, distance from the sensor, and its operating state. Changes in background/clutter and environmental conditions will also make the process of reliably detecting targets more difficult. Many practical systems are based upon deterministic template matching approaches. Template matching is attractive due to its simplicity and is an appropriate method under ideal conditions of "in the open" targets with no appearance variability. However, it has the disadvantage of being sensitive to several factors, such as occlusion, target appearance variation, template and test image geometry mismatches, and the presence of confounding clutter. More complex approaches attempt to overcome these difficulties by employing statistical pattern recognition, model vision based methods, or neural networks. Details of these techniques are reviewed in [2], [5].

Hidden Markov Models (HMMs) technology has been successfully applied to face recognition [10], optical character recognition [4], and recognition of objects defined by closed contours [6]. These approaches are intended for applications where the objects to be recognized do not rotate, translate, or where they can be described by their outer and inner boundaries. These assumptions are not appropriate for ATR. Burger and Bhanu [3] applied HMM to the ATR problem by using a model independent, heuristic decision rule for data sequencing. In this paper, we propose a novel approach for applying HMM to ATR. Our method directly addresses the problems associated with object rotation and translation by applying a rotationally and translationally invariant data sequencing. The method brings the following benefits: (1) A robust and flexible target representation is created in terms of a Hierarchical Target Grammar (HTG) constructed upon HMMs. (2) Global and local image information is combined for recognition. (3) Efficient implementation is made possible by applying a rich heritage of HMM technology.

The paper is organized as follow: In section 2, a framework for applying HMM to ATR is given as well as the details of each module in our system implementation. In section 3, we present experimental results for both the HMM-ATR method and a template matching approach. Finally, a conclusion and additional work are presented in section 4.

## 2. HMM-ATR SYSTEM

Figure 1 depicts both the training and recognition components of the HMM-ATR system. In the training stage, HMM target models are generated, based upon the presented data and the target grammar. These target models are used to recognize the targets by the process of target decoding. Each component of the HMM-ATR is described below.

### 2.1. Feature Extraction

The goal of feature extraction is to reduce noise and extract features which enable discrimination between target types and target from clutter. Note that feature extraction is not limited to a single sensor since we can utilize the HMM to efficiently integrate the feature information obtained from multiple sensors to get better performance. For the SAR-ATR problem, the images are segmented into four classes images by the Iterative Conditional Modes (ICM) method [1]. The four class image is denoted by $f(m,n)$, and the four classes correspond to background, target shadow, dim, and bright regions. The ICM scheme is an efficient approach for image segmentation based on both intensity and spatial information. Interested reader are referred to [1], [7] for further details.

*PREPARED THROUGH COLLABORATIVE PARTICIPATION IN THE ADVANCED SENSORS CONSORTIUM SPONSORED BY THE U.S. ARMY RESEARCH LABORATORY UNDER COOPERATIVE AGREEMENT DAAL01-96-2-0001



1058-6393/98 $10.00 © 1998 IEEE

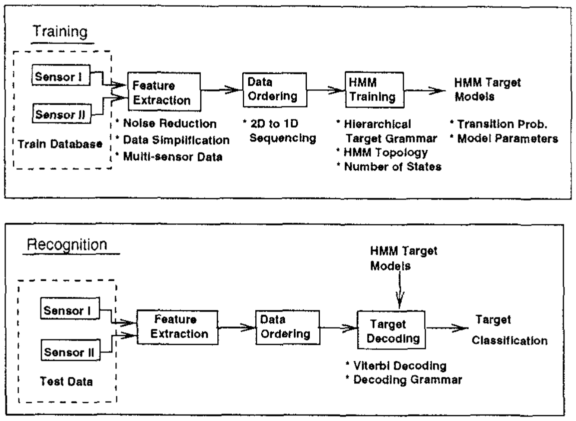


**Figure 1. Training and Recognition Components of the HMM-ATR.**

## 2.2. Data Ordering

The purpose of data ordering is to convert the 2D spatial image representation into a sequential representation. Such a representation allows application of existing HMM technology. This avoids the need for fully connected 2D HM models, which are exponentially complex in the size of the image and require intensive amounts of training data. To avoid the complexities of the 2D HMM problem, "Pseudo-2D" HM models have been proposed for hand-writing recognition and face recognition [10]. Since they are fully connected only in the horizontal direction, they are still sensitive to translation in the vertical direction and to rotation. In our approach, we apply a Radon transform (RT) to each segmentation class over a set of projection angles. The ouput of the RT at the current projection angle is concatenated to the previous projection angle outputs to create a sequential feature stream. At each projection angle, multiple RTs are applied, one for each segmentation class; this is termed a multi-channel RT or MCRT. The benefit is that the MCRT maps target translation and rotation in such as way that they can be accounted for within structure of an HMM.

The Radon transform (RT) maps the image data from a 2D space $(m-n)$ to a 2D parameter space $(\theta - s)$. The relationship between $(m-n)$ and $(\theta - s)$ coordinates are presented in Figure 2. The RT of an image $F(s, \theta)$ is the collection of its line integral along a line inclined at a angle $\theta$ from the $n$-axis and at a distance s from the origin for $0 \leq \theta < \pi$. It can be presented by

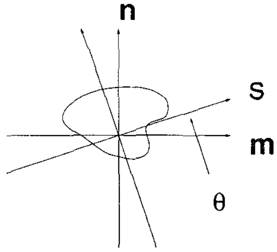


**Figure 2. Coordinates of $m-n$ and $\theta - s$.**

$$F(\theta, s) = \sum_m \sum_n f(m,n)\delta(s - mcos\theta - nsin\theta), \quad 0 \leq \theta < \pi. \tag{1}$$

Some basic properties of RT, such as linearity, shifting, and rotation are described below:

1. Linearity:

$$f(m,n) = \sum_q c_q f_q(m,n) \Rightarrow F(\theta, s) = \sum_q c_q F_q(\theta, s),$$

where $c_q$ are constants.

2. Shifting:

$$\begin{aligned} h(m,n) &= f(m-m_0, n-n_0) \Rightarrow \\ H(\theta, s) &= F(\theta, s - m_0 cos\theta - n_0 sin\theta) \end{aligned}$$

i.e., the shift in $m-n$ causes a shift in $s$ coordinate.

3. Rotation:

$$h(\phi, r) = f(\phi - \phi_0, r) \Rightarrow H(\theta, s) = F(\theta - \phi_0, s).$$

i.e., *rotation causes a shift in $\theta$ coordinate.*

Based on the shift and rotation properties of the RT and the efficiency of HMM modeling, the target translation and rotation problems are automatically absorbed into the HTG, as described later.

To create a sequential data stream, a multi-channel Radon Transform (MCRT) at a discrete set of projections, $\theta_i, i = 1, 2, \cdots, k$, is applied to the segmented image data $f(m,n)$. Each projection generates a $c$ dimensional vector called a projection segment (PS) via

$$\begin{aligned} F_1(\theta_i, s) &= \sum_{(m,n)\in R_1} f(m,n)\delta(s - mcos\theta_i - nsin\theta_i) \\ F_2(\theta_i, s) &= \sum_{(m,n)\in R_2} f(m,n)\delta(s - mcos\theta_i - nsin\theta_i) \\ &\vdots \\ F_c(\theta_i, s) &= \sum_{(m,n)\in R_c} f(m,n)\delta(s - mcos\theta_i - nsin\theta_i), \end{aligned}$$

where $R_j$ denotes the regions of $j$-th class and $c$ is the number of segmentation classes. The data in $(\theta_i - s)$ space is then simply concatenated to create a sequential feature stream; *e.g.*, the first data element in the $k^{\text{th}}$ projection segment is appended to the last data element in the $k^{\text{th}} - 1$ projection segment.

From the above equations, a MCRT can be simply interpreted as a conventional discrete Radon Transform applied to each segmentation class. Based on the linearity property of RT, we can obtain the RT by summing the MCRT over all channels; *i.e.*, $F(\theta, s) = \sum_{i=1}^{c} F_i(\theta, s)$. This process is in-reversible; we can get the data features of RT from MCRT but not the other way around. The projections of each channel represent the extracted feature of its associated class. The channel/class separation help capture the fine details of a target class during HMM training. In Figure 3, we present an example of MCRT. The number of classes determines the number channels; $k = 4$ in this case. Each class of the segmented image is projected downward and its associated profile is plotted on the bottom of the figure. The projection segments obtained from various projection angles are concatenated to create the observed projected feature data stream.

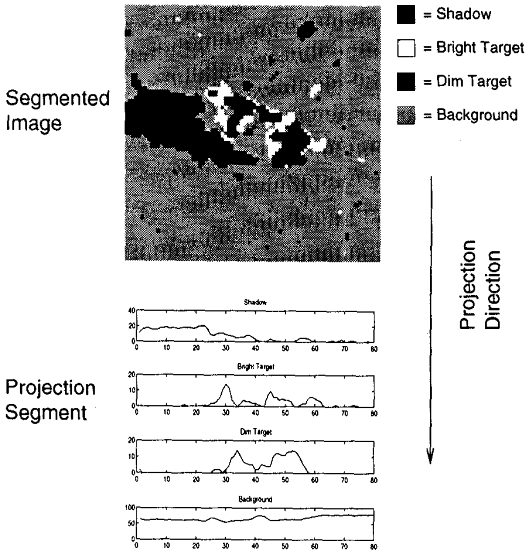


**Figure 3. An example of MCRT.**

To perform a simple visual validation, a backprojection is applied to reconstruct the segmented image from the extracted feature vectors via

$$
\begin{aligned}
r_1(\theta, s) &= \sum_{\theta} F_1(m,n)\delta(s - mcos\theta - nsin\theta) \\
r_2(\theta, s) &= \sum_{\theta} F_2(m,n)\delta(s - mcos\theta - nsin\theta) \\
&\vdots \\
r_k(\theta, s) &= \sum_{\theta} F_k(m,n)\delta(s - mcos\theta - nsin\theta)
\end{aligned}
$$

where the $r_i$, for $i = 1, 2, \cdots, k$, represents the reconstructed image associated with class $i$. The reconstructed image is obtained by applying a simple threshold on $r_i$, for $i = 1, 2, \cdots, k$, and combining the results of the different classes.

Figure 4 contains an example of the MCRT. On the top left is a SAR image of a BTR-70 taken from the MSTAR public target data set [8]. The four class ICM segmentation is displayed on the top right. The four classes represent shadow, dim target, bright target, and background. The extracted feature sequence obtained from different projection angles is displayed on the bottom left. To perform a simple visual validation, a backprojection is applied on each channel and the results combined. The reconstructed image is shown on the bottom right; it is visually similar to the original segmented image. This means that most of the 2D information is kept during feature sequencing.

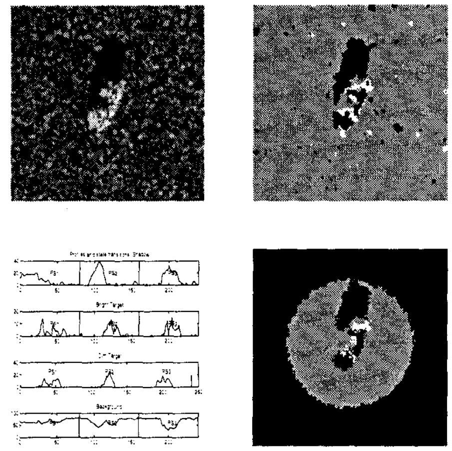

**Figure 4. Top left: original image. Top right: segmented image. Bottom left: extracted feature sequence for three projection angles. Bottom right: reconstructed image.**

### 2.3. HMM Training and Decoding

In the training stage, models are built that discriminate between different target types. The problem of target appearance variability is addressed through a *Hierarchical Target Grammar* (HTG). As shown in Fig. 5, the HTG represents each target as a network (object level) whose elements consist of projection segments (regional level). The HTG corresponds to a probabilistic network of HMMs connected in a hierarchical fashion. The HMMs are built upon projection segments. Minor variations in appearance are modeled at the low level through use of projection segments (i.e., "regional" view). Each projection segment is represented by an HMM which encodes localized information, such as edges, intensity peaks, or textured regions, as extracted through feature extraction and MCRT projection. Major appearance changes due to different poses or articulations are modeled at the high level of the hierarchy. The higher level of HTG specifies the manner in which projection sequence may be ordered; i.e., it describes how the regional views are interconnected to build the larger target model.

The hierarchical structure of the HTG facilitates solution of target rotation and translation. Translation in $m-n$ causes only a shift in $s$ coordinate. The shift in $s$ is absorbed into the observation density and state transition parameters of projection segment HMM. Rotation causes an ordering shift of projection segment (i.e., shift in $\theta$), which is absorbed into the decoding network. The translation problem is reconciled by the low level HMM model of HTG, whereas, the rotation problem is modeled by the high level of HTG.

At the high level of the HTG, a simply-constrained target network build upon 10 projection segments is implemented. At the low level, the training process entails specifying the target model (HMM topology, number of states) and estimating the model parameters (observation densities and transition probabilities). In the HMM experiments, a left-to-right model with skip transition allowed, 8 underlying states, and single mixture, continuous density model was chosen. The model parameters are estimated by the Baum-Welch algorithm [9]. The final estimated parameters for each different target are stored for decoding stages. Note that, our HMM training is an unsupervised training method which automates the tedious, data intensive training procedures. Unlike HMM techniques for speech processing, our training methods do not require hand labeling of data segments since the ICM segmentation generates class labels.

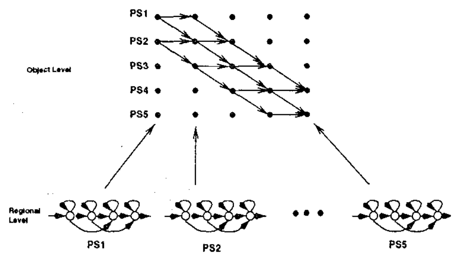


**Figure 5. A Hierarchical Target Grammar (HTG).**

The role of an HMM decoder is to map the input sequence of data into a target type which can be decoded by the Viterbi algorithm [9]. It searches for the most likely sequence of states over all possible state sequences according to a maximum likelihood criterion and target grammar. The criterion is computed from the probability density functions, state transition probabilities, and constrained by the decoding networks.

## 3. EXPERIMENT RESULTS

The HMM-ATR method was tested on a small three target SAR dataset, and compared against a multi-level template matching method based upon the same 4 class ICM segmentation algorithm. The template matching system was developed under the Army Research Labs, Advanced Sensor Federated Labs program for SAR ATR [7]. The ICM segmentation method also performed feature extraction for the HMM-ATR system. Providing the HMM and the template matcher with the same input features, ensures that the test is not biased to one method or the other due to differences in input data.

To evaluate the accuracy and robustness of the ATR performance, three different experiments were designed. The experiments were (I.) changing target's range of azimuth, (II.) increasing background noise (reducing SNR), and (III.) creating a mismatch between training and testing conditions. To evaluate the segmentation and template matching performance two different operating conditions were chosen. The first represents optimal conditions which are rarely seen in fielded systems, but are useful to determine the upper bound on system performance. The second is more realistic and is useful to predict actual system performance.

1. Optimal Conditions
   - Hand registration of target data (template matching only)
   - Reduced range of target appearance
   - Training/testing conditions are the same
   - High target to background contrast
2. Realistic Conditions
   - Automated registration of target data (template matching only)
   - Wide range of target appearance
   - Training/testing conditions are not the same
   - Reduced target to background contrast

Experiment I was designed to evaluate the recognition robustness as range of viewing angle increases. The Percentage of Correct Classification (PCC) results are shown in Figure 6. The multi-level template matching scheme performed well on small range (up to 10°) of azimuth angle. Its performance degraded rapidly when we further increased the azimuth angle. The reason is that the angle variation exceeds the tolerance range of template matching method. In contrast, the HMM method still performs well for wider angle variation. (The range of elevation was 2.5°.)

In experiments II and III, the range of elevation increased from 2.5° to 5.0°, and one of the targets shows two articulations. Additionally, these two experiments investigated how changes in SNR affect both HMM-ATR and template matching. In order to do so, it was necessary to augment the data provided by changing the SNR between target and background. The SNR is defined by $SNR = 10 \log \frac{\sigma_t^2}{\sigma_b^2}$ where $\sigma_t^2$ and $\sigma_b^2$ are the variances in background and target regions, respectively. The background and target variances are determined relative to the background mean and expressed as

$$\sigma_t^2 = \frac{1}{N_t} \sum_{s \in target} (y_s - \mu_b)^2$$

$$\sigma_b^2 = \frac{1}{N_b} \sum_{s \in background} (y_s - \mu_b)^2$$

where $N_t$ and $N_b$ are the numbers of pixels in target and background regions and $\mu_b$ is the sample mean of the background region. Based on the above definition, the average SNR for the unscaled SAR images was $SNR_0 = 6.64$ dB.

To manipulate the SNR values, the background intensities are scaled by a factor, $\sqrt{A}$, in three simple steps: (1) Generate a zero mean background region by subtracting the image's sample background mean from the background pixels. (2) Multiply the pixel values of the zero mean background region by $\sqrt{A}$. (3) Add back the image's background sample mean. Note that the new background variance is simply $\sigma_b^2 = A\sigma_{b0}^2$. This technique avoids trying to simulate the noise distribution of the background; only the level of noise is manipulated by scaling. The background noise was varied from 6.64 dB (no scaling, A = 1) to 0.62 dB (A = 4.0).

Experiment II was designed to evaluate the effects of decreasing SNR in both training and testing. The delta azimuth angle was limited to 10°, which is same as the real operating condition for template matching applications. The

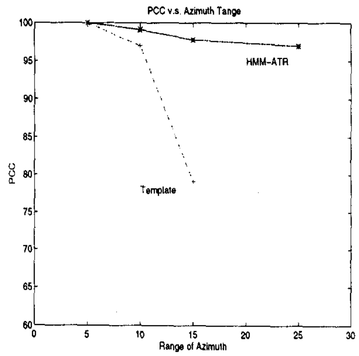


**Figure 6. Experiment I Percent Correct Classification (PCC) results:Hand registered template generation versus HMM-ATR.**

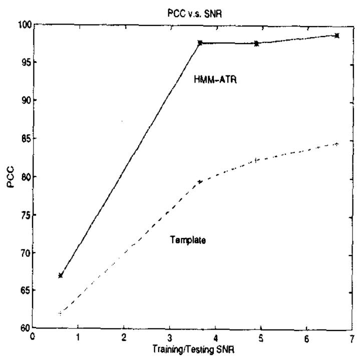


**Figure 7. Experiment II PCC results as SNR varied from 0.62 dB to 6.64 dB.**

results are shown in Figure 7. Clearly, the HMM-ATR outperforms the template technique over the range of conditions tested.

In a real ATR application, it is very likely that the training conditions are different from the testing conditions due to the unknown nature of the testing environment. Experiment III was designed to evaluate the effects of training at one SNR condition and testing at another. In Figure 8, we present results that are obtained by training the models at $SNR = 4.88dB$ and testing on images with other SNRs. The results show that HMM-ATR outperforms the template based method when there are mismatched conditions between training and testing.

## 4. CONCLUSION

The HMM-ATR system is able to tolerate variability introduced by various factors better than the template matching approach. The proposed system offers potential for advancing the state of the art in ATR.

Future work for HMM-ATR system will include a thorough investigation of applying HMM-ATR on an extensive dataset (MSTAR database), performance sensitivity on number of projection segments, and algorithm development/improvement.

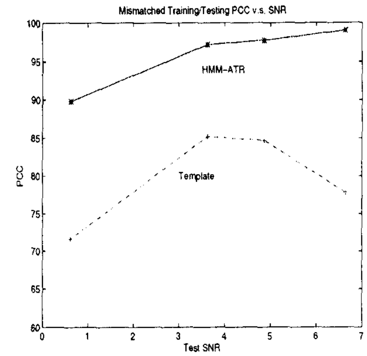


**Figure 8. Experiment III PCC: Mismatched training/testing conditions.**

The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies either expressed or implied, of the Army Research Laboratory or the U.S. Government.

## REFERENCES

[1] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society,* ser. B 48 no.3, pp. 259-302, 1986.

[2] B. Bhanu, D. E. Dudgean, E. G. Zelnio, A. Rosenfeld, D. Casasent, and I. S. Reed, *Special Issue on Automatic Target Recognition, IEEE Trans. IP,* vol. 6, 1997.

[3] W. Burger and B. Bhanu, "Signal-to-Symbol Conversion for Structural Object Recognition using Hidden Markov Models," *Proc. Image Understanding Workshop,* pp. 1287-1291, 1994.

[4] M.-Y. Chen, A. Kundu, and S. N. Srihari, "Variable Duration Hidden Markov Model and Morphological Segmentation for Handwritten Word Recognition," *IEEE Trans. IP,* vol. 4, pp. 1675-1688, 1995.

[5] D. E. Dudgean and R. T. Lacoss, *Special Issue on Automatic Target Recognition, The Lincoln Laboratory Journal,* vol. 6, no. 1, 1993.

[6] Y. He and A. Kundu, "Two Dimensional Shape Classification using Hidden Markov Models", *IEEE Trans. PAMI,* vol. 13, pp. 1172-1184, 1991.

[7] D. P. Kottke, J. K. Fwu, "Multi-level SAR Image Segmentation for Target Recognition", Advanced Sensors Federated Labs Consortium Quarterly Status Report, June 31, Army Research Labs, 1997.

[8] "MSTAR (Public) Targets: T-72, BMP-2, BTR-70, SLICY SAR Image Database", Wright Laboratory's Algorithm Integration Branch, `http://www.mbvlab.wpafb.af.mil/public/MBVDATA`.

[9] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. of the IEEE,* vol. 77, pp. 257-286, 1989.

[10] F. S. Samaria, "Face Recognition Using Hidden Markov Models," Ph.D. Dissertation, University of Cambridge, 1994.