Stochastic Models for Recognition of Articulated Objects

Bir Bhanu and Bing Tian
College of Engineering
University of California
Riverside, CA, 92521

Abstract

In this paper, we present a hidden Markov modeling (HMM) based approach for recognition of articulated objects in synthetic aperture radar (SAR) images. We develop multiple models for a given SAR image of an object and integrate these models synergistically using their probabilistic estimates for recognition and estimates of invariance of features as a result of articulation. The models are based on sequentialization of scattering centers extracted from SAR images. Experimental results are presented using 1440 training images and 2520 testing images for 4 classes.

1 Introduction

One of the critical problems for object recognition is that the recognition process be able to handle articulations of the object and spurious or noisy data. The recognition process working under articulation situations must be able to work with only portions of the *correct* spatial and amplitude information. Because of its stochastic nature, a hidden Markov model (HMM) is quite suitable for characterizing patterns. Its nondeterministic model structure makes it capable of collecting useful information from distorted or partially unreliable patterns. Many successful applications of HMM in speech recognition and character recognition attest to its usefulness.

However, the limit of traditional HMMs is that they are basically one dimensional models. So how to apply this approach to two dimensional image problems appropriately becomes the key. It remains largely an unsolved problem. In this paper we use features based on the image formation process to encode the 2-D image into 1-D sequences. We use information from both the relative positions of the scattering centers and their relative magnitude in SAR images. In this paper we address the fundamental issues of building object models and using them for recognition of articulated objects in SAR images.

Overview of the approach: Figure 1 provides an overview of the HMM based approach for recognition of articulated objects in SAR imagery. During an off-line phase, scattering centers are extracted in SAR images by finding local maxima of intensity. Both locations and magnitudes of these peak features are used in the approach. These features are viewed as *emitting patterns* of some hidden stochastic process. Multiple observation sequences based on both the relative geometry and relative amplitude of SAR return signal (obtained as a result of the physics of the SAR image formation process) are used to build the bank of stochastic models. At the end of the off-line phase, hidden Markov recognition models for various objects and azimuths are obtained. Similar to the off-line phase, during on-line phase features are extracted from SAR images and observation sequences based on these features are matched by the HMM forward process with the stored models obtained previously. Maximum likelihood decision is made on the classification results. Now the results obtained from multiple models are combined in a voting kind of approach that uses the object, azimuth label and its probability of classification and estimates of invariance of features as a result of articulation. This produces a rank ordered list of classifications of the test image and associated confidences.

Related work and our contribution: There is no published work on object recognition using HMM models. Fielding and Ruck [2] have used HMM models for spatio-temporal pattern recognition to classify moving objects in image sequences. Rao and Mersereau [3] have attempted to merge HMM and deformable template approaches for image segmentation. Template matching is not suitable to recognize articulated objects. Beinglan and Wofson [4] and Hel-Or and Werman [5] recognize articulated objects by finding constraints around a particular feature such as a joint for a table lamp. Such features are extremely difficult to extract in SAR images. Bhanu and Jones [1] have used a hashing based approach that exploits articulation invariants for object recognition.

The original contributions of this paper are:

- Hidden Markov modeling approach commonly used for recognizing 1-D speech signals is applied in a novel manner to 2-D SAR images to solve the articulated object recognition problem.
- Multiple models derived from various observation sequences, based on both the geometry and signal amplitude are used to capture the unique characteristics of patterns to recognize objects.
- Extensive amounts of data (1440 training images and 2520 testing images) generated by the XPATCH SAR simulator is used to test the approach.

2 Technical Approach

Extraction of Scattering Centers: Scattering

0-8186-8183-7/97 $10.00 © 1997 IEEE

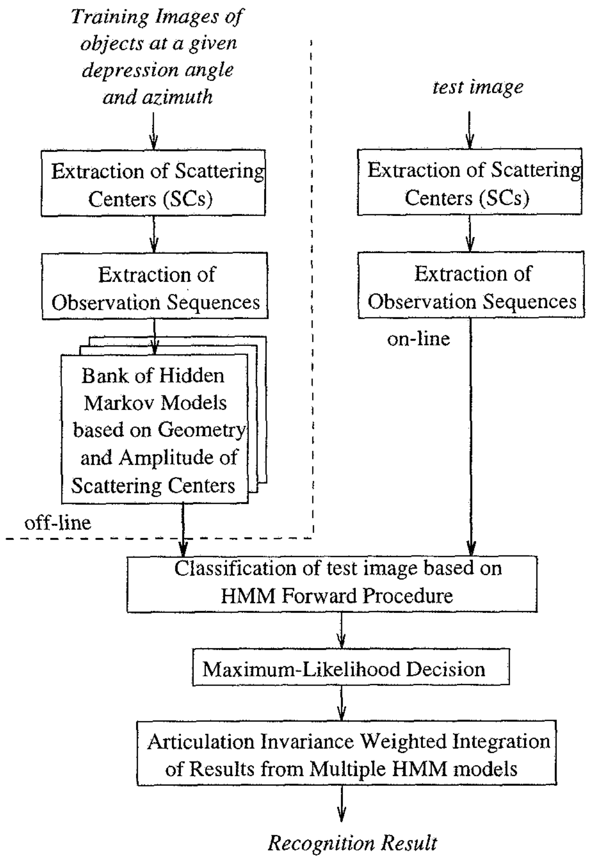

Figure 1: The HMM-Based approach for recognition of articulated objects.

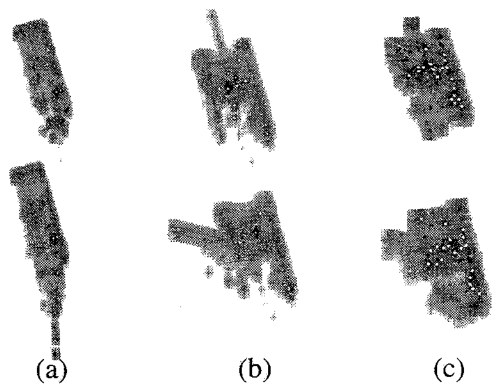

Figure 2: Examples of scattering centers (white dots) extracted from SAR images (15° azimuth) with and without articulation. (a) SCUD launcher with missile down and up, (b) T72 tank with turret at 0°, 60°, (c) T80 tank with turret at 0°, 60°.

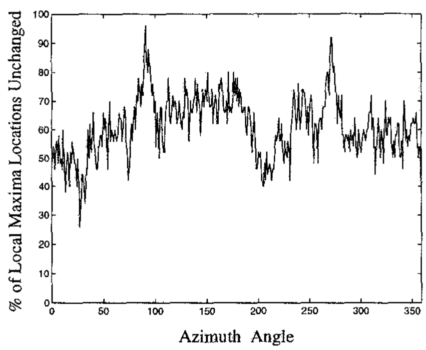

Figure 3: T72 tank articulation invariance with turret at 60°.

centers (location and magnitude) extracted from SAR images are used to train and test models for recognition. We consider a pixel as a scattering center if the magnitude of SAR return at this pixel is larger than all its eight neighbors. Figure 2 shows some examples of scattering centers extracted from SAR images (6″ resolution) of various objects at 15° depression angle and azimuths at 15°.

Rotation Variance: Unlike the visible images, SAR images are extremely sensitive to slight changes in viewpoint (azimuth and depression angle) and are not affected by scale. It has been found that the scattering centers for SAR images vary greatly with relatively small changes of azimuth angles. As a result we represent an object at a given depression angle by 360 azimuths taken in steps of 1°.

Articulation Invariants: Figure 3 shows the percentage of scattering centers that remain unchanged as a result of turret articulation (turret moving from 0° to 60°) for a T72 tank.

Hidden Markov Modeling Approach: It is well known that HMM can model speech signals well. Formally, a HMM is defined as a triple $\lambda = (A, B, \pi)$, where a_{ij} is the probability that state i transits to state j, $b_{ij}(k)$ is the probability that we observe symbol k in a transition from state i to state j, and π_i is the probability of i being the initial state.

Training Problem: To build a HMM is actually an optimization of the model parameters so that it can describe the observation better. This is a problem of training. The Baum-Welch re-estimation algorithm is used to calculate the maximum likelihood model [6]. The procedure for building the model base is to use the Baum-Welch algorithm to re-estimate the HMM parameters for each object and each azimuth angle for a given depression angle.

Testing Problem: During the testing phase we match each observation sequence against each model $(A, B, \Pi)_{(M_i^*, a_j^*)}$, in the model base, where M_i^* and a_j^* are ith object model and jth aspect of the object model. We use the Forward algorithm to compute the probability that a sequence is produced by a given model. The model that maximizes the probability of observation sequence is selected as the best match.

Forward Procedure: Suppose we have a HMM $\lambda = \{A, B, \pi\}$ and an observation sequence y_1^T. We define $\alpha_i(t)$ as the probability that the Markov process is in state i, having generated y_1^t.

$$\alpha_i(t) \;=\; 0, \text{ when t=0 and i is not an initial state.}$$

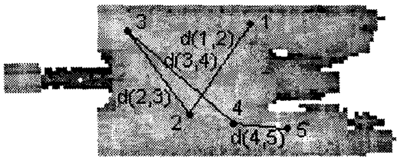

Figure 4: Example of an observation sequence superimposed on an image of T72 tank.

$$\alpha_i(t) = 1, \text{ when t=0 and i is an initial state.}$$
$$\alpha_i(t) = \Sigma_j[\alpha_j(t-1)a_{ji}b_{ji}(y_t)], \text{ when } t > 0.$$

The probability that the HMM stopped at the final state and generated y_1^T is $\alpha_{S_F}(T)$. After initialization of α, we compute it inductively. At each step the previously computed α is used, until the t reaches T. $\alpha_{S_F}(T)$ is the sum of probabilities of all paths of length T.

Extraction of Observation Sequences: After the scattering centers are extracted, we need to encode the data into a 1-D sequence as the input to a HMM process. We employ two approaches to obtain the sequences.

- Sequences based on relative amplitudes: $O_1 = \{Magnitude_1, Magnitude_2, ..., Magnitude_n\}$
- Sequences based on geometrical relationship: $O_2 = \{d(1,2), d(2,3), ..., d(n,1)\}$ $O_3 = \{d(1,2), d(1,3), ..., d(1,n)\}$ $O_4 = \{d(2,1), d(2,3), ..., d(2,n)\}$ $O_5 = \{d(3,1), d(3,2), ..., d(3,n)\}$

where $Magnitude_i$ is the amplitude of ith scattering center and $d(i,j)$ is the Euclidean distance between scattering centers i and j. Figure 4 gives an example to illustrate how we get the sequences. Sequence O_1 is obtained by sorting the scattering centers by their magnitude. We label the scattering centers 1 through n in descending order of their magnitude. Sequences O_2 through O_5 are obtained based on the relative locations of the scattering centers.

Invariance of Observation Symbols: Figure 5 shows the invariance in observation symbols by comparing the observation sequences O_1 extracted from the three sets of images (T72 tank with turret at 0°, 60° and 90°). Figure 5 (a1), (b1), and (c1) are obtained by counting the number of observation symbols in observation sequence of one image which are the same as its corresponding one in observation sequence of another image. Figure 5 (a2), (b2), and (c2) are obtained by counting the sum of differences between observation symbols in observation sequence of one image and its corresponding one in observation sequence of another image.

3 Experiments

Data: Using the XPATCH SAR simulator, we have generated one set of SAR images of 4 objects (SCUD missile launcher, T72 tank, T80 tank and M1a1 tank) at 15° depression angle and turret at 0° at each of the azimuth angles from 0° to 359°. We extract the 20 scattering centers (local maxima) with largest magnitude and use this data for training. We have also generated the data for the SCUD launcher with missile up, and the T72, T80 and M1a1 tank with turret rotated by 60° and 90° at 15° depression angle and each azimuth from 0° to 359°. This data, consisting of 2520 images, is used as the test data.

Table 1: Recognition results for articulated objects.

Training T72	Testing T72	performance (top 5)
Turret at 0° and 60°, 720 images	Turret at 90° 360 images	95.2
Turret at 0° and 90°, 720 images	Turret at 60° 360 images	94.0
Turret at 60° and 90°, 720 images	Turret at 0° 360 images	94.6
Average Performance		94.6

Table 2: Recognition rate % on 2520 testing cases.

	O_1	O_2	O_3	O_4	O_5	Integ.
Top 5	91.4	91.8	89.4	88.8	88.2	94.8
Top 10	97.3	97.8	96.1	96.5	95.6	99.2

Training: We perform experiments to choose the optimum of number of states and number of symbols of the HMM. We find 5 states and 24 symbols as the optimal number of states and symbols. For each of the 1440 images (4 classes × 360 azimuths) we build HMM models. These models are now put in the database and used during run-time recognition.

Testing: We have carried out 2 sets of experiments: (a) during testing we used two sets out of three sets of images as training data to train the HMM models, and tested the HMM models on the other set. Table 1 shows results for the T72 tank. These experimental results are obtained by using observation sequence O_1 only. These results can be improved by integrating the results from sequences O_2 through O_5 by following the approach given below for integration. (b) In the other experiment during the testing phase, each of the 2520 testing images is tested against all models in the database (1440 models). For each image the matching results are sorted in descending order of their probabilities. If the model with the probability in the top X ($X = 5, 10$, see Table 2) is the model (object) which produced a particular sequence (O_i), we count it as one correct recognition, otherwise we count it as one incorrect recognition. Note that sequence O_2 produces the best results.

Integration of results from multiple sequences: Since not all models based on various sequences for a particular object and azimuth will provide optimal recognition performance under occlusion, noise, etc., we improve the recognition performance by combining the results obtained from all five kinds of models.

We have developed a histogram-like method for integration. In this method for each test image we collect the ten highest possibilities in the testing results.

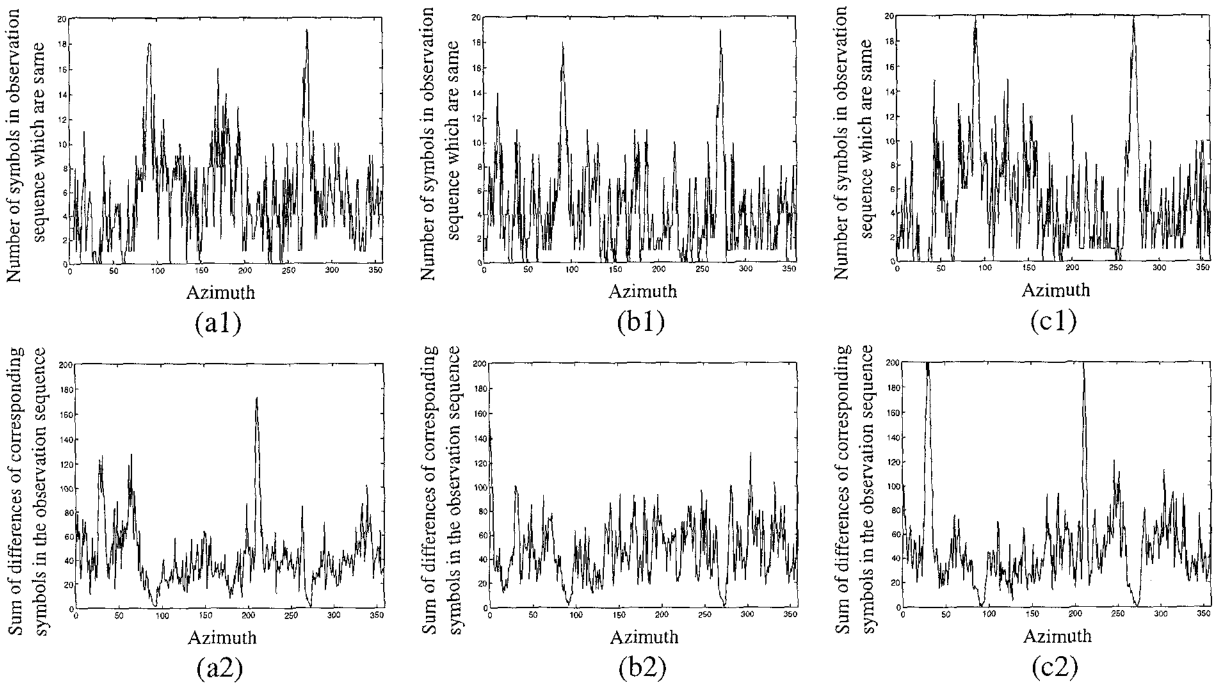

Figure 5: Comparison of observation sequence O_1 extracted from three sets of images for T72 tank. (a1, a2) 0° vs. 60°, (b1, b2) 0° vs. 90°, (c1, c2) 60° vs. 90°.

A normalization is done to the ten probabilistic estimates. Also for each test image, we get ten normalized values corresponding to each of the sequences $O_1, O_2, ..., O_5$. So we have 50 normalized numbers for each test image. Because each number corresponds to an object and a pose (the number is the probability that the test image is the image of that object at that pose), we draw a histogram with probability vs. object and pose. Now, this histogram is weighted by the average articulation invariance for cases of (0°, 60°), (0°, 90°), (60°, 90°) articulation. Articulation invariance is defined as the number of features that remain in exactly the same position as a result of articulation. If the object associated with the highest probability in the histogram is the same as the groundtruth, we count it as one correct recognition. The recognition results obtained using the integration algorithm are shown in Table 2. Note that integration results are better than any of the individual set of models. If we consider recognition in the top X ($X = 5, 10$) the results are in 90% range. If we consider top 1 only, then the recognition performance drops to 71.9%.

4 Conclusions

We have presented a novel conceptual approach for the recognition of articulated objects in SAR images. The number of observation sequences and the number of features are design parameters which can be optimized by following the approach presented in the paper.

Acknowledgements: This work was supported in part by DARPA grant MDA927-93-1-0010 and DARPA/AFOSR grant F49620-97-1-0184. The contents do not necessarily reflect the position or the policy of the U.S. governement.

References

[1] B. Bhanu and G. Jones, "Characterization of a model-based SAR target recognition system using invariants", *Proc. SPIE 3070*, Orlando, FL, April 21-24, 1997.

[2] K. H. Fielding and D. W. Ruck, "Spatio-temporal pattern recognition using hidden Markov models", *IEEE Trans. AES*, Vol. 31(4), pp. 1292-1300, 1995.

[3] R. R. Rao and R. M. Mersereau, "On merging hidden Markov models with deformable templates", *Proc. IEEE ICIP*, pp. 556-559, 1995.

[4] A. Beinglan and H. Wofson, "Articulated object recognition, or: how to generalize the generalized Hough transform", *Proc. IEEE Conf. on CVPR*, pp. 461-466, June, 1991.

[5] Y. Hel-Or and M. Werman, "Recognition and localization of articulated objects", *Proc. IEEE Workshop on Non-rigid and Articulated Objects*, pp. 116-123, Austin, Tx, Nov. 11-12, 1994.

[6] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. of the IEEE*, Vol. 77(2), pp. 257-285, 1989.