Recognition of Occluded Objects in SAR Images

Bir Bhanu, Yingqiang Lin and Bing Tian
Center for Research in Intelligent Systems
University of California, Riverside, CA 92521
Email: bhanu@cris.ucr.edu

Abstract

Recognition of occluded objects in synthetic aperture radar (SAR) images is a significant problem for automatic target recognition. Stochastic models provide some attractive features for pattern matching and recognition under partial occlusion and noise. In this paper, we present a discrete hidden Markov modeling (HMM) based approach for recognizing objects in synthetic aperture radar (SAR) images. We identify the peculiar characteristics of SAR sensors and using these characteristics we develop feature based multiple models for a given SAR image of an object. The models exploiting the relative geometry of feature locations *or* the amplitude of SAR radar return are based on sequentialization of scattering centers extracted from SAR images. In order to improve performance we integrate these models synergistically using their probabilistic estimates for recognition of a particular target at a specific azimuth. Experimental results are presented using both synthetic and real SAR images.

1 Introduction

One of the critical problems for object recognition is that the recognition approach should be able to handle partial occlusion of an object and spurious or noisy data.[1–4] In most of the object recognition approaches, the spatial arrangement of structural information of an object is the crucial part with the most important information. Under partial occlusion situations the recognition process must be able to work with only portions of the *correct*

spatial information. Rigid template matching and shape-based recognition approaches depend on good prior segmentation results. The structural primitive (e.g., line segments, point-like features, etc.) extracted from occluded and noisy images, however, may not have sufficient reliability, which will directly undermine the performance of those recognition approaches.

We want to suggest an object recognition mechanism that effectively makes use of all available structural information. Based on the nature of the problems caused by occlusion and noise, we view the spatial arrangement of structural information as a whole rather than view the spatial primitives individually. 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[5–7] and character recognition[8,9] attest to its usefulness. Thus, it is potentially an effective tool to recognize objects with partial occlusion and noise.

However, the limit of traditional HMMs is that they are basically one dimensional models. So how to appropriately apply this approach to two dimensional image problems becomes the key. It has been largely an unsolved problem. In this paper we use the features based on the image formation process to encode the 2-D image into 1-D sequences. These 1-D sequences or strings can be formed in many different ways. HMM models built off-line are used for matching during testing. In this paper, we use information from *both* the relative positions of the scattering centers and their relative magnitude in SAR images to address the fundamental

issues of building object models and using them for robust recognition of objects in SAR images.

1.1 Overview of the approach

Figure 1 provides an overview of the HMM based approach for recognition of occluded objects in SAR imagery. During an off-line phase, scattering centers are extracted from training SAR images by finding local maxima of intensity. Both locations and magnitudes of these peak features, viewed as *emitting patterns* of some hidden stochastic process, are used to generate multiple observation sequences. These multiple observation sequences, based on both the *relative* geometry and 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. From each training SAR image containing a particular object under a particular azimuth angle, five kinds of observation sequences are extracted to build five HMM models with one model taking one kind of observation sequence as input. These models provide robust recognition in the presence of occlusion and unstable features caused by scintillation phenomena where some of the features may appear/disappear at random in an image. 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 the 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. Then a maximum likelihood decision is made on the classification results. Finally the results obtained from multiple models are combined

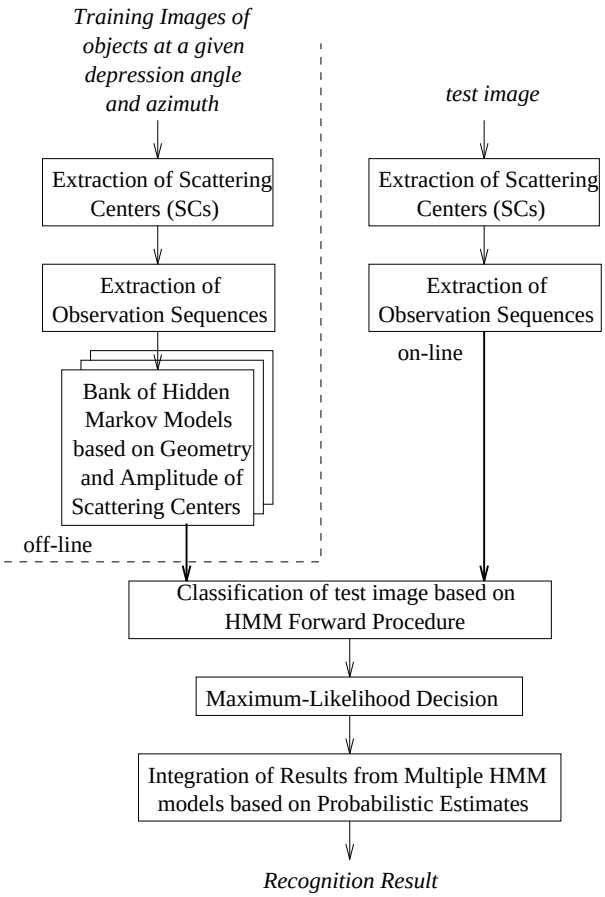


Figure 1: The HMM-Based approach for recognition of occluded objects.

in a voting kind of approach that uses both the object, azimuth label and its probability of classification. This produces a rank ordered list of classifications of the test image and associated confidences.

1.2 Related work and our contribution

Fielding and Ruck[10] have used HMM models for spatio-temporal pattern recognition to classify moving objects in image sequences. Rao and Mersereau[11] have attempted to merge HMM and deformable template approaches for image segmentation. Kottle et al.[12] propose a HMM-Based SAR automatic target recognition system. They first segment a SAR image,

and then extract features followed by Radon transforms. The feature sequences so obtained are input to HMMs. Template matching[13] and major axis based approaches[14] have been used to recognize and index objects in SAR images, however, they are not suitable to recognize occluded objects. This approach to recognizing occluded objects in SAR images uses significantly fewer models (typically 5) of an object at some aspect than our geometric hashing approach[15] (which uses 20 - 40 models).

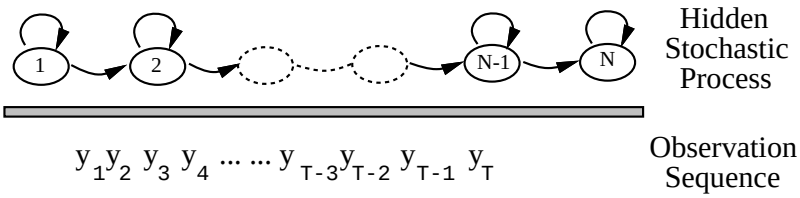
The contributions of this paper are:

- 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 occluded object recognition problem.

- Unlike most of the work for model building in pattern recognition and computer vision, our recognition models using hidden Markov modeling concepts are based on the peculiar characteristics of SAR images where the number of models used for recognition is justified by the quantification of the azimuthal variance in SAR images.

- Multiple models derived from various observation sequences, based on both the relative geometry and signal amplitude (four based on geometry and one based on amplitude) are used to capture the unique characteristics of patterns to recognize objects.

- Extensive amounts of data are used to test the approach for recognition of objects with various amounts of occlusion for both synthetic and real data.



N: the number of states.

M: the number of distinct observable symbols.

A: a_{ij} is the probability that state i will transit to state j.

B: $b_{ij}(k)$ is the probability that symbol k will be observed when there is a transition from state i to state j.

Π : π_i is the probability that state i is the initial state.

Figure 2: A N states forward-type HMM.

2 Hidden Markov Modeling Approach

It is well known that HMM can model speech signals well.[5–7] It is a model used to describe a doubly stochastic process which has a set of states, a set of output symbols and a set of transitions. Each transition is from state to state and associated with it are a probability and an output symbol. The word 'hidden' means that although we observe an output symbol, we cannot determine which transition has actually taken place. At each time step t, the state of the HMM will change according to a transition probability distribution which depends on the previous state and an observation y_t is produced according to a probability distribution which depends on the current state.

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. Figure 2 shows an example of a N states HMM.

Recognition Problem – Forward Procedure: The HMM provides us a useful mechanism to solve the problems we face for robust object recognition. Given a model and a sequence of observations, the probability that the observed sequence was produced by the model can be computed by the forward procedure.[16] Suppose we have a HMM $\lambda = \{A, B, \pi\}$ and an observation sequence $y_1^T = \{y_1, y_2, ..., y_t, ..., y_T\}$. We define $\alpha_i(t)$ as the probability that the Markov process is in state i, having generated $y_1^t = \{y_1, y_2, ..., y_t\}$.

$$\begin{aligned} \alpha_i(t) &= 0, \text{ when t=0 and i is not an initial state.} \\ \alpha_i(t) &= 1, \text{ when t=0 and i is an initial state.} \\ \alpha_i(t) &= \Sigma_j[\alpha_j(t-1)a_{ji}b_{ji}(y_t)], \text{ when } t > 0. \end{aligned} \tag{1}$$

The probability that the HMM stopped at the final state and generated y_1^T is $\alpha_{S_F}(T)$. The forward procedure is given below:

Let T be the length of an observation sequence and N is the number of states in the HMM.

1. Initialize $\alpha_i(0)$, where i = 1, 2, ..., N.

2. Compute $\alpha_i(t)$ inductively (equation 1), where t = 1, 2, ..., T. At each step, the previously computed $\alpha_i(t-1)$ is used. Repeat this process until t reaches T.

3. Output $\alpha_{S_F}(T)$, where $\alpha_{S_F}(T)$ is the probability that the HMM stopped at the final state and generated the observation sequence.

Usually, α becomes too small to be represented in a computer after several iterations. We take the logarithm of the α value in the computation.

Training Problem − Baum-Welch Algorithm: 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. But before we use the Baum-Welch algorithm, we must introduce the counterpart of $\alpha_i(t); \beta_i(t)$, which is the probability that the Markov process is in state i and will generate $y_{t+1}^T = \{y_{t+1}, y_{t+2}, ..., y_T\}$.

$$\begin{aligned} \beta_i(t) &= 0, \text{ when t=T and i is not a final state.} \\ \beta_i(t) &= 1, \text{ when t=T and i is a final state.} \\ \beta_i(t) &= \Sigma_j[a_{ij}b_{ij}(y_{t+1})\beta_j(t+1)], \text{ when } 0 \leq t < T. \end{aligned} \tag{2}$$

The probability of being in state i at time t and state j at time $t+1$ given observation sequence y_1^T and the model λ is defined as follows:

$$\begin{aligned} \gamma_{ij}(t) &= P(X_t = i, X_{i+1} = j \mid y_1^T) \\ &= \frac{\alpha_i(t-1)a_{ij}b_{ij}(y_t)\beta_j(t)}{\alpha_{S_F}(T)} \end{aligned} \tag{3}$$

Now the expected number of transitions from state i to state j given y_1^T at any time is simply $\Sigma_{t=1}^T \gamma_{ij}(t)$ and the expected number of transitions from state i to any state at any time is $\Sigma_{t=1}^T \Sigma_k \gamma_{ik}(t)$. Then, given some initial parameters, we could recompute $\overline{a_{ij}}$, the probability of taking the transition from state i to state j as:

$$\overline{a_{ij}} = \frac{\Sigma_{t=1}^T \gamma_{ij}(t)}{\Sigma_{t=1}^T \Sigma_k \gamma_{ik}(t)} \tag{4}$$

Similarly, $\overline{b_{ij}(k)}$ can be re-estimated as the ratio between the frequency that symbol k is emitted and the frequency that any symbol is emitted:

$$\overline{b_{ij}(k)} = \frac{\Sigma_{t:y_t=k}\gamma_{ij}(t)}{\Sigma_{t=1}^{T}\gamma_{ij}(t)} \tag{5}$$

It can be proved that the above equations are guaranteed to increase $\alpha_{S_F}(T)$ until a critical point is reached, after which the re-estimate will remain the same. In practice, we set a threshold as the ending condition for re-estimation.

So the whole process of training a HMM is as follows:

1. Initially, we have only an observation sequence y_1^T and blindly set (A, B, π).

2. Use y_1^T and (A, B, π) to compute α and β (equations (1), (2)).

3. Use α and β to compute γ (equation (3)).

4. Use y_1^T, (A, B, π), α, β and γ to compute A and B (equations 4, 5). Go to *step* 2.

A HMM is able to handle pattern distortions and the uncertainty of the locally observed signals, because of its nondeterministic nature. However, a HMM is primarily suited for sequential, one-dimensional patterns and it is not obvious that how a HMM can be used on 2-D patterns in object recognition. The basic ideas to apply a HMM for our purpose are (a) training the HMM λ by samples of SAR images of a certain object, and (b) recognizing an unknown object in a given SAR image. These two problems are addressed in the following. The key questions are what we shall use as observation data and how we get the observation sequences.

3 Hidden Markov Models for SAR Object Recognition

3.1 SAR Imagery and Target Recognition

The differences between SAR and other imagery include the source of illumination, image resolution, imaging geometry, and noise. SAR imagery provides its own source of illumination. However, microwaves reflect off of objects in a scene differently (specular vs. diffuse reflections) than sunlight. The signal returned from an object depends upon its microwave reflectivity, as well as relative position, size, and texture. The reflected microwaves are measured to produce the image. One of the useful properties of SAR imagery is that the spatial resolution is independent of the distance. In the range direction, the time resolution determines slant range resolution, and both are dependent only on bandwidth. The azimuth resolution is dependent only on the length of the antenna in the azimuth direction. The ability to collect high-resolution images from extreme distances is a particularly attractive characteristic of SAR. Although there are many positive aspects to the properties of SAR, there are negative aspects as well. For example, the imaging geometry makes understanding SAR imagery and its formation unintuitive and unnatural. In addition, the interaction of microwaves on surfaces is not entirely understood. Another obstacle for SAR imagery applications is speckle which results from the interference of many scatterers within the same resolution cell. Although speckle can be considered as multiplicative noise, when the image is converted to a dB image speckle statistics become additive and techniques for additive noise removal, such as spatial averaging, can be applied to remove the noise. However, spatial averaging degrades resolution

which is important for many applications which use SAR imagery. Because of the fundamental differences between SAR and other types of imagery, algorithms developed for application to other imagery are not directly applicable to SAR.

The goal of automatic target recognition in SAR imagery is to recognize the identity (ID) of a target, its location and azimuth or pose. The challenge is to recognize occluded targets, articulated targets, and targets with different configurations.[2] In this paper, we focus on the occlusion problem and assume that the potential targets have already been extracted. Thus, we are given the region-of-interest (or chip) of a target and the task is to determine the ID and pose or only the ID of the target.

3.2 Extraction of Scattering Centers

Scattering centers (location and magnitude) extracted from SAR images are used to train and test models for recognition. At six to twelve inch resolution (See Figure 3), there exist a large number of peaks corresponding to scattering centers. We have selected peaks as features in this work since we wanted to evaluate the limits of our approach using six inch resolution synthetic SAR images (generated using the XPATCH code[17]) and real SAR images at one foot resolution from the MSTAR data[18] before more complicated features are used. 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 3 shows some examples of top 20 scattering centers (based on their magnitude) extracted from XPATCH generated SAR images of various objects at 15° depression angle and azimuths at 0°, 60°, and 90°. The squint angle, the angle between the

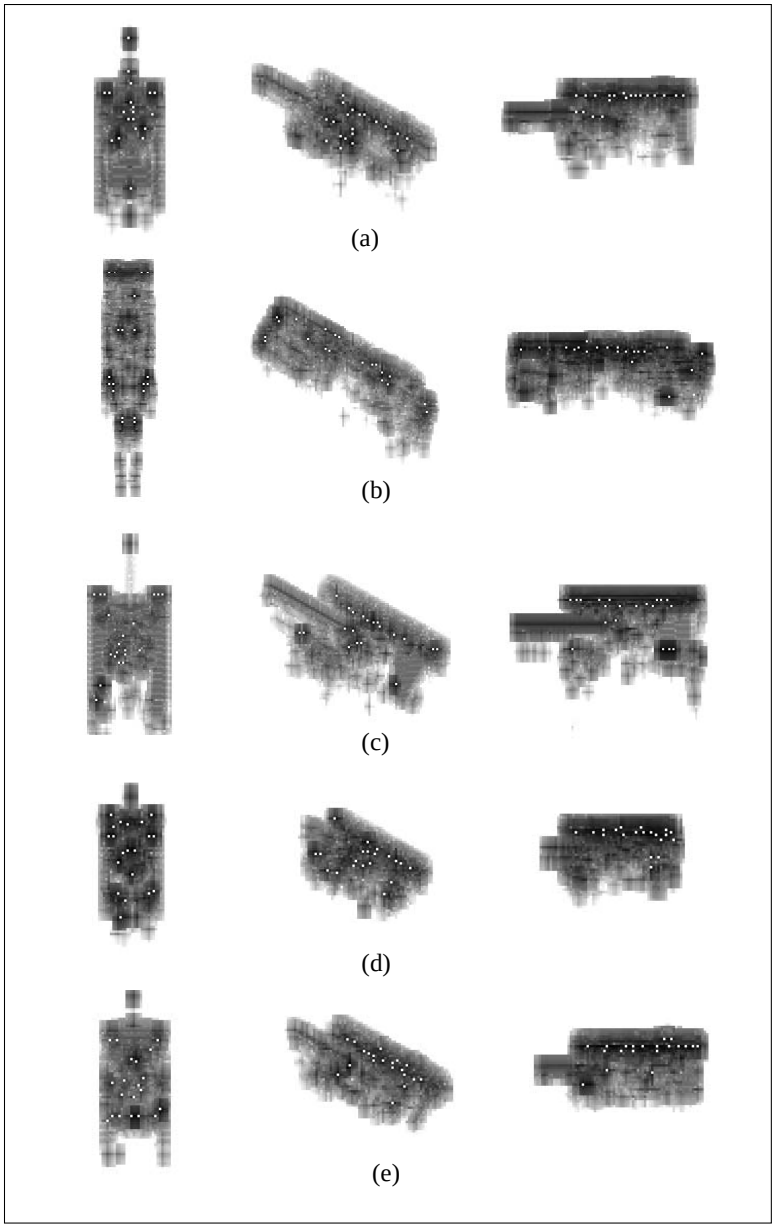
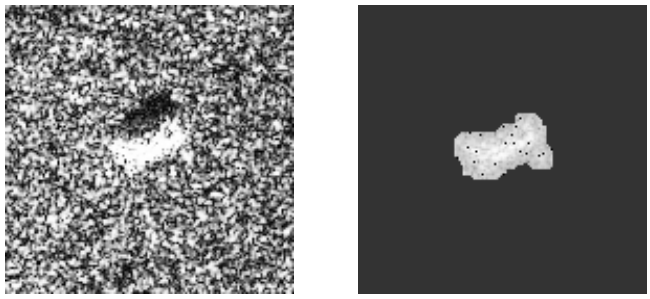


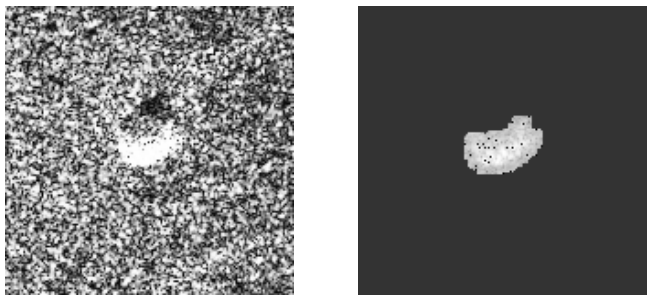
Figure 3: Examples of scattering centers (white dots) extracted from SAR images at azimuths $0°, 60°, 90°$. (a) Fred tank, (b) SCUD launcher with missile down, (c) T72 tank, (d) T80 tank, (e) M1a1 tank.

flight path and radar beam, is known (90° here) and kept fixed for all the images used in this paper. Figure 4 shows examples of scattering centers extracted for MSTAR SAR data. It is to be noted that in this paper we have used different number of scattering centers for different experiments for variety. It is possible to determine an optimal number of scattering centers for a given task.



(a) SAR image of a T72 tank

(b) Features extracted for T72 tank



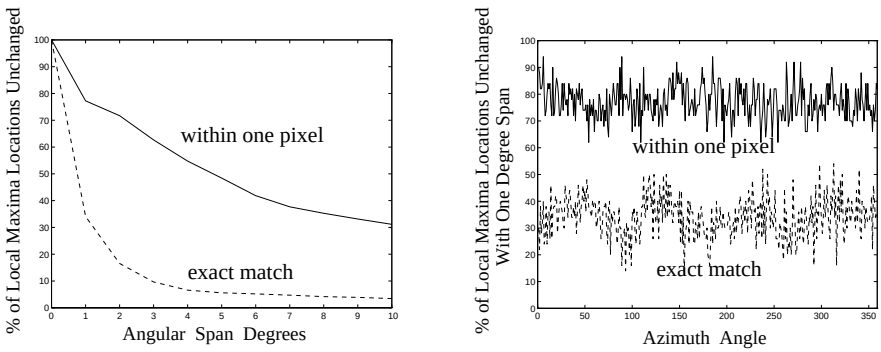
(c) SAR image of a ZSU gun

(d) Features extracted for ZSU gun

Figure 4: Real SAR images and region-of-interests (ROIs) (with peaks shown as black dots superimposed on the ROI) for T72 tank #a64 and ZSU anti-aircraftgun #d08.

3.3 Rotation Variance of Scattering Centers and Representation of 3-D Objects

In order to determine the number of azimuths needed to represent a 3-D object in a systematic manner, we evaluate the characteristics of SAR images. Unlike the visible images, SAR images



(a) (b)

Figure 5: XPATCH Data (a)T72 tank rotational invariance.(b)T72 tank rotational invariance with 1° Angular Span.

are extremely sensitive to slight changes in viewpoint (azimuth and depression angle) and are not affected by scale.[19] We evaluate[15] the characteristics of scattering centers to find out what kind of location invariance exists among scattering centers. Figure 5 shows the rotation invariance of the T72 tank with XPATCH data. The data is obtained by rotating the image at azimuth i° (for a fixed depression angle) by x° (angular span), and comparing the rotated image with the image of $(i + x)^\circ$ to see how many scattering centers do not change their location. Since the object chip is 256×256 pixels, we rotate the image with respect to the center point $(127.5, 127.5)$. The distance measurement criteria "exact match" and "within one pixel" are defined in the following:

$$\begin{cases} \mathbf{x_r}\ \textbf{exactly matches}\ \mathbf{x} \text{ if } MAX(|x - x_r|, |y - y_r|) < \frac{1}{2} \text{ pixel} \\ \mathbf{x_r}\ \textbf{and}\ \mathbf{x}\ \textbf{are within one pixel} \text{ if } MAX(|x - x_r|, |y - y_r|) < 1\frac{1}{2} \text{ pixel} \end{cases}$$

Figure 5(a) shows the average (over all the 360 azimuth angles) results for the number of scattering centers that remain unchanged within some error tolerance (as defined in the above) for T72 tank images for angular span of 1° to 10°. The 50 strongest scattering centers are used for each image. Note that for a 1° change in azimuth (angular span), approximately 1/3

and 3/4 scattering centers remain unchanged for the exact match case and within one pixel case, respectively. Figure 5(b) gives the percentage of scattering center locations unchanged vs. azimuth angle, for each of the 360 azimuths, with 1° angular span for the exact match and within one pixel match. For example, after rotating the image at azimuth 100° by 1° and comparing the rotated image with the image at azimuth 101° using "exact match" criterion, we get the point whose horizontal coordinate is 100 on the "exact match" curve in Figure 5(b). The results given in Figure 5 show that scattering centers for XPATCH SAR images for all azimuths vary greatly with relatively small changes of angles. Similar results are obtained with real MSTAR SAR data at one foot resolution.[20] Therefore, to recognize occluded objects, we need to get SAR images of an object at a given depression angle under various finely sampled azimuth angles. It is possible to collect SAR images which are 1° apart in azimuth. Thus, under ideal conditions, we can have 360 SAR images of an object with one image corresponding to one and only one azimuth angle between 0° and 359°. Consequently, we treat an object under different azimuth angles separately and build a model for each azimuth angle whenever possible. Note that when building models for an object azimuth, a single HMM model is inadequate because of noise, articulation, occlusion, etc. Therefore, to increase robustness, we build multiple HMM models for a given object at a specific azimuth with one model taking a particular kind of observation sequence as input, as discussed in the next section.

3.4 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 model based HMM process for recognition. It is one of the *key* factors which affects the performance of HMM approach for object recognition. There are many ways to choose sequences, but we want to use information from both the magnitude and the relative spatial location of the scattering centers extracted from a SAR image. Also the sequentialization method should not be affected by distortion, noise, or partial occlusion and should be able to represent the image efficiently. Based on the above considerations, we build five kinds of sequences.

- Sequences based on amplitudes: $O_1 = \{Magnitude_1, Magnitude_2, ..., Magnitude_n\}$,

  where $Magnitude_i$ is the amplitude of ith scattering center and $Magnitude_i \geq Magnitude_{i+1}$, $\forall\ i = 1, ..., n-1$.

- Sequences based on *relative* geometrical relationships:

  $O_2 = \{d(1,2), d(2,3), ..., d(n,1)\}$ $(length\ n)$

  $O_3 = \{d(1,2), d(1,3), ..., d(1,n)\}$ $(length\ n-1)$

  $O_4 = \{d(2,1), d(2,3), ..., d(2,n)\}$ $(length\ n-1)$

  $O_5 = \{d(3,1), d(3,2), ..., d(3,n)\}$ $(length\ n-1)$,

  where $d(i,j)$ is the Euclidean distance between scattering centers i and j. The scattering centers are sorted in descending order (1, ..., n) of their magnitude.

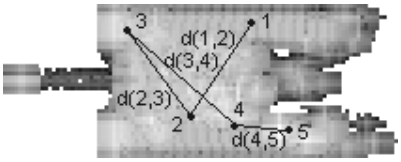


Figure 6: Example of an observation sequence superimposed on an image of T72 tank.

Figure 6 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. So in this approach, we do not use the location information and thus can avoid the instability caused by the error in localization of scattering centers. Sequences O_2 through O_5 are obtained based on the relative locations of the scattering centers and the magnitude of the scattering centers only affects the ordering. Note that they are not scale dependent, since the SAR images are scale independent.[19] In experiments described in section 4, we only consider the top 20 scattering centers for XPATCH data and the top 30 scattering centers for MSTAR data (sorted in descending order of their magnitude). This method of ordering is done for convenience, for repeatability and because we expect that the scattering centers with larger magnitude are generally more stable than the weaker ones.

Since we use discrete HMMs, each element in the sequence should be converted to an observation symbol. Labels from 1 to M represent the symbols which can be observed for a HMM. We use the K-means algorithm[21] to classify the magnitude values (or distance values) of all the scattering centers in the database into K classes (here K = M, it is experimentally determined, see Section 4.1.1). Once we know to which class each of the elements of a sequence belongs, we label the element with the label of its class. Thus, for a given sequence, we obtain

a sequence of observation symbols and call the resulting sequence as observation sequence..

3.5 Off-line Training and On-line Recognition Phases

The procedure for building a model based on a particular kind of observation sequence is described as follows:

1. Loop (for a given depression angle) lines 2-4 for each object and each azimuth angle.

2. Generate images which simulate occlusion and clutter with scattering centers occluded from different directions (see Section 4).

3. Extract one observation sequence from each image generated in (2), including the original image.

4. Run Baum-Welch algorithm on all the observation sequences generated in (3) to estimate the HMM parameters. First, randomly set the parameter values and run Baum-Welch algorithm on the first observation sequence (Use equations (1) to (5)). Then the parameter values resulting from the first observation sequence are used as the initial parameter values when the second observation sequence is input and so on. Exit the loop when there is no further change in parameter values.

The recognition procedure based on a particular kind of observation sequence is described as follows:

1. Extract the observation sequence from the testing image.

2. Loop line 3 for all the models in the model base.

3. Feed the observation sequence into the model, $(A, B, \Pi)_{(M_i^*, a_j^*)}$, where (M_i^*, a_j^*) represents the model of object i with azimuth angle j. Use Forward algorithm to compute the probability that this sequence is produced by this model.

4. The models with the H highest probabilities of generating the observation sequence are selected as the best possible matches to be used by the next integration step. The model with maximum probability is considered the best match.

3.6 Integration of Results from Multiple Sequences

Since a model based on a particular kind of observation sequence for a particular object and azimuth cannot provide optimal recognition performance under occlusion, noise, etc., we improve the recognition performance by combining the results obtained from multiple models corresponding to multiple kinds of observation sequences. Thus, when a testing image cannot be recognized correctly by a model based on a particular observation sequence, say O_1, it may be recognized correctly by models based on other kinds of observation sequences.

We have developed a voting-like method shown in Figure 7 to integrate the results from models based on five kinds of observation sequences. The algorithmic steps are:

(1) For each kind of observation sequence, we collect the H highest possibilities in the test results. Each possibility is the probability that the test image is the image of that object at that azimuth.

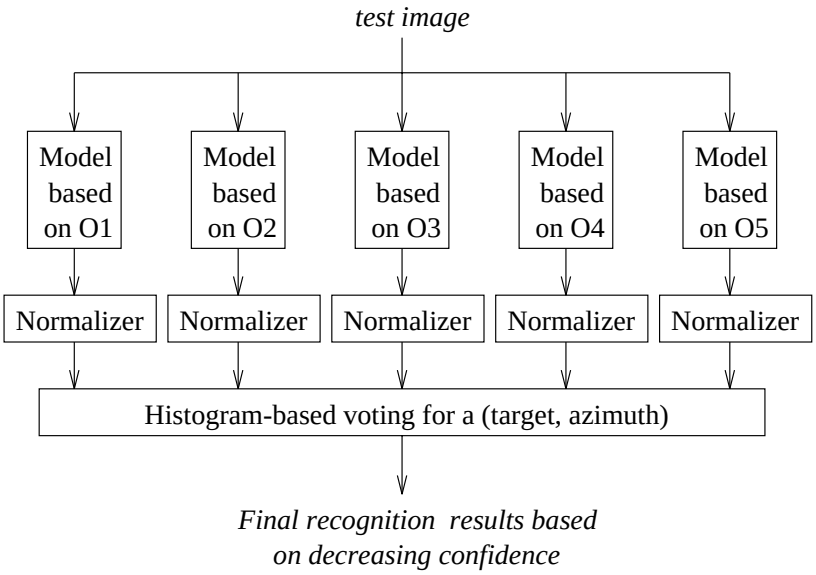


Figure 7: Integration of results by histogram-based method.

(2) A normalization is done to the H probabilistic estimates for each kind of observation sequence so that estimates from five kinds of observation sequences can be compared.

(3) We draw a histogram that sums the normalized probabilities of the various object/azimuth instances for the results obtained in step (2).

(4) If the object associated with the highest frequency in the histogram is the same as the ground truth, we count it as one correct recognition.

4 Experiments

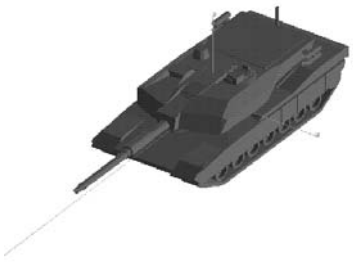
Both simulated and real-world data is used to test the performance of our approach.

4.1 XPATCH SAR Data

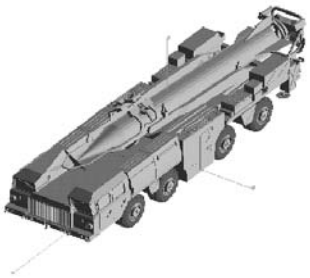
We use the XPATCH SAR simulator[17] to generate the data to perform controlled experiments. XPATCH is considered the best radar signature prediction code available; it has been extensively validated against primitive objects and complex full scale vehicles.

In our research, we selected six inch resolution data to get a significant number of "pixels-on-target" to facilitate solving the occluded target recognition problem. It is possible to collect real data at this resolution and also super resolution techniques[22] exist to achieve such data. We generate one set of SAR images of 5 objects (Fred tank, SCUD missile launcher, T72 tank, T80 tank and M1a1 tank, shown in Figure 8.) at 15° depression angle, and 90° squint angle (fixed), at each of the azimuth angles from 0° to 359°. We extract the 20 scattering centers (local maxima) with largest magnitudes. In the experiments, since we want to test the performance of our approach under partial occlusion and spurious data, we simulate realistic occlusion situations and generate images for training and testing.

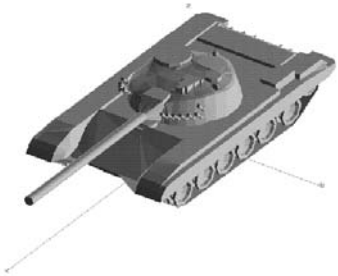
Simulating occlusion and clutter: There are no accepted empirical models of object occlusion in SAR imagery. So, we consider the occlusion to occur possibly from 9 different directions as shown in Figure 9. Scattering centers being occluded are not available. Moreover, we add some spurious data into the image to simulate noise or clutter. For instance, 20 scattering centers are shown in each image of Figure 9. They are obtained by removing 4 scattering centers (20% occlusion) from the center of one object or from one particular direction (simulated occlusion) and adding 4 spurious scattering centers (clutter or noise) into the image.



(a) Fred tank

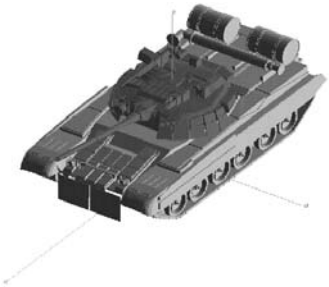


(b) SCUD missile launcher



(c) T72 tank

(d) T80 tank



(e) M1a1 tank

Figure 8: Targets.

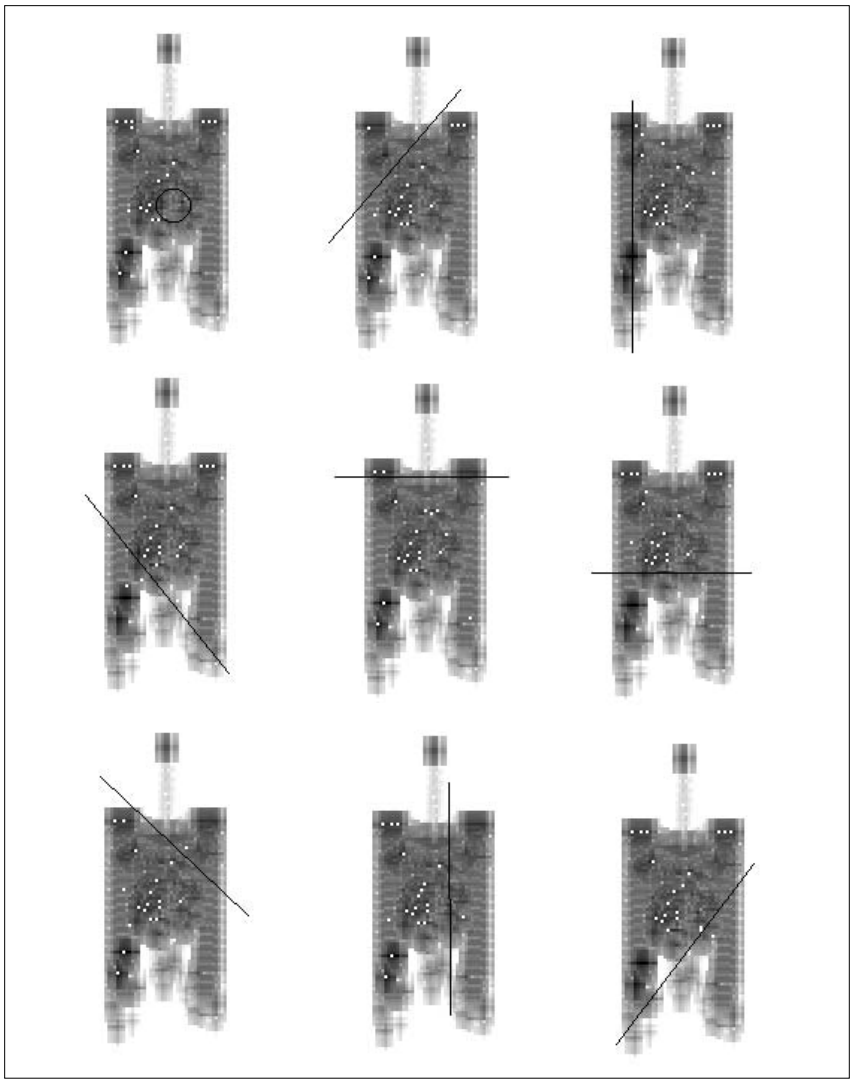


Figure 9: Scattering centers of T72 tank at azimuth 0°, part of scattering centers are occluded from a particular direction (0-8, left to right, top to bottom).

The spurious scattering centers are added based on the following rules:

- The location of the scattering center is generated as a pair of random numbers.
- The magnitude of the scattering center depends on a random number r between 1 and 50. We use the magnitude of the rth brightest scattering center as the magnitude of the spurious scattering center. Top 50 vs top 20 biases clutter to low values.

Training Data: Based on the method of simulating occlusion and clutter described above, we generate 90 images from the original image (10 samples for each of 9 directions) at 5% occlusion and another 90 images at 10% occlusion. Including the original image, we have 181 images per object per azimuth angle to train multiple HMM models. Thus, we have a total of $325,800$ (5 objects, 360 azimuths, 181 occluded images) samples for training.

Testing Data: We generate one image with o scattering centers occluded ($o = 2, 4, 6, 8$ or 10) from direction d ($d = 0, 1, ..., 8$) per azimuth angle per object. So there are 1800 images (5 objects $\times$ 360 degrees) generated for testing of occlusion with o scattering centers occluded from direction d. Thus, we have a total of $81,000$ (5 objects, 360 azimuths, 5 different occlusions $10\% - 50\%$, and 9 directions) samples for testing.

4.1.1 Training – Building Bank of HMM Models for Recognition

The results of experiments to choose the appropriate number of states and symbols of the HMM are shown in Table 1. We use data from five azimuth angles of five objects (Fred tank, SCUD missile launcher, T80 tank, T72 tank, and M1a1 tank). Table 1 shows that with an increase in the number of states and symbols, recognition performance increases. Considering both the recognition performance and the computation cost, we choose 8 states and 32 symbols as the "desired" number of states and symbols for our HMM models. Figure 10 illustrates an example of a 5 state, 4 symbol HMM.

Using the algorithm in Section 3.5 we built recognition models. For a selected sequence type we have 1800 (= 360 azimuths $\times$ 5 object classes) HMM models. Since we have defined

Table 1: Recognition rate of HMM with different number of states and symbols. N = # of states. M = # of symbols. R = Recognition rate % (top answer is correct). I = Indexing rate % (correct answer is in the top 5).

		id only		*id with pose*	
N	M	R	I	R	I
4	8	76.1	96.5	62.6	79.9
4	16	89.6	98.4	85.4	93.1
4	24	95.1	99.3	91.8	97.3
4	32	96.6	99.9	94.8	99.0
4	64	99.7	100.0	99.6	100.0
5	8	80.1	97.4	67.3	84.0
5	16	91.9	98.6	86.7	93.7
5	24	96.6	99.7	94.6	98.6
5	32	97.8	99.8	96.7	99.3
5	64	99.9	100.0	99.9	100.0
6	8	82.5	96.9	71.7	84.8
6	16	93.8	99.5	90.1	96.7
6	24	98.5	99.8	97	99.7
6	32	98.9	100.0	98.5	99.9
6	64	100.0	100.0	100.0	100.0
8	8	84.3	97.6	77.4	87.6
8	16	96.4	99.8	94.6	98.3
8	24	99.4	100.0	99	99.9
8	32	99.8	100.0	99.8	100.0
8	64	100.0	100.0	100.0	100.0
10	8	100.0	100.0	100.0	100.0
10	16	98.3	99.9	97.3	99.6
10	24	99.9	100.0	99.9	99.9
10	32	100.0	100.0	99.9	100.0
10	64	100.0	100.0	100.0	100.0

five kinds of observation sequences for each image $(O_1, O_2, O_3, O_4, O_5)$, we get models based on each kind of observation sequence.

4.1.2 Testing Results

During the testing phase, for a given observation sequence type each of the 81,000 testing images is tested against all models (1800 models: 5 objects, each has 360 models for each

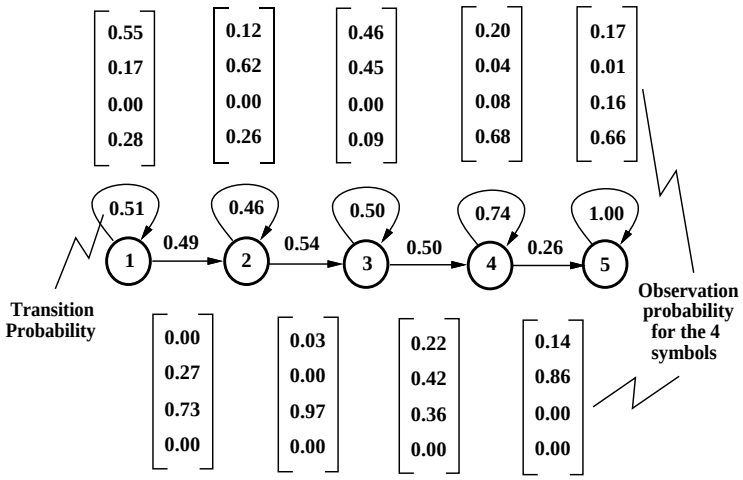


Figure 10: An example: parameters of a 5 states, 4 symbols HMM. The number on edges represents the transition probability, and the vector associated with each transition represents $b_{ij}(k)$. In our case, we use HMM with 8 states and 32 symbols.

azimuth angle). If the model with the maximum probability is associated with the object and azimuth angle from which the observation sequence is extracted, we count it as one correct recognition (object type and its pose). Otherwise, we count it as one incorrect recognition. After we get the results on all the images with a given percentage of occlusion (occluded from all 9 directions), we average these results and associate this recognition performance with the selected model for this particular percentage of occlusion.

Figure 11 shows the testing results taken directly from the output of each of the five kinds of observation sequences: $O_1, O_2, ..., O_5$ (section 3.4). The top curve, a dotted line, is the percentage that the test case object and pose is among the top ten recognition results, and the lower curve, in solid line, indicates the percentage that the recognition result with the highest probability is the same as the test case object and pose.

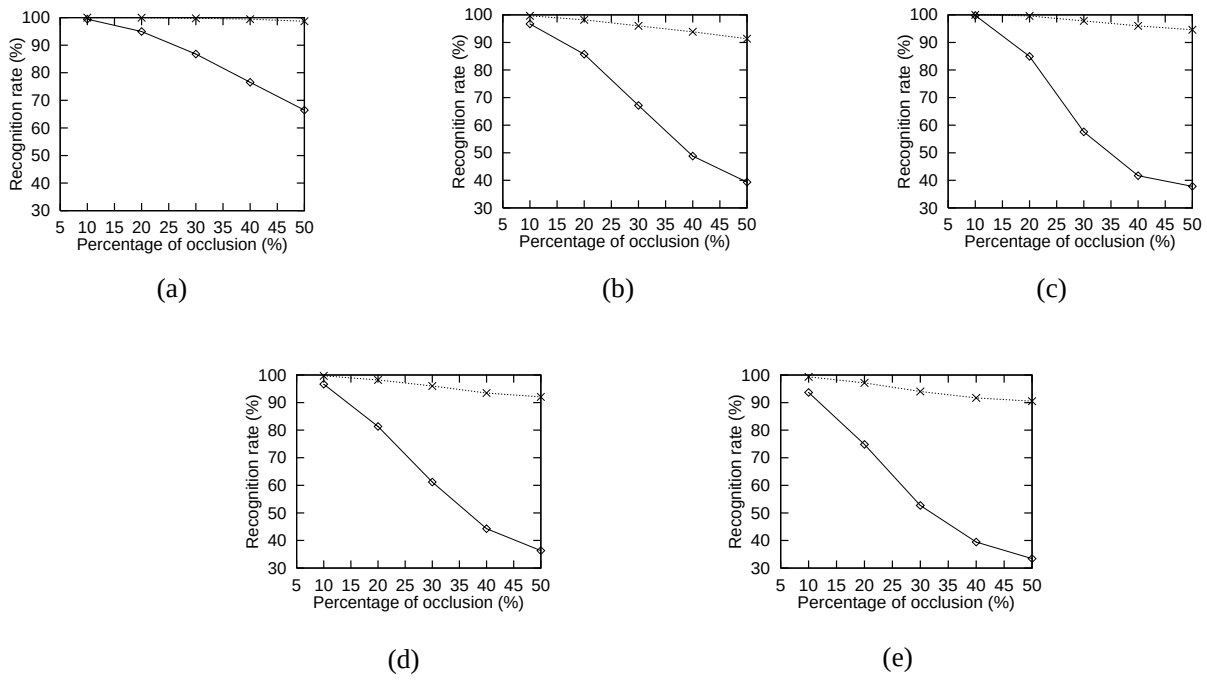


Figure 11: Recognition rate vs. percentage of occlusion for HMM models based on (a) O_1, (b) O_2, (c) O_3, (d) O_4, and (e) O_5.

4.1.3 Integration of Results from Multiple Observation Sequences

In order to understand the integration process better, we have done a set of testing experiments in which both geometric (O_2 to O_5) and magnitude (O_1) observation sequences are used. The results are shown in Table 2 where each row, for a given percentage of occlusion, shows the average results for occlusion that takes place from 9 different directions. The table shows how many incorrect recognitions, made by using models based on observation sequence O_2 can be correctly recognized ("captured") by models based on other observation sequences. The table demonstrates that each of these observation sequences add complimentary information. This implies the independent nature of these observation sequences and their results are integrated by means of a histogram based voting mechanism. We *define* the "upper bound"

Table 2: Testing results for occluded object recognition using 81,000 testing cases. Average results of 9 directions based on integration of O_1 to O_5.

Percent. occlusion	*Errors with model* O_2	*Errors Captured by models*				*Errors using models* O_1 *to* O_5	*% Correct Recognition ("upper bound")*	*% Based on Integrated Recognition*	*% Based on Integrated Indexing*
		O_1	O_3	O_4	O_5				
10%	4	0	1	0	1	2	100.0	99.9	99.9
20%	271	19	53	74	101	7	99.6	98.9	99.6
30%	763	111	294	339	418	25	98.6	93.4	97.6
40%	1050	265	580	629	675	79	95.6	79.4	91.8
50%	1119	397	726	755	784	148	91.8	62.2	83.3
Average Recognition Rate							97.1	86.8	94.4

as the highest possible recognition performance that can be achieved using the five observation sequences, considering *only* the top candidate. The percentage of correct recognition corresponding to "upper bound" is shown in the 8th column. The results shown in column 9 are obtained by integrating the results from observation sequences O_1 to O_5. The integrated recognition performance is 86.8%. Note that these results are for both object and azimuth (object/azimuth), that is, both the object identity and pose should be correctly recognized.

In the rest of this section, we demonstrate the results using 5 kinds of observation sequences (O_1 to O_5). We draw two curves (Figure 12(a)) to show the "upper bound" and "lower bound" of recognition rate. The curve on the top is obtained by considering all 5 kinds of observation sequences, if a model based on one of them correctly recognizes the test data, we count it as a correct recognition. The "lower bound" or the bottom curve is the worst recognition result out of the five models.

The second curve from the bottom in Figure 12(b) is the result for integrated recognition.

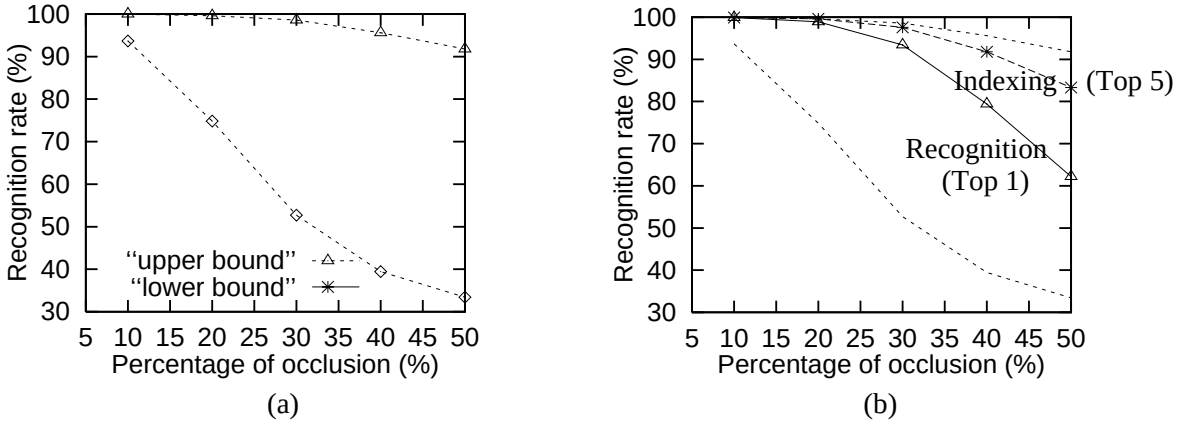


Figure 12: (a) "Upper" and "lower" bound of recognition rate vs. percentage of occlusion. (b)Performance of integrated models: using integrated models O_1 to O_5. The results for recognition (Top 1) and indexing (Top 5 candidates) are superimposed on the figure shown in (a).

The corresponding confusion matrix for various amounts of occlusion is shown in Table 3. On the average, we find 86.8% correct recognition performance when the objects are occluded from $10-50\%$. The second curve from the top in Figure 12(b) is obtained by counting a correct indexing result when the ground truth object/azimuth is in the objects/azimuth associated with the highest 5 probabilities in the histogram. For the purpose of comparison, we have also superimposed the curves of Figure 12(a) into Figure 12(b) with "lower/upper" bounds. Considering the correct indexing answer in the top 5 responses, the average performance is 94.4% for 5 objects occluded from $10\% - 50\%$. The difference between top 5 and top 1 recognition performance is 7.6%.

4.2 Real SAR Data

The methods used here are the same as those used in the previous subsection. The only difference is that here we experiment on real data instead of XPATCH data. We use MSTAR

Table 3: Confusion Matrix for 5 object classes at varying amounts of occlusion ($10\% - 50\%$).

	% Occlusion	*Fred*	*SCUD*	*T72*	*T80*	*M1a1*
Fred	10	100.0	0.0	0.0	0.0	0.0
	20	99.2	0.0	0.1	0.4	0.3
	30	95.9	0.2	0.6	1.9	1.4
	40	87.1	0.7	2.8	5.5	3.9
	50	73.2	1.6	7.1	12.1	6.0
SCUD	10	0.0	100.0	0.0	0.0	0.0
	20	0.0	99.7	0.2	0.1	0.0
	30	0.9	97.3	1.2	0.4	0.3
	40	3.1	88.8	4.9	1.9	1.3
	50	5.6	77.9	11.9	2.7	1.9
T72	10	0.0	0.0	100.0	0.0	0.0
	20	0.4	0.2	99.2	0.1	0.2
	30	2.4	0.5	95.3	1.1	0.6
	40	9.1	2.1	82.5	3.8	2.4
	50	16.8	5.2	65.9	6.8	5.4
T80	10	0.0	0.0	0.0	100.0	0.0
	20	1.2	0.0	0.1	98.6	0.1
	30	6.9	0.0	0.6	91.1	1.4
	40	21.5	0.1	1.6	72.6	4.2
	50	37.4	0.8	3.1	50.9	7.8
M1a1	10	0.0	0.0	0.0	0.0	100.0
	20	1.6	0.0	0.1	0.3	98.0
	30	8.5	0.2	0.7	2.9	87.8
	40	22.5	0.8	2.0	8.5	66.1
	50	36.9	1.1	5.2	13.8	42.9

public real SAR images (at one foot resolution and depression angle 15°) of 2 objects (T72 tank with serial number #a64 and ZSU anti-aircraft gun with serial number #d08). Ideally, we should have 360 object models for each azimuth for each object. However, we don't have 360 SAR images for each object in the MSTAR data set. For the T72 tank, there are 288 images available for different azimuths. Also for the ZSU gun, 288 images are available. Thus, each object consists of 288 azimuths (or aspects) which we call object models. Each object model consists of HMM models based on observation sequences (O_1 to O_5). We extract 30 scattering centers with the largest magnitudes from each SAR image. Figure 4 shows some

examples of SAR imagery, regions-of-interest and scattering centers superimposed on the SAR image. ROIs are obtained using an automatic thresholding and dilation/erosion process.[20]

We consider the occlusion to occur possibly from 9 different directions (center, 4 sides and 4 corners of the image). Scattering centers being occluded are not available. Moreover, we add back into the image at random locations a number of spurious scattering centers, equal to the number of occluded scatterers, of random magnitude. The random magnitude could be equal to the magnitude of any of the top 30 scatterers. This simulates noise or clutter from other objects. For example, for 30% occlusion, we remove 9 scattering centers from the center of one object or from one particular direction and add randomly 9 spurious scattering centers back into the image with magnitude randomly determined as stated in the above. We compute the observation sequences based on the scattering centers available after the occlusion process has taken place. Note that we assume there is only one object present in one image. However, the technique we use to synthetically generate occluded data adds an equivalent number of clutter features which could be viewed as a surrogate for multiple objects. No real SAR data with multiple objects exist in the public domain.

4.2.1 Training and Testing Data

Training Data: We generate 91 training images from each SAR image. The first one is the original SAR image without occlusion. Then we occlude the SAR image from 9 directions. For each direction, the occlusion level is 5% and 10%. For each occlusion level, we extract 5 sample images. So 91 images are obtained from each original testing SAR image. We have

Table 4: Confusion matrix for various occlusions using MSTAR data.

Target Type	Occlusion Level			
	20%		30%	
	T72	ZSU	T72	ZSU
T72	280(97.2%)	8(2.8%)	233(80.9%)	55(19.1%)
ZSU	5(1.7%)	283(98.3%)	64(22.2%)	224(77.8%)
Target Type	Occlusion Level			
	40%		50%	
	T72	ZSU	T72	ZSU
T72	200(69.4%)	88(30.6%)	164(56.9%)	124(43.1%)
ZSU	98(34.0%)	190(66.0%)	117(40.6%)	171(59.4%)

two objects and 288 SAR images of each object, thus, the number of training images is 52,416. Since there are 5 kinds of observation sequences that are extracted from each image, the total number of observation sequences is 262,080.

Testing Data: From each SAR image, we generate 36 testing images. We occlude the SAR image from 9 directions. For each direction, the occlusion level is from 20% to 50% with 10% increments. That is, the numbers of occluded scattering centers are 6, 9, 12, and 15 respectively. Thus, we have total 20736 ($2 \times 288 \times 36$) testing images and corresponding to each image, there are 5 types of observation sequences.

4.2.2 Experimental Results

• **Training:** For consistency, we use the same 8 states and 32 symbols used for XPATCH data for our HMM models for the real SAR data. Using the algorithm presented in Section 3.5, we build recognition models using the real SAR training data described in Section 4.2.1. For one observation sequence type we have 576 (288 azimuths $\times$ 2 object classes) HMM models.

Table 5: Recognition results for various occlusions using MSTAR data.

Target Type	Occlusion Level					
	20%			30%		
	correct	error	rejection	correct	error	rejection
T72	279(96.9%)	7(2.4%)	2(0.7%)	230(79.9%)	47(16.3%)	11(3.8%)
ZSU	281(97.6%)	4(1.4%)	3(1.0%)	216(75.0%)	59(20.5%)	13(4.5%)
Target Type	Occlusion Level					
	40%			50%		
	correct	error	rejection	correct	error	rejection
T72	188(65.3%)	75(26.0%)	25(8.7%)	156(54.2%)	110(38.2%)	22(7.6%)
ZSU	168(58.3%)	91(31.6%)	29(10.1%)	164(56.9%)	102(35.5%)	22(7.6%)

Since we have defined five kinds of observation sequences for each image $(O_1, O_2, O_3, O_4, O_5)$, we get models based on each kind of observation sequence.

• **Testing Results:** During the testing phase, we apply the recognition algorithm (described in Section 3.6) to the real SAR test data (described in Section 4.2.1). Here, we consider only the kind of an object, which we call ID. We count a recognition result as a correct recognition if the HMM model with the maximum probability is associated with the object from which the testing observation sequence was extracted, we do not consider the corresponding azimuth angles of the HMM models.

Tables 4 and 5 show the experimental results on MSTAR real data. These results are obtained by integrating the results from 5 different types of observation sequences O_1, O_2, O_3, O_4, and O_5. Table 4 shows the results for recognizing 20% to 50% occluded T72 tank and ZSU gun. The confusion matrix shows how many of them are correctly identified and how many are incorrectly recognized. Table 5 shows the results similar to Table 4. The difference

is that here we use "probability ratio threshold" instead of considering only the maximum probability. Only when the ratio between the maximum and second maximum (other object type) probabilities is greater than the probability ratio threshold (1.01 used here) we accept the recognition result. Otherwise, it is rejected and the test data is labeled as unknown object.

We also performed an additional experiment to determine the optimal number of observation sequences. Table 6 shows the results when only observation sequences O_1 and O_2 are used in integration. The results of Table 6 based on O_1 and O_2 are slightly better than the results of Table 4 based on all the five observation sequences (O_1 to O_5). The reason is that the results from observation sequences O_3, O_4, and O_5 are less reliable than others. Each of these observation sequences measures distances from a specific scattering center, for example, O_3 measures distances of other scattering centers from (reference) scattering center 1, O_4 measures distances of other scattering centers from scattering center 2, and so on. There is a significant probability that these prominent scattering centers (used as a reference) get occluded at higher levels of occlusion. If the first scattering center is occluded, then the entire observation sequence (O_3, O_4 or O_5) is subject to error, which is reflected in the integration results. An alternative is to use all the scattering centers as reference points to build models to achieve increased performance at the expense of increased computation, not just the top three scattering centers that we used in observation sequences O_3, O_4, and O_5.

From Tables 4, 5, 6, we can see that the performance degrades as the occlusion level increases. These results show that reasonable recognition performance can be achieved when

the occlusion is as much as 30% - 40%. The time needed to recognize a test case is 6ms on a SUN Ultra 2.

Table 6. Results of Integration (Observation Sequences O_1 and O_2).

Occlusion Level	# of Correct Recognition				# of test sequences
	20%	30%	40%	50%	
T72 tank	284(98.6%)	249(86.5%)	208(72.2%)	186(64.6%)	288
ZSU gun	285(99.0%)	238(82.6%)	194(67.4%)	175(60.8%)	288

5 Conclusions

Recognition of occluded objects has been a significant problem for automatic target recognition. In this paper, we have presented a novel conceptual approach based on hidden Markov modeling for the recognition of occluded objects in SAR images. The approach uses multiple models for various observation sequences that are chosen based on the SAR image formation and account for both the geometry and magnitude of SAR image features. We have shown the results on both XPATCH and real SAR data. 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. We have demonstrated that the HMM approach makes use of available structural information to solve the problem caused by occlusion and noise. It takes the spatial arrangement of structural information as a whole and is able to collect useful information from distorted or partially unreliable patterns. Reasonable recognition can be achieved for up to 30% - 40% occlusion with both the simulated and real SAR data.

References

[1] B. Bhanu. Automatic target recognition: state of the art survey. *IEEE Trans. Aerospace and Electronic Systems*, 22:364–379, July 1986.

[2] B. Bhanu, D. E. Dndgeon, E. G. Zelnio, A. Rosenfeld, D. Casasent, and I. S. Reed. Introduction to the special issue on automatic target detection and recognition. *IEEE Trans. Image Processing*, 6(1):1–6, January 1997.

[3] B. Bhanu and T. L. Jones. Image understanding research for automatic target recognition. In *Proc. ARPA Image Understanding Workshop*, pages 249–254, January 26-29, 1992.

[4] D. E. Dndgeon and R. T. Lacoss. An overview of automatic target recognition. *The Lincoln Laboratory Journal*, 6(1):3–9, 1993.

[5] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. of the IEEE*, 77(2):257–285, Feb 1989.

[6] L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(1):4–16, Jan 1986.

[7] L. R. Rabiner, B. H. Juang, S. E. Levinson, and M. M. Sondhi. Recognition of isolated digits using hidden Markov models with continuous mixture densities. *AT&T Technical Journal*, 64(6):1211–1233, July-August 1985.

[8] O. E . Agazzi and S. S. Kuo. Hidden Markov model based optical character recognition in the presence of deterministic transformations. *Pattern Recognition*, 26(12):1813–1826, November 1993.

[9] O. E . Agazzi and S. S. Kuo. Pseudo two-dimensional hidden Markov models for document recognition. *AT&T Technical Journal*, 72(5):60–72, 1993.

[10] K. H. Fielding and D. W. Ruck. Spatio-temporal pattern recognition using hidden Markov models. *IEEE Trans. Aerospace and Electronic Systems*, 31(4):1292–1300, 1995.

[11] R. R. Rao and R. M. Mersereau. On merging hidden Markov models with deformable templates. In *Proc. International Conf. on Image Processing*, pages 556–559, 1995.

[12] D. P. Kottle, P. D. Fiore, K. L. Brown, and J. K. Fwu. A design for HMM-based SAR ATR. In *SPIE Conf. on Algorithms for Synthetic Aperture Radar Imagery V*, pages 541–551, April 1998.

[13] L. M. Novak, G. J. Owirka, and C. M. Netishen. Performance of a high-resolution polarimetric SAR automatic target recognition system. *The Lincoln Laboratory Journal*, 6(1):11–24, 1993.

[14] J. H. Yi, B. Bhanu, and M. Li. Target indexing in SAR images using scattering centers and Hausdorff distance. *Pattern Recognition Letters*, 17:1191–1198, September 1996.

[15] G. Jones III and B. Bhanu. Recognition of articulated and occluded objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):603–613, July 1999.

[16] K. F. Lee. *Automatic Speech Recognition - The development of the SPHINX system.* Kluwer Academoc Publishers, Boston, 1989.

[17] D. J. Andersh, S. W. Lee, H. Ling, and C. L. Yu. XPATCH: A high frequency electromagnetic scattering prediction code using shooting and bouncing ray. In *Proc. Ground Target Modeling and Validation Conference*, pages 498–507, August 1994.

[18] T. Ross, S. Worrell, V. Velten, J. Mossing, and M. Bryant. Standard SAR ATR evaluation experiments using the MSTAR public release data set. *SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery V*, 3370:566–573, April 1998.

[19] J. P. Fitch. *Synthetic Aperture Radar.* New York: Springer-Verlag, 1988.

[20] B. Bhanu and G. Jones III. Recognizing target variants and articulations in synthetic aperture radar images. *Optical Engineering*, 39(3):712–723, March 2000.

[21] J. T. Tou and R. C. Gonzalez. *Pattern Recognition Principles.* Addison-Wesley, London, 1974.

[22] L. Novak, G. Benitz, G. Owirka, and L. Bessette. ATR performance using enhanced resolution SAR. In *SPIE Conf. on Algorithms for Synthetic Aperture Radar Imagery*, April 10 - 12 1996.