Bounding SAR ATR Performance Based on Model Similarity

Michael Boshra and Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside, California 92521

ABSTRACT

Similarity between model targets plays a fundamental role in determining the performance of target recognition. We analyze the effect of model similarity on the performance of a vote-based approach for target recognition from SAR images. In such an approach, each model target is represented by a set of SAR views sampled at a variety of azimuth angles and a specific depression angle. Both model and data views are represented by locations of scattering centers, which are peak features. The model hypothesis (view of a specific target and associated location) corresponding to a given data view is chosen to be the one with the highest number of data-supported model features (votes). We address three issues in this paper. Firstly, we present a quantitative measure of the similarity between a pair of model views. Such a measure depends on the degree of structural overlap between the two views, and the amount of uncertainty. Secondly, we describe a similarity-based framework for predicting an upper bound on recognition performance in the presence of uncertainty, occlusion and clutter. Thirdly, we validate the proposed framework using MSTAR public data, which are obtained under different depression angles, configurations, and articulations.

Keywords: Model-based SAR ATR, model similarity, performance prediction, data uncertainty, occlusion and clutter, upper bound on performance, MSTAR public data

1. INTRODUCTION

Model-based automatic target recognition (ATR) involves comparing features extracted from scene data with those of targets in a model database. The data acquired by a sensor are corrupted by distortion factors, such as uncertainty (error in estimating feature locations and attributes), occlusion (missing features) and clutter (spurious features). Performance of recognition depends on these data factors, in addition to model factors such as number of model targets, and degree of similarity between them. The model-similarity factor plays a fundamental role in determining the recognition performance. Intuitively, a model target that is distinct from all other model ones can be correctly recognized, even in the presence of high levels of data distortion. On the other hand, if the target is similar to some other model targets, then there is a high chance of it being "confused" with one of the similar targets, even in the presence of moderate distortion levels. Accordingly, we can say that recognition performance is inversely proportional to the degree of model similarity.

In this paper, we consider the above-mentioned data and model factors within a framework for predicting performance of a vote-based approach for ATR using synthetic aperture radar (SAR) data. In such an approach, each model target is represented by a set of SAR views that sample its signature at a variety of azimuth angles and a specific depression angle. Recognition is performed by matching locations of scattering centers (peaks) extracted from data and model views. Each model hypothesis (view of a specific target and associated location) is evaluated based on the number of data-supported model features (votes). The hypothesis that has the maximum number of votes is selected as the valid one. We address the following three issues in this paper. Firstly, we present a quantitative measure of similarity between model views. Secondly, we use the proposed measure along with statistical data-distortion models to predict an upper bound on the probability of correct recognition (PCR) as a function of data-distortion rates (prediction of a tight lower bound on performance is reported elsewhere by Boshra and Bhanu). Thirdly, we validate the proposed similarity-based framework using MSTAR public data obtained under different depression angles, configurations and articulations.

The remainder of the paper is organized as follows. The next section reviews relevant efforts for predicting performance of SAR ATR, and highlights our contributions. Section 3 describes the proposed similarity measure in
Figure 1. An illustration of performance prediction approaches: (a) empirical, using real data, (b) empirical, using synthetic data, (c) formal.

detail. The performance-prediction framework is presented in Section 4. Its validity is demonstrated in Section 5 through comparing experimentally-determined PCR plots with predicted bounds. Finally, conclusions are drawn in Section 6.

2. RELEVANT RESEARCH AND OUR CONTRIBUTIONS

The approaches for ATR performance prediction can be classified into two categories:

- **Empirical Approaches**: Performance is determined through experimentation using either real data (as illustrated in Figure 1(a)), or synthetic data (e.g., Monte Carlo techniques, see Figure 1(b)). Examples of these approaches are: 1) Catlin et al. present a method to estimate the function relating the probability of correct identification to aspect and depression angles using SAR data. Each point on the function is determined experimentally through passing a number of real images to an ATR system. The resulting function is modeled using either linear regression or an artificial neural network. 2) Diehl et al. describe a method to determine a number of performance metrics, such as the confusion matrix as a function of image resolution. These metrics are estimated experimentally using synthetic target images, which are degraded according to a noise model. 3) Irving et al. use Monte Carlo simulation to estimate classification performance using peak features represented by their locations. Performance is estimated as a function of a number of factors such as number of target classes, peak positional uncertainty, and clutter rate.

- **Formal Approaches**: A statistical framework is developed to directly predict performance for a given set of statistical data-distortion models (see Figure 1(c)). For example, Grenander et al. present a method to predict fundamental performance of template-based ATR for a given image noise model (a probabilistic likelihood function) using Hilbert-Schmidt estimator. The expected error of this estimator (Hilbert-Schmidt bound) forms a lower bound on the error of any other estimator. Predicted bounds are not validated experimentally in that work. Irving et al. present a formal method for predicting performance of a relevant problem, which is target detection. Their method, which considers a target detection approach involving peak features, is based on statistical modeling of both data distortion factors, and model target views.

Most ATR performance prediction approaches belong to the first category. There is a lack of formal methods that predict performance based on the amount of information provided by sensory data (Ratches et al.). Such approaches can provide a valuable insight into the roles played by different data and model factors (such as those mentioned in the previous section) in determining the recognition performance.
The prediction method presented in this paper belongs to the second category. Compared to previous relevant work, the framework presented in this paper is unique in the following aspects:

- An upper bound on performance of vote-based ATR (using point features) is predicted.
- Data uncertainty, occlusion, clutter and model similarity are considered simultaneously.
- Real SAR data (MSTAR) are used for validating the method.

3. MODEL SIMILARITY

In this section, we define a quantitative measure of the structural similarity between a pair a model views represented by locations of point features (e.g., scattering centers in SAR images).

We define the following entities:

- \( \mathcal{MD} \): a database of model views belonging to a set of model targets,
- \( \mathcal{M}_i \in \mathcal{MD} \): \( i \)th model view in the database of model views,
- \( \mathcal{C}_i \): the coordinate frame associated with \( \mathcal{M}_i \),
- \( F_{ik} \in \mathcal{M}_i \): \( k \)th feature of \( \mathcal{M}_i \), represented by its location with respect to \( \mathcal{C}_i \),
- \( \mathcal{M} \): a data view of some model target,
- \( \mathcal{C} \): the coordinate frame associated with \( \mathcal{M} \),
- \( \mathcal{F}_i \in \mathcal{M} \): \( i \)th feature of \( \mathcal{M} \) represented by its location with respect to \( \mathcal{C} \),
- \( \mathcal{T} \): the space of applicable data/model transformations,
- \( \tau \in \mathcal{T} \): an applicable transformation (e.g., if \( \mathcal{T} \) is the space of 2-D translations, then \( \tau = \begin{pmatrix} \tau_x & \tau_y \end{pmatrix} \)),
- \( 0 \in \mathcal{T} \): origin of \( \mathcal{T} \),
- \( \text{TRANS}(\mathcal{C}, \mathcal{C}') \): the transformation of coordinate frame \( \mathcal{C} \) with respect to \( \mathcal{C}' \),
- \( \mathcal{M}^*_i \): a hypothesis of view \( \mathcal{M}_i \) at location \( \tau \) with respect to \( \mathcal{C} \), i.e., \( \tau = \text{TRANS}(\mathcal{C}_i, \mathcal{C}) \),
- \( \mathcal{M}^*_j \): a hypothesis of view \( \mathcal{M}_j \) at location \( \tau \) with respect to \( \mathcal{C}_i \), i.e., \( \tau_i = \text{TRANS}(\mathcal{C}_i, \mathcal{C}) \), and
- \( \text{AREA}(R) \): area of region \( R \).

The proposed similarity measure is dependent upon the amount of uncertainty in the data. This agrees with the intuitive observation that as views get "blurred", it becomes more difficult to distinguish between them, which is equivalent to saying that they become more "similar" to each other. In the following, we introduce a sequence of definitions that lead to a formal definition of model similarity.

1) Data Uncertainty: The uncertainty associated with data feature \( \mathcal{F}_i \) is a probabilistic description of the location of \( F_{ik} \), the model feature that \( \mathcal{F}_i \) corresponds to. In this paper, we assume that this uncertainty is a uniform distribution. Accordingly, it can be represented by an uncertainty region, \( R_u(\cdot) \). In particular, we have \( F^*_k \in R_u(\mathcal{F}_i) \), where \( F^*_k \in \mathcal{M}^*_i \). Furthermore, we assume that the uncertainty distributions associated with the data features are independent.

2) Feature Consistency: A pair of data and model features, \( \mathcal{F}_i \) and \( F^*_k \), is said to be consistent if \( \mathcal{F}_i \) can be interpreted as an uncertain measurement of \( F^*_k \). It can be easily shown that the condition for consistency is \( F^*_k \in R_u(\mathcal{F}_i) \), or equivalently \( \mathcal{F}_i \in R_u(F^*_k) \), where \( R_u(\cdot) \) is the reflection of \( R_u(\cdot) \) about the origin. We refer to \( R_u(\cdot) \) as a consistency region. In most practical situations, region \( R_u(\cdot) \) is symmetric about the origin (e.g., a circle, ellipse or square). In such a case, we have \( R_u(\cdot) = R_u(\cdot) \). Notice that consistency region \( R_u(\cdot) \) defines the number of votes.
for a specific hypothesis. In order to demonstrate that, let $\text{VOTES}(\mathcal{M}_i^t; \hat{\mathcal{M}})$ be the number of votes for hypothesis $\mathcal{M}_i^t$ given data view $\hat{\mathcal{M}}$. We can express $\text{VOTES}(\mathcal{M}_i^t; \hat{\mathcal{M}})$ as follows:

$$\text{VOTES}(\mathcal{M}_i^t; \hat{\mathcal{M}}) = \{ F_{ik} : F_{ik} \in \mathcal{M}_i^t \text{ and } \exists \hat{F}_i \in \hat{\mathcal{M}} \text{ s.t. } \hat{F}_i \in \hat{R}_u(F_{ik}) \}.$$ 

3) Feature/Feature Similarity: The similarity between a pair of model features, $F_{ik} \in \mathcal{M}_i$ and $F_{jil} \in \mathcal{M}_j^{\tau_l}$, denoted by $S_{ff}(F_{ik}, F_{jil})$, is defined as the probability that an uncertain measurement of $F_{ik}$ is consistent with $F_{jil}$. Formally,

$$S_{ff}(F_{ik}, F_{jil}) = \frac{\text{AREA}(\hat{R}(F_{ik}) \cap \hat{R}(F_{jil}^{\tau_l}))}{\text{AREA}(\hat{R}(F_{ik}))}.$$ 

Notice that $S_{ff}(F_{ik}, F_{jil}^{\tau_l})$, which lies in the range $[0, 1]$, is proportional to the extent of overlap between $\hat{R}(F_{ik})$ and $\hat{R}(F_{jil}^{\tau_l})$. We sometimes refer to feature pairs with overlapping/non-overlapping consistency regions as similar/dissimilar features, respectively. Figure 2 shows $S_{ff}(F_{ik}, F_{jil}^{\tau_l})$ as a function of $\tau_l$ for a sample of three uncertainty regions, assuming that $T$ is the space of 2-D translations.

4) View/Feature Similarity: The similarity between a view, $\mathcal{M}_i$, and a feature, $F_{jil}^{\tau_l} \in \mathcal{M}_j^{\tau_l}$, is defined as the probability that an uncertain measurement of any feature in $\mathcal{M}_i$ is consistent with $F_{jil}^{\tau_l}$. Formally, view/feature similarity, $S_{vf}(\mathcal{M}_i, F_{jil}^{\tau_l})$, can be defined as follows:

$$S_{vf}(\mathcal{M}_i, F_{jil}^{\tau_l}) = 1 - \prod_k (1 - S_{ff}(F_{ik}, F_{jil}^{\tau_l})).$$

Figure 2: An illustration of feature/feature similarity for three uncertainty regions: (a) a circle of unit radius, (b) a discrete four-neighbor region, (c) a point region (implying that there is no uncertainty). For the sake of simplicity, we assume in this figure that $F_{ik} = F_{jil}^{\tau_l} | r_i = 0$. 

719
5) View/Hypothesis Similarity: The structural similarity between a view, $\mathcal{M}_i$, and a view hypothesis, $\mathcal{M}_j^r$, is defined as the number of features in $\mathcal{M}_j^r$ consistent with uncertain measurements of features in $\mathcal{M}_i$. Obviously, view/hypothesis similarity, denoted as $S_{\text{sh}}(\mathcal{M}_i, \mathcal{M}_j^r)$ or simply $S_j^r$, is a random variable. It can be shown that $S_j^r$ is bounded as follows:

\[
\begin{align*}
\min(S_j^r) &= \{ \{ F_{ji}^r : F_{ji}^r \in \mathcal{M}_j^r \text{ and } S_{\text{sh}}(\mathcal{M}_i, F_{ji}^r) = 1 \} \\ 
\max(S_j^r) &= \{ \{ F_{ji}^r : F_{ji}^r \in \mathcal{M}_j^r \text{ and } S_{\text{sh}}(\mathcal{M}_i, F_{ji}^r) > 0 \} 
\end{align*}
\]

Figure 3(a) shows an example of view/hypothesis similarity and associated bounds.

In order to simplify the estimation of view/hypothesis similarity, and subsequently the prediction of performance in the next section, we make the following reasonable assumptions about $\mathcal{M}_i$ and $\mathcal{M}_j^r$ and the similarity between their features:

1. There is no overlap between consistency regions of the features that belong to each of $\mathcal{M}_i$ and $\mathcal{M}_j^r$,

2. There is a one-to-one correspondence between similar features in $\mathcal{M}_i$ and $\mathcal{M}_j^r$, and

3. The similarity between every pair of similar features in $\mathcal{M}_i$ and $\mathcal{M}_j^r$ is a constant value, which is the average view/feature similarity.

The above assumptions result in a “uniform” view of the structural similarity between $\mathcal{M}_i$ and $\mathcal{M}_j^r$. As an example, Figure 3(b) illustrates the uniform similarity corresponding to the view/hypothesis pair shown in Figure 3(a). Based on the above assumptions, we can express the expected value of $S_j^r$ as

\[
E(S_j^r) = \sum_i S_{\text{sh}}(\mathcal{M}_i, F_{ji}^r).
\]

In addition, we can represent the probability mass function (PMF) of $S_j^r$ by a binomial distribution defined as follows:

\[
\Pr[S_j^r = s] = B_{S_j^r}(s; N_j^r, Q_j^r)
\]

where $N_j^r = \max(S_j^r)$, $Q_j^r = \frac{E(S_j^r)}{N_j^r}$, $B_X(x; n, p) = C(n, x)p^x(1-p)^{n-x}$, and $C(a, b) = \frac{a!}{(a-b)!b!}$.

6) View/View Similarity: The similarity between a pair of views, $\mathcal{M}_i$ and $\mathcal{M}_j$, is defined as the view/hypothesis similarity $S_{\text{sh}}(\mathcal{M}_i, \mathcal{M}_j^r)$ (or $S_j^r$), for all $r \in T$. Thus, view/view similarity can be viewed as a probabilistic function. As an example, Figures 4(a) and 4(b) show a pair of simple model views and corresponding expected similarity function, respectively. Peaks in this function correspond to view hypotheses that have a higher degree of similarity with $\mathcal{M}_i$ than neighboring hypotheses. A sample of three peak hypotheses is shown in Figure 4(c). In the next section, peak hypotheses are used for predicting an upper bound on ATR performance.
Figure 4. An illustration of view/view similarity assuming a four-neighbor discrete uncertainty region, and a discrete 2-D translation space: (a) a pair of model views, $M_i$ and $M_j$, (b) corresponding expected-similarity function, $E(S_{ij}^T)$, (c) a sample of three peak hypotheses.

4. PERFORMANCE PREDICTION

In this section, we present a similarity-based framework for predicting an upper bound on performance of vote-based ATR in the presence of data distortion. We assume that locations of model and data features are discretized at some resolution. Furthermore, the space of applicable transformations $T$ is assumed to be discrete 2-D translations.

We define the following entities:

- $\mathcal{M}_i$: a data view that is a distorted version of $M_i$,
- $O$: number of features in $M_i$ that are occluded in $\mathcal{M}_i$,
- $C$: number of clutter features in $\mathcal{M}_i$,
- $M^T_{pq}$: a view hypothesis corresponding to a peak in expected-similarity function $E(S_{ij}^T)$,
- $(N^T_{pq}, Q^T_{pq})$: similarity parameters associated with peak hypothesis $M^T_{pq}$,
- $V^T_{ij}$: votes for view hypothesis $M^T_{ij}$, given $\mathcal{M}_i$ (i.e., $V^T_{ij} = \text{VOTES}(M^T_{ij}; \mathcal{M}_i)$),
- $V_i$: votes for $M_i$, given $\mathcal{M}_i$ (notice that $V_i \equiv V^T_{ii}|_{r=0}$),
- $V^C_{ij}$: votes for $M^T_{ij}$, given $\mathcal{M}_i$, due to similarity with $M_i$,
- $V^C_{ij}$: votes for $M^T_{ij}$, given $\mathcal{M}_i$, due to coincidence with clutter features,
- $V^C_{pq}$: votes for peak hypothesis $M^T_{pq}$, given $\mathcal{M}_i$,
- $\mathcal{H}^T_{ij}$: set of hypotheses of $M_j$ relative to $M_i$ (i.e., $\mathcal{H}^T_{ij} = \{M^T_{ij} : \forall r_i \in T\}$),
- $\mathcal{H}^T_i$: set of hypotheses of $M_i$ relative to itself excluding $M_i$ (i.e., $\mathcal{H}^T_i = \{M^T_{ij} : \forall r_i \in T\} - \{M_i\}$),
- $\mathcal{H}^T$: set of all hypotheses relative to $M_i$ excluding $M_i$ (i.e., $\mathcal{H}^T = \bigcup_j \mathcal{H}^T_{ij}$),
- $\mathcal{H}^T_{pj}$: set of peak hypotheses of $M_j$ relative to $M_i$,
- $\mathcal{H}^T_{pj}$: set of peak hypotheses of $M_j$ relative to itself excluding $M_i$, and
- $\mathcal{H}^T$: set of all peak hypotheses relative to $M_i$ excluding $M_i$ (i.e., $\mathcal{H}^T = \bigcup_j \mathcal{H}^T_{pj}$).
Figure 5. A simple example showing the sequence of distortion processes that transform a model view, $\mathcal{M}_i$, to a corresponding data view, $\widehat{\mathcal{M}}_i$. The circles shown represent consistency regions associated with features of $\mathcal{M}_i$. Clutter features are shown as crosses to distinguish them from perturbed features of $\mathcal{M}_i$ observed in $\widehat{\mathcal{M}}_i$.

4.1. Modeling of Data Distortion

A model view, $\mathcal{M}_i$, is assumed to be distorted to form a data view, $\widehat{\mathcal{M}}_i$, in the following sequence (refer to the example shown in Figure 5):

1. A group of $O$ model features are occluded. We assume that occlusion is uniform; i.e., each subset of features is equally likely to be occluded as any other subset of the same size.

2. Uncoccluded model features are uniformly perturbed within associated consistency regions, $\mathcal{R}_e()$. Since we are considering discretized features, their consistency regions are also discrete (e.g., the four-neighbor region shown in Figure 2(b)).

3. Resulting (occluded and perturbed) model view is cluttered by adding $C$ spurious features to form $\widehat{\mathcal{M}}_i$. We assume that the distribution of those features is uniform within some clutter region, $\mathcal{R}_c$, encompassing the original model view (e.g., convex hull, bounding box). In order to control the number of votes for $\mathcal{M}_i$, we exclude the consistency regions associated with the occluded features from the original clutter region. The resulting region, $\mathcal{R}_c$, is referred to as an effective clutter region.

4.2. Predicting an Upper Bound on Performance

Let us assume that we are given a data view, $\widehat{\mathcal{M}}_i$. The condition for correct recognition of $\widehat{\mathcal{M}}_i$ as $\mathcal{M}_i$ (notice that $\mathcal{M}_i = \mathcal{M}_i^T | r_i=0$) is that the number of votes for $\mathcal{M}_i$ is greater than the number of votes for all erroneous hypotheses in $\mathcal{H}^i$. Accordingly, we can express the probability of correctly recognizing $\mathcal{M}_i$ in a distorted view $\widehat{\mathcal{M}}_i$ as

$$\Pr[\mathcal{M}_i; \widehat{\mathcal{M}}_i] = \Pr[\mathcal{M}_i^T \in \mathcal{H}^i : V_i > V_{j_i}^T; \widehat{\mathcal{M}}_i],$$

where $V_i$ and $V_{j_i}^T$ are the numbers of votes for $\mathcal{M}_i$ and $\mathcal{M}_i^T$, respectively. According to our modeling of the data distortion process, described in Section 4.1, it is easy to obtain $V_i$:

$$V_i = |\mathcal{M}_i| - O.$$  

On the other hand, $V_{j_i}^T$ is a random variable that depends on the degree of similarity between $\mathcal{M}_i$ and $\mathcal{M}_i^T$, and the number of clutter features that happen to be consistent with features in $\mathcal{M}_i^T$. The PMF of $V_{j_i}^T$ is derived in the next section.
Determining $\Pr[M_i; \widetilde{M}_i]$ appears to be a very complicated task. This is because the random variables $V_j^{\tau_i}$, for different $\mathcal{M}_j \in \mathcal{H}^i$, are not independent. This is especially true for votes of adjacent hypotheses, due to the spatial overlap between their consistency regions. This is the reason that we are resorting to estimation of performance bounds, rather than attempting to estimate the actual performance.

We present three possible upper bounds on $\Pr[M_i; \widetilde{M}_i]$, which vary in their tightness and reliance on assumptions. From (2), it is easy to see that a recognition error occurs when any of the hypotheses in $\mathcal{H}^i$ reaches or exceeds the votes for $M_i$. Accordingly, the maximum of the associated probabilities forms a lower bound on the probability of recognition error. This directly leads to the following upper bound on $\Pr[M_i; \widetilde{M}_i]$ (refer to (2) and (3)):

$$\Pr[M_i; \widetilde{M}_i] < 1 - \max_{\tau, j} \Pr[V_j^{\tau_i} \geq |M_i| - |O|; \widetilde{M}_i],$$

(4)

where $\mathcal{M}_j \in \mathcal{H}^i$. Obviously, such a bound is not expected to be tight, since it only considers a single hypothesis in $\mathcal{H}^i$. In order to obtain a tighter bound, one possible approach is to select a subset of hypotheses, and assume that the votes for those that belong to the same view are independent. In our work, we consider hypotheses that correspond to peaks in the expected similarity function (refer to Section 3). The rationale behind such a choice can be stated as follows:

1. Peak hypotheses are more similar to corresponding model views than neighboring hypotheses, and so, a distorted model view is generally more likely to be misinterpreted as a peak hypothesis than an off-peak neighboring one.
2. Peak hypotheses are not very close to each other, thus making the vote-independence assumption reasonable.

Based on such an approach, we can obtain the following upper bound:

$$\Pr[M_i; \widetilde{M}_i] < \min_j \prod_{\tau_i} (1 - \Pr[V_j^{\tau_i} \geq |M_i| - |O|; \widetilde{M}_i])$$

(5)

where $\mathcal{M}_j \in \mathcal{H}^i$. Obviously, the above bound is expected to be considerably tighter than the one expressed in (4), since it considers a representative subset of hypotheses that belong to a model view (peak ones), not just a single hypothesis as in (4). A tighter bound can be obtained by considering a representative subset of hypotheses that belong to all model views, not just a single view as in (5). In order to achieve this goal, we further make the assumption that votes for peak hypotheses belonging to different views are independent. Based on such an assumption, we obtain:

$$\Pr[M_i; \widetilde{M}_i] < \prod_{\mathcal{M}_j \in \mathcal{H}^i} (1 - \Pr[V_j^{\tau_i} \geq |M_i| - |O|; \widetilde{M}_i])$$

(6)

where $\mathcal{M}_j \in \mathcal{H}^i$. While this assumption is invalid in some extreme cases, such as when we have identical or very similar views, we argue that it is reasonable in most practical SAR ATR scenarios. This is because, even in the presence of highly similar targets, their corresponding SAR views are not expected to be as similar, due to the high sensitivity of the SAR imaging process to slight geometric changes in target shape.

4.3. Determining the Probability Distribution of Hypothesis Votes

In this section, we determine the PMF of $V_j^{\tau_i}$, the number of votes for hypothesis $\mathcal{M}_j$ given distorted data view $\widetilde{M}_i$. The votes for $\mathcal{M}_j$ come from two different sources:

1. **Similarity**: This source is due to the similarity between $\mathcal{M}_j$ and $M_i$. As discussed in Section 3, this similarity is represented by the number of similar feature pairs and the average view/feature similarity. In addition to similarity, the number of similarity votes for $\mathcal{M}_j$ depends on the number of unoccluded features of $M_i$ that are similar to features in $\mathcal{M}_j$.

2. **Clutter**: This source is due to random coincidence between clutter features and consistency regions of features in $\mathcal{M}_j$. The number of clutter votes for $\mathcal{M}_j$ depends on both the number of clutter features and the number of features of $\mathcal{M}_j$ that lie inside the clutter region, $R_c$.

An example of similarity and clutter votes is shown in Figure 6(a).

From the above discussion, we can express $V_j^{\tau_i}$ as

$$V_j^{\tau_i} = V_{Oj}^{\tau_i} + V_{Cj}^{\tau_i}$$

(7)
where \( V_{ij}^s \) and \( V_{ij}^c \) are the numbers of similarity and clutter votes for \( M_j \), respectively. Let \( N_{ij}^s \) be a random variable denoting the number of unoccluded features in \( M_i \) that are similar to features in \( M_j \). The PMF of \( V_{ij}^s \) can be expressed as

\[
\Pr[V_{ij}^s = v_i; \tilde{\mathcal{M}}_i] = \sum_{n_o} \Pr[N_{ij}^s = n_o; \tilde{\mathcal{M}}_i] \Pr[V_{ij}^s = v_i; N_{ij}^s = n_o, \tilde{\mathcal{M}}_i].
\]  
(8)

From (7) and (8), we can express the PMF of \( V_{ij}^c \) as

\[
\Pr[V_{ij}^c = v_i; \tilde{\mathcal{M}}_i] = \sum_{n_o} \Pr[N_{ij}^c = n_o; \tilde{\mathcal{M}}_i] \sum_{v_s} \Pr[V_{ij}^s = v_s; N_{ij}^c = n_o, \tilde{\mathcal{M}}_i] \Pr[V_{ij}^c = v - v_s; N_{ij}^c = n_o, V_{ij}^s = v_s, \tilde{\mathcal{M}}_i].
\]  
(9)

In order to simplify the process of computing the PMF of \( V_{ij}^c \), we consider the uniform view of similarity described in Section 3 (see Figure 3(b)). First, we determine the PMF of \( N_{ij}^s \) and the conditional PMF of \( V_{ij}^c \). It can be easily shown that \( N_{ij}^c \) is described by the following hypergeometric distribution:

\[
\Pr[N_{ij}^c = n_o; \tilde{\mathcal{M}}_i] = H_{N_{ij}^c}(N_{ij}^c = n_o; O, N_{ij}^c, | \mathcal{M}_i | - N_{ij}^c)
\]  
(10)

where \( H_X(x; a, b) = \frac{C(a, x)C(b, n-x)}{C(a+b, n)} \). The conditional PMF of \( V_{ij}^c \) is described by the following binomial distribution:

\[
\Pr[V_{ij}^c = v_i; N_{ij}^c = n_o, \tilde{\mathcal{M}}_i] = B_{v_i}(v_i; n_o, Q_{ij}^c).
\]  
(11)

Next, we estimate the conditional PMF of \( V_{ij}^c \). This appears to be a complex task, due to the following reasons. Firstly, in some cases, clutter features have to be at a minimum distance from each other. For example, if point features extracted from a SAR image correspond to local peaks, then no peak can lie in the eight-neighbor region of another one. Secondly, clutter features that fall within the same consistency region contribute only a single vote. Thirdly, since consistency regions of occluded features are excluded from the clutter region (as described in Section 4.1), features of \( M_j \) that are similar to occluded ones in \( M_i \) effectively have smaller consistency regions (depending on the average view/feature similarity, \( Q_{ij}^c \)). One simple approximation of the conditional PMF of \( V_{ij}^c \) can be outlined as follows.

- The effective clutter region, \( R'_c \), is split into two sub-regions, \( R'_c \) and \( (R'_c - R'_c) \), such that a clutter feature falling within the first (second) sub-region does (not) vote for \( M_j \). It can be shown that \( R'_c \), referred to as the clutter-vote region, is the intersection of \( R'_c \) and the consistency regions of features of \( M_j \) that are not contributing similarity votes. Figure 6(b) shows an example of a clutter-vote region, which corresponds to the view/hypothesis pair shown in Figure 6(a).
• Each clutter feature that falls in \( R_k \) is assumed to contribute a vote for \( M^*_j \).

• Locations of clutter features are assumed to be independent of each other.

This approach results in approximation of the conditional PMF of \( V^*_j \) by the following binomial distribution:

\[
\Pr[V^*_j = v_c; N^*_j = n_o, V^*_j = v_s, \widetilde{M}_i] \approx B(v_c; C, \text{AREA}(R^*_j)/\text{AREA}(R^*_k))
\]

where

\[
\text{AREA}(R^*_j) = \text{AREA}(R_c) - \text{AREA}(R_o) + \text{O}, \quad \text{and} \\
\text{AREA}(R^*_k) = \text{AREA}(R_o)((M^*_j \cap R_c) - v_s - Q^*_j(N^*_j - n_o)).
\]

In the Appendix, we present a more accurate approximation that considers some of the above-mentioned constraints.

4.4. Algorithm

In this section, we present the algorithm used to determine an upper bound on average PCR, given a database of model views, \( \text{MD} \), and statistical models for data distortion. The algorithm consists of two steps. In the first step, similarity information involving all model-view/peak-hypothesis pairs, \( M_i/H^*_j \) for all \( i \), is calculated and accumulated in a peak similarity histogram. In the second step, the peak similarity histogram along with the distortion models are used to predict PCR upper bound. These two steps are described as follows.

1) Construction of Peak Similarity Histogram: From the analysis presented in the previous section, it can be observed that the PMF of \( V^*_j \) depends on four view-dependent parameters only: \( |M_i|, |M_j^* \cap R_c| \) (size of the section of \( M_j^* \) that lies inside clutter region \( R_o \)), and similarity parameters \( (N^*_j, Q^*_j)^* \). Accordingly, only these parameters need to be stored when calculating the similarity between model views. In particular, for each view/peak-hypothesis pair \( M_i/H^*_j \), where \( M_i \in \text{MD} \), entry \( (|M_i|, |M_j^* \cap R_c|, N^*_j, Q^*_j) \) in the 4-D peak similarity histogram (PSH) is incremented by one.

2) Computation of PCR Bound: We consider the upper bound defined in (6). Let

\[
T(a, b, c, d) = \Pr[V^*_j \geq |M_i| - O, \widetilde{M}_i]
\]

such that \( a = |M_i|, b = |M_j^* \cap R_c|, c = N^*_j, \) and \( d = |N^*_j + Q^*_j + 0.5| \). The upper bound on average PCR for model database \( \text{MD} \), the main result of this paper, can be approximated as follows:

\[
\text{PCRUB}(\text{MD}) \approx \prod_a \prod_b \prod_c \prod_d (1 - T(a, b, c, d))^{\text{PSH}(a, b, c, d)/|\text{MD}|}.
\]

5. EXPERIMENTAL VALIDATION

In this section, we validate the proposed performance prediction method by comparing actual PCR plots, as a function of occlusion and clutter rates, with predicted upper bounds.

1) Model Databases: We have selected two model databases consisting of military targets from the MSTAR public data domain. The first database, \( \text{MD}_1 \), consists of three targets, T72, BMP2, and BTR70, while the second one, \( \text{MD}_2 \), consists of two targets, T72 and ZSU. Each target is represented by a number of SAR views, which sample its signature at a variety of azimuth angles and a specific depression angle. The two databases are described in Table 1. Each model view is processed to extract eight-neighborhood peaks, which are the scattering centers used for recognition. The strongest 30 scattering centers in magnitude are chosen to represent each model view. Figure 7 shows a sample of three model views (from \( \text{MD}_1 \)) along with extracted scattering centers superimposed on corresponding regions of interest (ROI's).

2) Test Sets: We have chosen five sets of test views (see Table 2). The first two sets, \( \text{TS}_1 \) and \( \text{TS}_2 \), are obtained by introducing distortion to model databases, \( \text{MD}_1 \) and \( \text{MD}_2 \), respectively. Distortion is introduced in a manner similar to that described in Section 4.1 (see also Figure 5). Details of this process are:

*In the actual implementation, we have included features of \( M_j^* \) close to the boundary of \( R_c \) on the outside that are similar to features in \( M_i \).
Table 1. Description of the model databases

<table>
<thead>
<tr>
<th>Database</th>
<th>Depression Angle</th>
<th>Targets</th>
<th>No. of Views</th>
<th>Total No. of Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD_1</td>
<td>17°</td>
<td>T72 (#812) BMP2 (#c21) BTR70 (#c71)</td>
<td>231 233 233</td>
<td>697</td>
</tr>
<tr>
<td>MD_2</td>
<td>30°</td>
<td>T72 (#a64) ZSU (#d08)</td>
<td>288 288</td>
<td>576</td>
</tr>
</tbody>
</table>

Figure 7. Examples of three MSTAR model views at azimuth angle = 132°, and corresponding ROI's with extracted scattering centers superimposed on them.

- The chosen distortion parameters are $O \in \{9, 10, \ldots, 18\}$, $C = O$, $R_u(\cdot)$ = four-neighbor region, and $R_c = \text{convex hull of the features representing the model view under consideration (note that in Figure 7, we show a sample of ROI's, not convex hulls)}. Notice that since $C = O$, test views have the same number of features as the model ones (30 features).

- The locations of clutter features are restricted such that no feature lies within the eight-neighbor region of another one. This constraint is introduced to simulate the process of scattering-center extraction.

- For each $O/C$ value, each model view is distorted four times. Accordingly, for a model database of size $N$ and a specific $O/C$ value, the size of corresponding test subset is $4N$.

The remaining three test sets involve real distortion. The third set, $T S_3$, is a variant of $M D_1$ obtained under a different depression angle (15°, versus 17° for model views). The fourth one, $T S_4$, is a variant of $M D_1$ obtained under different configurations (e.g., difference in number of fuel barrels, different flash lights). The fifth set, $T S_5$, is a variant of $M D_2$ obtained under articulation changes (e.g., rotation of a turret or a gun). Each test view is represented by the strongest 30 scattering centers. We assume that $R_u(\cdot)$ and $R_c$ are the same as in the first two sets. For each test view, the corresponding model view/location is estimated by finding the best match between the test view and model ones within ±3° in azimuth. If no model view exists within such a range, the test view is compared with the model view that is closest in azimuth. The resulting test/model match directly determines the amount of distortion (occlusion and clutter) in the test view. Since test and model views are represented by the same number of features, test-view distortion is described by a single $O/C$ value.
Table 2. Description of the test sets

<table>
<thead>
<tr>
<th>Set</th>
<th>Distortion</th>
<th>Reference Database</th>
<th>Depression Angle</th>
<th>Targets</th>
<th>No. of Views</th>
<th>Total No. of Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>(TS_1)</td>
<td>controlled</td>
<td>(MD_1)</td>
<td>17(^\circ)</td>
<td>T72 (#812)</td>
<td>10 \times 4 \times 231</td>
<td>10 \times 4 \times 697</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BMP2 (#c21)</td>
<td>10 \times 4 \times 233</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BTR70 (#c71)</td>
<td>10 \times 4 \times 233</td>
<td></td>
</tr>
<tr>
<td>(TS_2)</td>
<td>controlled</td>
<td>(MD_2)</td>
<td>30(^\circ)</td>
<td>T72 (#64)</td>
<td>10 \times 4 \times 288</td>
<td>10 \times 4 \times 576</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ZSU (#d08)</td>
<td>10 \times 4 \times 288</td>
<td></td>
</tr>
<tr>
<td>(TS_3)</td>
<td>depression-angle difference</td>
<td>(MD_1)</td>
<td>15(^\circ)</td>
<td>T72 (#812)</td>
<td>193</td>
<td>581</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BMP2 (#c21)</td>
<td>194</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BTR70 (#c71)</td>
<td>194</td>
<td></td>
</tr>
<tr>
<td>(TS_4)</td>
<td>configuration differences</td>
<td>(MD_1)</td>
<td>17(^\circ)</td>
<td>T72 (#132)</td>
<td>231</td>
<td>464</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BMP2 (#9563)</td>
<td>233</td>
<td></td>
</tr>
<tr>
<td>(TS_5)</td>
<td>articulation changes</td>
<td>(MD_2)</td>
<td>30(^\circ)</td>
<td>T72 (#64)</td>
<td>124</td>
<td>242</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ZSU (#d08)</td>
<td>118</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8. Actual PCR plots versus predicted upper bounds for controlled-distortion sets: (a) \(TS_1\), (b) \(TS_2\).

3) Determination of Actual and Predicted PCR Plots: An uncertainty-accommodating recognition system has been implemented for experimentally determining PCR as a function of \(O/C\). Such a system examines all the relevant problem space, which is four-dimensional (target, azimuth, and translations along the range and cross-range directions). Accordingly, its recognition performance is optimal, for the given vote-based criterion (four-neighbor region). Each test set is passed to the recognition system. Results are accumulated in a histogram whose bins correspond to \(O/C\) values. This histogram is directly used to determine the actual PCR plot. The predicted PCR plot (an upper bound) is determined as explained in Section 4.4. Notice that, in our experiments, the shape of the clutter region is view-dependent (the convex hull). Accordingly, the area of the clutter region, \(\text{AREA}(R_c)\), needed to calculate the conditional PMF of clutter votes (Section 4.3) is not of a constant value. In our case, \(\text{AREA}(R_c)\) is substituted by the average of convex hull areas corresponding to views of the appropriate model database. This average is found to be 532 pixels for \(MD_1\), or 482 pixels for \(MD_2\).

4) Results: Figures 8(a), 8(b), and 9(a)-(c) show actual and predicted PCR plots for the five test sets, \(TS_1\) to \(TS_5\), respectively. From these plots, we observe that the proposed performance method succeeds in predicting relatively tight upper bounds on ATR performance. We also observe that the predicted upper bounds in the controlled-distortion cases (Figures 8(a) and 8(b)) are tighter than those in the real-distortion ones (Figures 9(a)-(c)). This is because of some differences between the actual distortion models and the assumed ones (i.e., uniform uncertainty, occlusion and clutter models, where clutter region is convex hull of view features). Development of methods to learn accurate distortion models for a given set of data is an interesting subject for future research.
6. CONCLUSIONS

A novel statistical framework has been presented for predicting an upper bound on performance of vote-based SAR ATR. The framework is based on a quantitative measure of structural similarity between model target views. It considers data distortion factors such as uncertainty, occlusion and clutter. The validity of the method has been demonstrated by comparing experimentally-determined PCR plots with predicted upper bounds in the presence of both controlled and real distortion. Test sets involving real distortion are obtained by varying depression angle, configuration, and articulation. Future work involves the development of tighter upper bounds.

Acknowledgments

This work was supported by DARPA/AFOSR grant F49620-97-1-0184; the contents and information do not reflect positions or policies of the U.S. Government.

Appendix: Estimation of the Conditional Distribution of Clutter Votes

We present an approximate method to estimate $\Pr[V_{c}^j = v_i; N_{c}^{j} = n_o; \hat{M}_i]$. Let $R_s(\cdot)$ be a separation region, which defines the minimum degree of adjacency between features. For example, if point features are local image peaks, then $R_s(\cdot)$ will be an eight-neighbor region. Considering the separation region, we can approximate the area of the new effective clutter region, $\mathcal{R}_s$, as

$$\text{AREA}(\mathcal{R}_s) \approx \text{AREA}(R_c) - \text{AREA}(R_a(\cdot))O - \text{AREA}(R_a(\cdot))\{1 - |M_i| - O\}.$$ 

Features of $M_i$ that can contribute clutter votes are classified into three categories: 1) Features dissimilar to all features of $M_i$, 2) features similar to occluded features of $M_i$, and 3) features similar to unoccluded features of $M_i$, but are not contributing similarity votes (refer to Figure 6(b)). Assuming $\text{AREA}(R(a)(\cdot)) \leq \text{AREA}(R_a(\cdot))$ (as is the case in our experiments), we ignore the third type of features. It can be observed that a feature of the second type has a “truncated” consistency region of area $(1 - Q_j^i)\text{AREA}(R_a(\cdot))$, since consistency regions of occluded features of $M_i$ do not receive clutter features. To simplify the computations, we consider a single “effective” consistency region, $\mathcal{R}_e(\cdot)$, whose area is the weighted average of the areas of full and truncated consistency regions. From the above discussion, it can be shown that there are $|M_i^j \cap R_c| - N_j^i$ full consistency regions and $N_j^i - n_o$ truncated ones. Accordingly, we get:

$$\text{AREA}(\mathcal{R}_e(\cdot)) = \text{AREA}(R_a(\cdot))\left|\frac{|M_j^i \cap R_c| - N_j^i + (1 - Q_j^i)(N_j^i - n_o)}{|M_j^i \cap R_c| - n_o}\right.$$ 

Given $m$ model features, $n$ clutter features, a clutter region of area $a$, an uncertainty region of area $b$, and a separation region of area $c$, the probability that $z$ model features are consistent with clutter features can be represented as follows:

$$G_x(z; m, n, a, b, c) \approx C(m, x)P(n, z)p^z(1 - (m - x)p)^{n - z}.$$
where \( P(x, y) = \frac{\exp(-x - y)}{(\pi \sigma^2)^{1/4}} \), \( p = \frac{a}{b} \) and \( \hat{p} = \frac{b}{a - c} \). Mapping that PMF to our case, we obtain the following estimate of the conditional PMF of \( V_{e}^{i} \):

\[
\Pr[V_{e}^{i} = v_{e}; N_{e}^{i} = n_{e}, V_{e}^{i} = v_{e}, \mathcal{M}_{e}] \approx \nabla_{V_{e}^{i}}(v_{e}, \mathcal{M}_{e}^{i} \cap R_{e}^{c} \mid -n_{e}, C, \text{AREA}(R_{e}^{c})), \text{AREA}(R_{e}^{c})), \text{AREA}(R_{e}^{c})).
\]

REFERENCES