Composite Class Models for SAR Recognition

Bir Bhanu, Grinnell Jones III and Rong Wang
Center for Research in Intelligent Systems
University of California, Riverside, California 92521 USA

ABSTRACT

This paper focuses on a genetic algorithm based method that automates the construction of local feature based composite class models to capture the salient characteristics of configuration variants of vehicle targets in SAR imagery and increase the performance of SAR recognition systems. The recognition models are based on quasi-invariant local features: SAR scattering center locations and magnitudes. The approach uses an efficient SAR recognition system as an evaluation function to determine the fitness of candidate members of a genetic population of new models and synthetically generates composite class models. Experimental results are given on the fitness of the composite models and the similarity of both the original training model configurations and the synthesized composite models to the test configurations. In addition, results are presented to show the SAR recognition performance and pose accuracy for training models and composite class models of configuration variants of MSTAR vehicle targets.

Keywords: automatic target recognition, recognizing configuration variants, genetic algorithms, synthetic aperture radar, automatic model construction

1. INTRODUCTION

In this paper we are concerned with methods to automate the construction of composite recognition models that capture the salient characteristics of configuration variants of real vehicles in Synthetic Aperture Radar (SAR) imagery to improve the performance of a recognition system. The recognition system starts with real SAR chips of actual military vehicles from the MSTAR public data16 and ends with the identification of a specific vehicle type (e.g., a T72 tank).

A major challenge is that the vehicles can be in articulated configurations (such as a tank with its turret rotated), have significant external configuration variants (fuel barrels, searchlights, etc.) or they can be partially occluded. The detection theory,7,8 pattern recognition14,15,17 and neural network9 approaches to SAR recognition all tend to use global features that are optimized for standard, non-articulated, non-occluded configurations. Approaches that rely on global features are not appropriate for recognizing occluded (or articulated) objects because occlusion (or articulation) changes global features like the object outline and major axis.10 Previous work by Jones and Bhanu1,2,10,11,12 relied on local features to successfully recognize articulated and highly occluded objects. It started by using invariant locations of SAR scattering centers as features and later developed techniques using quasi-invariant locations and magnitudes of the scattering centers. In this prior work, the most difficult challenge for the local features approach was not occlusion, articulation, or depression angle changes; it was always the configuration differences. Other work, by Boshra and Bhanu on predicting the performance of recognition systems,5,6 introduced the idea that recognition performance depends on the distortion in the test data and the inherent similarity of the object models. Bhatnager et al.3,4 determined that there

Further author information -
Email: {bhanu, grinnell or rwang} @cris.ucr.edu
URL: http://www.cris.ucr.edu

264

was significant intra-class variability among the MSTAR T72 configurations and took an approach to building class models that used multiple templates from different configurations.

In this paper we develop a genetic algorithm based approach that incorporates the recognition system in the fitness evaluation to determine the similarities among the configurations of object models and use this a priori knowledge to synthesize new local feature based composite recognition models. We then evaluate the subsequent recognition performance of both the original training models and the new composite models.

The key contributions of this paper are:

1. Develops an approach using a genetic algorithm to synthesize class models of objects.

2. Uses a SAR recognition engine to evaluate the fitness of synthesized models with respect to various object configurations.

The remainder of the paper is organized as follows: The next section gives a description of the basic SAR recognition system. Section 3 describes the genetic algorithm approach used to synthesize composite class models. Section 4 gives experimental results. Finally conclusions are drawn and areas of future work are given in Section 5.

2. SAR RECOGNITION SYSTEM

The basic SAR recognition system is an off-line model construction process and a subsequent on-line recognition process. The approach is designed for SAR and is specifically intended to accommodate recognition of articulated and occluded objects. Standard non-articulated models of the objects are used to recognize these same objects in non-standard, articulated and occluded configurations. The models are a look-up table and the recognition process is an efficient search for positive evidence, using relative locations of the scattering centers in the test image to access the look-up table and generate votes for the appropriate object (and azimuth pose).

The relative locations and magnitudes of the N strongest SAR scattering centers (local maxima in the radar return signal) are used as characteristic features (where N, the number of scattering centers used, is a design parameter). Because of the specular radar reflections in SAR images, a significant number of features do not typically persist over a few degrees of rotation. Consequently, we model each object at 1° azimuth increments. Any local reference point, such as a scattering center location, can be chosen as a 'basis point' to establish a reference coordinate system for building a model of an object at a specific azimuth angle pose. The relative distance and direction of other scattering centers can be expressed in radar range and cross-range coordinates and naturally tessellated into integer buckets that correspond to the radar range/cross-range bins. For ideal data, picking the location of the strongest scattering center as the basis point is sufficient. However, for potentially corrupted data where any scattering center could be spurious or missing (due to the effects of noise, target articulation, occlusion, non-standard target configurations, etc.), we use all N strongest scattering centers in turn as basis points to ensure that a valid basis point is obtained. Thus, to handle articulation and occlusion, the size of the look-up table models (and also the number of relative distances that are considered in the test image during recognition) are increased from N to N(N − 1)/2. Using a technique like geometric hashing, the models are constructed using the relative positions of the scattering centers in the range and cross-range directions as the initial indices to a look-up table of labels that give the associated target type, target pose, basis point range and cross-range positions and the magnitudes of the two scatterers. Since the relative distances are not unique, there can be many of these labels (with different target, pose, etc. values) at each lookup table entry.
The recognition process uses the relative locations of the $N$ strongest scattering centers in the test image to access the look-up table and generate votes for the appropriate object, azimuth, range and cross-range translation. Constraints are applied to limit the allowable percent difference in the magnitudes of the data and model scattering centers to $\pm L\%$. (Usually, the design parameters $N$ and $L$ are optimized, based on experiments, to produce the best recognition results; however, for convenience, values of $N=40$ and $L=10\%$ were used throughout this research. Given the MSTAR targets are ‘centered’ in the chips, a $\pm 5$ pixel limit on allowable translations is imposed for computational efficiency.) To accommodate some uncertainty in the scattering center locations, the eight-neighbors of the nominal range and cross-range relative location are also probed and the translation results are accumulated for a 3x3 neighborhood in the translation subspace. This voting in translation space, in effect, converts the consideration of scatterer pairs back into a group of scatterers at a consistent translation. The recognition process is repeated with different scattering centers as basis points, providing multiple ‘looks’ at the model database to handle spurious scatterers that arise due to articulation, occlusion or configuration differences. The recognition algorithm actually makes a total of $9N(N - 1)/2$ queries of the look-up table to accumulate evidence for the appropriate target type, azimuth angle and translation. The models (labels with object, azimuth, etc.) associated with a specific look-up table entry are the “real” model and other models that happen by coincidence, to have a scatterer pair with the same (range, cross-range) relative distance. The constraints on magnitude differences filter out many of these false matches. In addition, while these collisions may occur at one relative location, the same random object-azimuth pair doesn’t often keep showing up at other relative locations with appropriate scatterer magnitudes and mapping to a consistent 3x3 neighborhood in translation space, while the “correct” object does. The basic decision rule used in the recognition is to select the object-azimuth pair (and associated “best” translation) with the highest accumulated vote total. (The details about the basic model construction and recognition algorithms for the recognition system are given in the paper by Bhanu and Jones.\textsuperscript{1})

3. SYNTHESIS OF COMPOSITE MODELS

The objective is to automatically synthesize effective class models of objects with various configurations for subsequent use in the SAR recognition system. For example, given two configurations of objects in the MSTAR data (such as T72 tank serial number (#) 132 and #812, as well as BMP2 armored personnel carrier #C21 and #9565), construct synthetic composite T72 and BMP2 models for use in subsequent tests to recognize a third configuration (T72 #S7 and BMP2 #9566). Since we model objects at $1^\circ$ azimuth increments, we synthesize a specific class model of an object at a particular azimuth. Because all azimuths are not available in the MSTAR data, only azimuths where both configurations of the same object exist are modeled and only azimuths where all three configurations exist are used for testing.

Our basic approach is to use genetic algorithms (GA) to generate and evaluate composite recognition models. GAs are an evolutionary computation technique, inspired from the principles of natural selection, that are an efficient way to automatically conduct a search of a high dimensional solution space, especially when a closed form solution is not explicitly known. The overall schematic for a GA is shown in Figure 1. Basically, an initial population is subjected to a fitness evaluation, then if the fitness of the best individual is acceptable (or in our case, a given number of generations has passed) the process stops. Otherwise, a new population is generated by crossover and mutation operations and the process continues for another generation.

Each individual in the population is a synthetic model with 40 scattering centers, where each scatterer is a triplet of range, cross-range location and magnitude. The scattering center triplets are maintained in order by descending magnitude for convenience with the recognition engine. The population size is fixed at 100 individuals. Each of the individuals in the initial population is chosen in their own random draw of 40 scatterers from a pool of 80 scattering centers that comprise the strongest 40 scatterers from each of the two configurations of interest. To ensure that this pool of scatterers is consistent, an initial
pre-processing step (involving using the recognition engine to find the optimum translation between the two configurations) is used to remove the relative translation between the two configurations. If any of the 40 scatterers in an initial draw are duplicates (same or adjacent locations), then one of the duplicates is discarded and additional random draws are made until 40 legal non-duplicate scatterers are obtained for that individual.

The key idea for the GA to generate composite models is that we use the SAR recognition system to evaluate the fitness of candidate synthesized models with respect to the two given training configurations of the object of interest. For example, we evaluate the fitness of synthetic T72 models (at some particular azimuth angle) using a recognition engine with models of T72 configuration #132 and configuration #812 (at that same azimuth). The fitness evaluation of an individual in the population is the sum of the votes for the two configurations (at the appropriate azimuth) that the individual receives from the recognition engine.

A new population is generated using standard genetic algorithm operations for crossover and mutation with some modifications for this specific application. We apply elitism to crossover (where only a fitter child will replace a parent) and mutation (only a fitter mutant replaces the original) in order to prevent destructive effects. Single-point crossover is used (with a probability of 0.6) and scatterer triplets are kept intact. Because of the constraint that two scatterers can not be in the same location (or immediately adjacent), if crossover or mutation produce duplicates, then one of each of the duplicates is discarded and additional random draws are made from the initial pool of 80 scatterers until 40 legal non-duplicate scatterers are obtained for that individual. Mutation (with a probability of 0.01) is applied to scatterers and if a scatterer is selected for mutation all three of its features are changed randomly.

Figure 2 shows the best fitness for composite models of the T72 and BMP2 (averaged over 103 and 108 available azimuths, respectively) as a function of the number of generations. In both cases the best initial fitness is relatively good, because we drew from an initial pool of scatterers that represented one or the other of the two configurations. The elitism technique prevented destructive effects of crossover and mutation so the fitness never declined. The fitness continually improved out to at least 50 generations, although the rate of improvement from 30 to 50 generations was very small (average 0.5 votes per generation). The fitness of the composite BMP2 model in Figure 2(b) is initially and consistently
higher than the composite T72 model in Figure 2(a). The higher initial fitness implies that there is more similarity between the two BMP2 configurations used in training than the two T72 configurations. Based on the higher overall fitness, we would expect the composite BMP2 model to be a better representation of the two BMP2 training configurations than the T72 composite model is a representation of the T72 training configurations.

**4. EXPERIMENTAL RESULTS**

Figure 3 compares the average similarity of the evolving synthesized composite models of the T72 and BMP2, as well as models using the training configurations (T72s #132 and #812, BMP2s #C21 and #9565), to their respective test configurations: T72 #S7 and BMP2 #9566. The recognition system is used to measure the similarity, which is in votes received by the particular model at the appropriate azimuth. The results in Figure 3 are an overall average over the azimuths that are available for all three configurations of the T72 and BMP2 (49 and 53 azimuths, respectively). The vote ratio is defined as votes normalized by the average number of votes received by the two training configurations. Figure 3(a) shows that, on average, T72 #S7 has more similarity to #132 than #812. In Figure 3(b) BMP2 #9566 is more similar to #9565 than #C21, but the differences among the BMP2s are much less than among the T72s.

For both the T72 and the BMP2, the composite models were able to achieve better results than both the average of the two training configurations (a 1.0 vote ratio) and the best of the two training configurations (T72 #132 and BMP2 #9565). However, the behavior of the composite model curves in Figure 3, where the results broadly peak and then decline with increasing number of generations, indicates that in this application the genetic learning technique is overfitting the training data. This is not unusual with many applications and learning approaches. Typically after learning with training data sets, some other data is then used to determine when to stop the learning to avoid overfitting. Since these results show that the configuration differences of the T72 are a harder problem than the BMP2, we used Figure 3(a) to pick 15 generations to avoid the composite models overfitting the data.

Figure 4 shows the similarity of the various T72 models (#132, #812 and a 15 generation composite model) to T72 #S7 for the 49 azimuths where all three T72 configurations exist in the dataset. While
Figure 3. Average similarity of various models to test configurations.

Figure 4. Model similarity to T72 #S7 (composite model at 15 generations).

Figure 3(a) shows that on average T72 #132 is a better model than #812, Figure 4 shows that #132 is better than #812 for only 65% of the azimuths. The synthesized composite T72 model is better than the best training configuration (#132) for 61% of the azimuths, but better than both configurations (#132 and #812) for only 45% of the azimuths.

Table 1 gives three confusion matrices that compare the forced recognition performance of models using the two training configurations with the results using 15 generation composite models of the BMP2 and T72. Excellent recognition results were achieved for the BMP2 and the results using BMP2 #9565 as a model were better than when using #C21. The results for the T72 using #132 as a model were better than both #812 and the composite model. The resulting 0.9216 probability of correct identification (PCI) for the composite models was better than the 0.9068 average PCI of the two sets of training models, but not as good as the 0.9412 PCI for the best set of training models (BMP2 #C21 and T72 #132). Correct recognition is achieved in these results, if the test object is recognized, even though it is not necessarily at the correct azimuth pose. Figure 5 shows the pose accuracy of the recognition results using the composite models and also the best set of training models. The pose accuracy results
Table 1. Forced recognition confusion matrix comparisons.

<table>
<thead>
<tr>
<th>test targets [serial number]</th>
<th>Identification results (configuration modeled)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BMP2 (#C21)</td>
</tr>
<tr>
<td>BMP2 [#956]</td>
<td>52</td>
</tr>
<tr>
<td>T72 [#S7]</td>
<td>5</td>
</tr>
<tr>
<td>PCI</td>
<td>0.9412</td>
</tr>
<tr>
<td>Average</td>
<td>0.9068</td>
</tr>
</tbody>
</table>

Figure 5. Pose accuracy.

allow for a 180° directional ambiguity (e.g. if the test azimuth is 30°, a result of 211° is a 1° error). The composite model results in Figure 5 have higher percentage with the correct pose (32.4%) than the training models (29.4%) and up to 4.9% more with small errors of 3 degrees or less (70.6% vs. 65.7%).

5. CONCLUSIONS AND FUTURE WORK

A genetic algorithm based approach can successfully build local feature based composite class models that capture salient characteristics of vehicle targets in SAR imagery. The resulting synthesized composite class models can be better representatives than individual configuration models. While the current composite class models provide better than average identification results, they are not the best for a single test configuration. Because the test T72 #S7 configuration is more similar to #132 than #812, a composite model that incorporates characteristics of #812 is less similar to #S7 than #132. However, there are 11 different configurations of T72s available in the MSTAR data and, in the future, one would not expect that a single T72 configuration would continue to be better than a synthesized composite model, because a single T72 configuration is less likely to resemble many different configurations than a composite T72 model that is derived from several configurations. Another significant extension of the current class model approach would be to not only capture the similarities among configurations of the same object, but also to emphasize the differences between objects.

ACKNOWLEDGMENTS

This work was supported by AFOSR grant F49620-02-1-0315; the contents and information do not necessarily reflect the position or policy of the U.S. Government.
REFERENCES


