# MODEL-BASED ATR USING SYNTHETIC APERTURE RADAR

*Robert Hummel, DARPA/SPO, Arlington, VA, USA*

## Introduction

The Moving and Stationary Target Acquisition and Recognition (MSTAR) program was initiated by the U.S. Defense Advanced Research Projects Agency (DARPA) and the U.S. Air Force Research Laboratory (AFRL) in the summer of 1995. The goal of this project was to advance the state of Automatic Target Recognition (ATR) using synthetic aperture radar (SAR) imagery by developing the technology of model-based vision. Now, in the year 2000, the project has largely achieved its goals, and is prepared to transition technology to operational systems that exploit SAR imagery. While new approaches and technology development for the purpose of target recognition will continue to be developed and pursued in other programs, the MSTAR program provides a case study in the progress that can be achieved through a concerted effort. The central ingredients to this effort were:

- Data collections, and sufficient data at a sufficient resolution;
- System development in modules, together with incremental build cycles;
- Thorough performance evaluations with specific goals and milestones.

The keystone to the project was a synthetic SAR prediction capability, which permits the generation of synthetic SAR images from parameterized target models. This latter piece relies, in part, on the Xpatch software developed in cooperation with other programs.

This paper provides a retrospective discussion of the progress made in the course of the MSTAR project.

## ATR Problem

The automatic target recognition (ATR) problem addressed by the MSTAR is the recognition of ground military vehicles in one-foot resolution SAR imagery. Figure 1 shows a SAR scene with

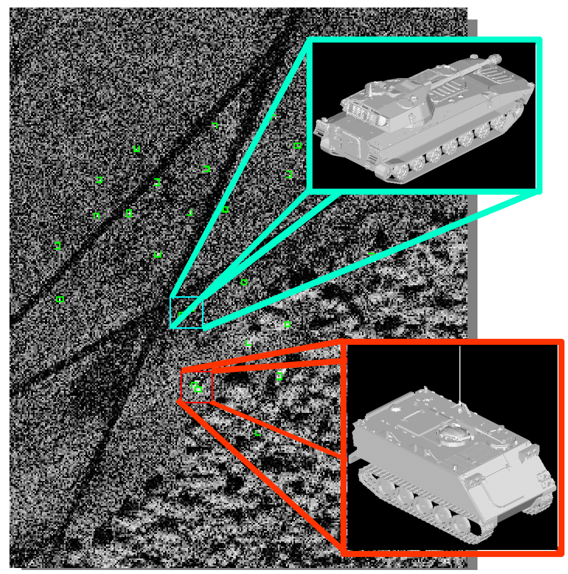

**Figure 1.** The recognition problem in synthetic aperture radar data.

embedded vehicular targets, and depicts recognition of a couple of those targets. SAR imagery was collected using X-band sensors operating in small aircraft, observing carefully ground-truthed military vehicles. In order to achieve one-foot resolution, the radar imager must have approximately 600 MHz of bandwidth, and must observe the target array in spotlight mode for approximately 2.85 degrees of arc. Eventually, hundreds of thousands of instances of targets were imaged, and hundreds of square kilometers of background imagery (without target vehicles) were also collected. The background imagery is termed “clutter” data. Some of this data was approved for public release, and made available to the general public; other data was made available only to MSTAR developers and others under contract to the U.S. government, and still other data was held "sequestered" and used only for evaluation purposes.

Figure 2 shows a close-up of a target in the SAR imagery, still at one-foot resolution. The ATR problem can be summarized by the task of converting the observed signature (as in Figure 2) into a 3-D description of the vehicle that has been imaged.

With one-foot resolution, the typical vehicular target will have several hundred pixels in the target region. In addition, there will be information in the shape of the shadow. With coarser resolution, say one-meter resolution, there will only be a few dozen pixels on the typical target. With hundreds of pixels, one can reasonably expect to extract dozens of "features," such as locations of peaks, shape information, and locations of internal edges. These are among the features that MSTAR uses to perform recognition.

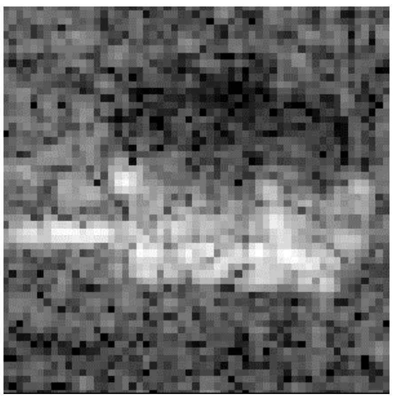

**Figure 2.** A target signature.

# Architecture

For computational efficiency, an ATR system will generally be constructed hierarchically. That is, rather than processing each pixel in the image uniformly, testing every possible target hypothesis, a system can dismiss most pixel locations from consideration by applying ATR filters. The MSTAR system consists of three stages: The "Focus of Attention" module, the "Indexer," and the "PEMS" subsystem (Figure 3). "PEMS" stands for "Predict, Extract, Match, and Search," and refers to an iterative subsystem that refines and verifies target identifications by making use of a prediction capability using high-fidelity models of targets. We discuss each of the three stages below.

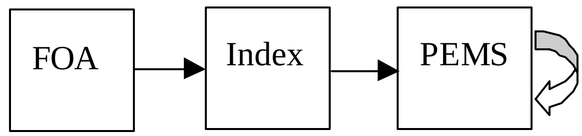


**Figure 3.** The MSTAR Architecture.

The "Focus of Attention" (FOA) stage filters out all but "interesting" regions of the input image. It does this by looking for bright regions in the image, and assuring that those bright regions have the size and shape characteristic of targets. The output of the FOA stage is a set of "regions of interest" (ROI's). Even in "clutter" imagery with no targets in the scene, we expect the FOA stage to produce a stream of ROI's. The rate of ROI's that do not actually contain true targets is the "False Alarm Rate" (FAR) of the FOA module. The FAR will depend on the complexity of the clutter data. This stage is in fact composed of several sub-stages. The first is a standard "CFAR" algorithm, which provides a "constant false alarm rate," and looks for regions that are sufficiently brighter than the surround, where the surround is chosen as an annulus that surrounds a target-size region. Subsequent sub-stages examine global features of the putative bright region, such as the size, shape, contrast, and fill ratio, and perform classic pattern recognition discrimination functions so as to filter out clutter data, without eliminating true targets.

Notably, the FOA module makes certain assumptions about the environment of the target. A particular difficulty arises if the target is in close proximity to another vehicle, or to clutter that causes the bright region of the target response to be adjacent to other bright return. In order to handle proximity, the FOA algorithm should pass as regions of interest bright regions that are target-size or larger. Depending upon how well one filters out large-size clutter, this will increase the processing load on subsequent stages.

The Index stage, also known as the "Indexer," examines each region of interest (ROI) to produce a list of hypotheses as to the target type and orientation of each target hypothesis. The Indexer accomplishes this function in much the same way that a standard ATR produces an identification result. That is, it compares the observed signature in the region of interest with a set of pre-stored signature "templates." The signature information is not necessarily the complex image values, nor even the magnitude of those values, in the SAR image. In the MSTAR system, the features used in the comparison are the zero-crossings of the "Laplacian of Gaussian" (LoG) of the imagery magnitude data, and also ridges in the magnitude data. These features are extracted in the "target region" extracted from the observed image, as well as features that have been previously extracted from the pre-stored exemplars. The exemplars cover all "mission" targets (i.e., possible hypothesis target types), as well as samples of the possible aspect and depression angles at which the targets can be observed. Features are compared using a "two-sided" metric, which first compares the target region of the observed image with the pre-stored data, and then compares the stored features from the exemplar target region with the observed data. The comparison makes use of a penalty function that measures the degree of match of the features, and also provides for some robustness by smoothly dropping the penalty when the distance becomes large. The Indexer takes the scores that it obtains by comparing the data in an ROI against all possible hypotheses, and producing a list of candidate hypotheses. If

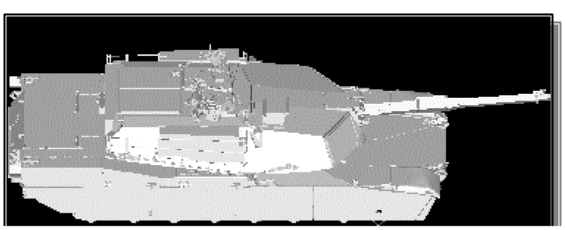

**Figure 2.** A CAD model of a target.

the top candidate has a sufficiently high score, and its nearest competition is sufficiently small, the Indexer module can make the call without any further processing. If there is some ambiguity, then the list of hypotheses with scores above some threshold are passed to the final stage.

The PEMS module then examines operates on each ROI and the associated list of hypotheses. The purpose of the PEMS module is to reorder the hypothesis list by reasoning about the three-dimensional structure of the target environment, and to optimize the match between each candidate hypothesis and the observed data. Notionally, this requires an iterative process with each hypothesis, to optimize the position, aspect, configuration, and other parameters that compose the target hypothesis, and affect the expected observed data. The fundamental tenet of the MSTAR system is that it is necessary to reason about the structure of the target in order to achieve robustness to "extended operating conditions" (EOC's). Example EOC's include:

- Position, aspect, squint;
- Configuration variations;
- Manufacturing variations;
- Ground conditions;
- Articulation parameters;
- Revetments, obscurations, and other neighboring confounding conditions.

The PEMS system needs to consider not only these EOC's, but combinations of EOC's. Of course, when it is possible to perform recognition using features that are independent of a particular EOC, then it is desirable to use those features. However, variability due to EOC's is greater than the separation of the target signatures due to target type differences, and thus it fundamental in the MSTAR philosophy that a certain amount of reasoning, particularly three-dimensional reasoning, is critical to robust, accurate identification.

Since the PEMS module is so fundamental to the MSTAR system, we discuss this module in greater detail in the next section.

## PEMS

The PEMS stage depends on a SAR signature prediction capability. We explain this capability, which is embodied in the "Predictor" module, very briefly below; algorithmic and mathematical details are beyond the scope of this brief summary.

The Predictor module begins with a CAD model (computer-aided design model) of a target, as depicted in Figure 4. Typical CAD models that are used by the Predictor can have a million triangular facets. These models are obtained by "climbing over top" of an example target, recording accurate three-dimensional positions, and facetizing the resulting structure. The modeling process requires an instance of the target, and a labor-intensive facetizing process. As a result of the DARPA-funded "Rapid Target Model Insertion" program, we have examples of the construction of a CAD model in two weeks.

Using the CAD model, the Xpatch software code is applied to produce synthetic SAR data from a discretized collection of viewpoint aspects. The Xpatch code uses the "Shooting and Bouncing Ray" approach to computational electromagnetics. The far-field response of the target is computed based on shooting an array of rays at the target, and understanding the response by performing a ray-tracing process to locate the exit path of each ray. The result is pruned to produce a collection of "scatterers" in 3-D space, together with tags as to which aspect directions in which the scatterer can be viewed. Further, each scatterer contains information as to the rays that contributed to the response, which ultimately permits trace-back from pixels in the SAR image to facets in the target model.

Once the intermediate data structure of scatterers is formed, an online prediction can be computed. The online predictor accepts as input the "pose" of the target model, and parameters such as articulation and configuration variations. By manipulating the location of the scatterers in 3-D space, a synthetic SAR image can be formed rapidly. Although the prediction fails to take into account large-scale interactions between components of the target model, extensive validation has shown that the predictions are useful for object identification.

Figure 5 shows an example of a predicted and observed target pair. Note that the prediction is based purely on a CAD model, whereas the observed data is an actual target sitting in a field, imaged by a true SAR imaging system. Both images are samples from probability distributions, and so one cannot expect identical pixel values. However, the qualitative information is reflected in the predicted image, and features such as peak locations, edges, and shapes tend to match up well.

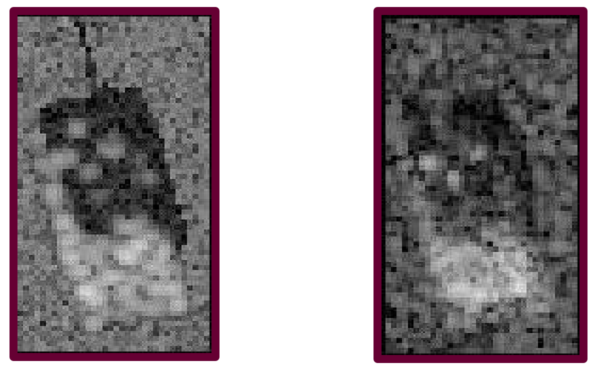

**Figure 3.** A predicted (left) and observed (right) SAR image of a target.

The PEMS module uses predictions to compare with extracted features. The comparison is performed by the matching module, which is discussed briefly in the next section. Where there are mismatches, the Search module can then use cues to posit changes to the hypothesis parameters in order to refine the prediction. For each hypothesis, several iterations are applied in order to assess the target hypothesis in light of potential EOC's.

When all hypotheses presented by the Indexer have been considered, the Search module makes a target type call. One of the possible calls is that the ROI contains something other than a mission target. That is, the upshot of the PEMS stage can be a label of "Other."

## Matching

There are many methods for comparing a predicted set of features with an observed collection of features. Even using a Bayesian model for evaluating a probability of a match given the observed feature, there are many choices that are possible. The MSTAR system uses a "Diffusive scattering model" for explaining how observed features are expected to occur given predicted features. Suppose that the prediction X consists of a set of "features:" $X = \{X_1, \cdots, X_n\}$. In the MSTAR model, each predicted feature $X_i$ represents a spatially varying Poisson distribution, so that zero, one, two, or more features will be generated, with each feature chosen according to a spatial distribution $g_i(\mathbf{x} - \mathbf{x}_i)$. The spatial distribution is a density function with unit mass, so that $\int g(\mathbf{x}) d\mathbf{x} = 1$. The expected number of features generated by prediction $X_i$ is $l_i$, and so the probability that $k$ features will be generated by $X_i$ is given by $e^{-l_i} \cdot l_i^k / k!$. Thus $X_i$ consists of the information $(l_i, g_i, \mathbf{x}_i)$. Generally $l_i$ is quite small, so that usually either zero or one feature will be generated, but it is entirely possible for multiple observed features to be associated with a single prediction. We also posit a background distribution of $g_0(\mathbf{x})$ with an expected number of spurious features equal to $l_0$. Taken together, the entire set of feature generators turns out to be a single Poisson process, with an expected number of features equal to $l = l_0 + \sum_{i=1\cdots n} l_i$, and density distribution

$$\frac{l_0 g_0(\mathbf{x}) + \sum_{i=1\cdots n} l_i g_i(\mathbf{x} - \mathbf{x}_i)}{l_0 + \sum_{i=1\cdots n} l_i}.$$

Accordingly, the density distribution function, for an observed set of features $Y = \{\mathbf{y}_1, \cdots, \mathbf{y}_m\}$, is given by:

$$f(Y \mid H) = \frac{1}{m!} e^{-l} \prod_{j=1\cdots m} \left[ l_0 g_0(\mathbf{y}_j) + \sum_{i=1\cdots n} l_i g_i(\mathbf{y}_j - \mathbf{x}_i) \right].$$

For a given set of extracted features Y, this density function provides an unnormalized score for the match of $Y$ with the prediction $X$. Normalization is required to provide some stability to the scores, and can be provided by taking the ratio of the unnormalized score with a hypothesis that the observed data is "clutter."

## Results

Thorough evaluations of the MSTAR system have been provided by an Evaluation Team, which scores the system against sequestered data, carefully recording results as a function of the conditions of the test imagery. Sufficient numbers of tests in similar conditions (a "bin") must be evaluated in order to provide statistically meaningful numbers. As a result, hundreds of thousands of instances of targets have been run through the system. At the same time, hundreds of square kilometers of clutter data have been processed.

Sadly, it is impossible to give a single number, or even just a few numbers, to state system performance. Even "ROC" curves are inadequate. It is a truism that performance evaluation measures as much (if not more) about the test conditions as the quality of the system. Accordingly, we make some qualitative remarks here, and refer the interested reader to the MSTAR Book (to appear in 2000) for more details.

Certain subproblems have been "solved," as evidenced by the MSTAR system. If the targets are nominally configured, not subjected to "EOC's," and are thus "in the clear," they can be detected and identified in one-foot resolution SAR, using a single view, with near certainty. This statement remains true up to 30 or so targets in the mission set, and there is good reason to expect that it remains true for much larger mission sets. If the system is restricted to single-target missions, such as a mission of finding all instances of Scud missile launchers, near certain detection and identification is again generally possible.

Certain EOC's are easier than other EOC's. Articulation of the turret and barrel of the M109 causes serious variability to the signature, but the MSTAR system has demonstrated a good ability to reason about M109 articulation. Other targets that articulate often can be handle through invariance.

Configuration changes, such as the attachment of fuel barrels to a tank, can also be handled successfully. Target variants have also caused few problems. The system has been shown to work, with few or no changes, on a variety of sensor sources.

Obscuration by walls, and targets in close proximity to one another, remain challenges, and are research topics in the current year 2000. Approaches to treat these conditions have been identified.

False alarms rate vary according to the complexity of the background, and system settings that depend upon the desired detection rates. When set so that most targets (90%) of targets including those with considerable EOC's are detected, FAR can be as high as one per square kilometer. When the system is set so that 50% of the easy targets are detected, false alarms can be eliminated.

Under all these conditions, correct identification rates in the range of 80% to 90% are typical, for those EOC's that "work." Whenever the system explicitly reasons about an EOC, system performance tends to be good. Problems are generally confined to cases where the system attempts to achieve robustness by ignoring variability. This conclusion largely validates the notion that model-based technology holds the key to transition of useful ATR to operational status.

## Acknowledgements

Many people at Air Force Research Labs Sensor Directorate were instrumental in management and shepherding of the MSTAR program. In addition, MSTAR evaluations were performed by a team that included personnel at AFRL and Sverdrup Technologies. Special thanks go to Tim Ross, Mark Minardi, and John Mossing.

There were many developers who were instrumental to the design and execution of the MSTAR system. If any are mentioned, then many will be slighted. Fortunately, the MSTAR Book will acknowledge the work of all key people on the program. Of course, there are many people who assisted in the development of ATR technology over the years, and the MSTAR system builds upon this work, and continues to borrow liberally from developments in the Image Understanding community. Work by Tom Binford deserves special mention: Model-based technology owes much to his insights.