Flexible Histograms: A Multiresolution Target Discrimination Model

Jeremy S. De Bonet, Paul Viola & John Fisher

Learning & Vision Group
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
545 Technology Square
Cambridge, MA 02139

ABSTRACT

In previous work we have developed a methodology for texture recognition and synthesis that estimates and exploits the dependencies across scale that occur within images.[1,2] In this paper we discuss the application of this technique to synthetic aperture radar (SAR) vehicle classification. Our approach measures characteristic cross-scale dependencies in training imagery; targets are recognized when these characteristic dependencies are detected. We present classification results over a large public database containing SAR images of vehicles. Classification performance is compared to the Wright Patterson baseline classifier.[3] These preliminary experiments indicate that this approach has sufficient discrimination power to perform target detection/classification in SAR.

1. INTRODUCTION

The detection and classification of targets in imagery is a difficult problem. Any successful algorithm must be able to correctly classify the very wide range of images generated by a single target class. The sources of image variations in SAR are many: target rotation, translation, articulation, sensor noise, depression angle, overlay, camouflage, and many others. Typically classifiers attempt to deal with these variations in one of two ways: invariance and duplication. For example, translation invariance could be achieved by only measuring the distance between detected scatters, since distance is invariant to translation. However, such a method does not work as well as the target rotates, since scatters may appear and then disappear. Alternatively, a duplicative technique could be used at the cost of additional computation – by using several separate models for a single target class, each tuned to a different target orientation.

In this paper we will compare with a baseline classifier proposed and implemented by a group at Wright Patterson AFB.[3] Their intent was to provide a reasonable standard for comparison of new SAR target recognition approaches. Their approach is duplicative: each target class is modeled by a number distinct templates each tuned to a different orientation. Targets are recognized if one of the target templates is nearby in the vector space defined by the cells of a SAR image*. The history of templates in recognition is extensive (see Duda and Hart[4] for an early review). Measuring the distance to a template is equivalent to measuring the negative log likelihood of a Gaussian process whose mean value is the template. Since each target is modeled as a collection of templates, the model is essentially a mixture of Gaussians. This is the optimal classification approach when we believe that rotation has been sampled finely enough, and that the other variations in SAR imagery can be modeled as a Gaussian noise. Over many years template matching has proven to be a very effective technique. It is not surprising that the template based classifier yields very good performance. However, it may be infeasible in practice to acquire the large number of example images needed.

In this paper we present an alternative statistical model for the variations in SAR imagery. Models are developed from far fewer target vehicle examples than are needed by template techniques. We will show that it effectively classifies targets in the presence of a difficult set of confusion vehicles. These results have been quantified and compared with the performance of our reimplementation of the Wright Patterson baseline classifier.

EMAIL: `jsd@ai.mit.edu`, `viola@ai.mit.edu` & `fisher@ai.mit.edu`
HOMEPAGE: `http://www.ai.mit.edu/projects/lv`

*In this paper we will refer to the amplitude of the complex signal in a resolution cell as a *cell*.

2. MULTI-RESOLUTION MODELS OF TARGETS

Recently it has been shown that multi-resolution models of statistical processes can be effective for target detection in SAR data.[5–7] These model have decomposed the SAR signal into a number of different images at different scales from coarse to fine. Irving *et al.* define a coarse to fine statistical process in which the distribution of higher frequency data is conditioned on lower frequency data. This conditional distribution is modeled as a Gaussian. Using these conditional distributions a multi-scale Markov chain is defined which can model complex dependencies both across scale and space. A target detection algorithm is constructed from a pair of such distributions, one for targets and one for clutter. Each statistical model is trained from example data: a set of chips containing targets and another set of chips containing only clutter. From these chips multi-scale representations are computed, cross-scale conditional distributions are observed, and the parameters of the cross-scale Gaussian distribution are estimated.

In many ways our classification approach is a generalization of these multi-scale detection algorithms, with three key differences: i) our multi-resolution representation separates and distinguishes information represented at different orientations; ii) we consider full (non-Markovian) cross-scale conditional distributions; and iii) our cross-scale model is non-parametric and as a result, can model non-Gaussian processes.

2.1. A Coarse to Fine Generative Process

As in previous work[6,7] the generative model that underlies our statistical classification approach is a conditional probability distribution across scale. This probability chain defines a statistical distribution for the entire SAR signal (including information at every frequency). This is accomplished by modeling the dependency of the highest frequencies on lower frequencies.

Unlike previous approaches we define the conditional chain on a multi-scale wavelet transformation. The wavelet transform is most efficiently computed as an iterative convolution using a bank of filters. First a "pyramid" of low frequency downsampled images is created: $G_0 = I$, $G_1 = 2\downarrow(g \otimes G_0)$, and $G_{i+1} = 2\downarrow(g \otimes G_i)$, where $2\downarrow$ downsamples an image by a factor of 2 in each dimension, $\otimes$ is the convolution operation, and g is a low pass filter. At each level a series of filter functions are applied: $F^i_j = f_i \otimes G_j$, where the f_i's are various types of filters (see Figure 1). Computation of the F^i_j's is a linear transformation that can thought of as a single matrix W. With careful selection of g and f_i (as done in[8]) this matrix can be constructed so that $W^{-1} = W^\top$. Computation of the inverse wavelet transform is algorithmically similar to the computation of the forward wavelet transform.

For every resolution cell in an SAR data we define the *parent vector*:

$$
\begin{aligned}
\vec{V}(x,y) = \; & \left[F^0_0(x,y), F^1_0(x,y), \cdots, F^M_0(x,y),\right. \\
& F^0_1\left(\left\lfloor \tfrac{x}{2} \right\rfloor, \left\lfloor \tfrac{y}{2} \right\rfloor\right), F^1_1\left(\left\lfloor \tfrac{x}{2} \right\rfloor, \left\lfloor \tfrac{y}{2} \right\rfloor\right), \ldots, F^M_1\left(\left\lfloor \tfrac{x}{2} \right\rfloor, \left\lfloor \tfrac{y}{2} \right\rfloor\right), \ldots, \\
& \left. F^0_N\left(\left\lfloor \tfrac{x}{2^N} \right\rfloor, \left\lfloor \tfrac{y}{2^N} \right\rfloor\right), F^1_N\left(\left\lfloor \tfrac{x}{2^N} \right\rfloor, \left\lfloor \tfrac{y}{2^N} \right\rfloor\right), \ldots, F^M_N\left(\left\lfloor \tfrac{x}{2^N} \right\rfloor, \left\lfloor \tfrac{y}{2^N} \right\rfloor\right)\right]
\end{aligned}
\tag{1}
$$

where N is the top level of the pyramid and M is the number of features. The parent vector can be visualized schematically as in Figure 2, in which the image is decomposed from coarsest (on the left) to finest resolution (on the right). Each element in these arrays represents the collected feature responses at that location and resolution.

We define a probability distribution over SAR images as a multiresolution wavelet tree across scale. In this tree the generation of the lower levels depend on the higher levels:

$$
\begin{aligned}
p(\vec{V}(x,y)) = \; & p(\vec{V}_M(x,y)) \times p(\vec{V}_{M-1}(x,y)|\vec{V}_M(x,y)) \\
& \times p(\vec{V}_{M-2}(x,y)|\vec{V}_{M-1}(x,y), \vec{V}_M(x,y)) \\
& \times \ldots \\
& \times p(\vec{V}_0(x,y)|\vec{V}_1(x,y), \ldots, \vec{V}_{M-1}(x,y), \vec{V}_M(x,y))
\end{aligned}
\tag{2}
$$

where $\vec{V}_l(x,y)$ is the a subset of the elements of $\vec{V}(x,y)$ computed from G_l. Usually we will assume ergodicity, i.e. that $p(\vec{V}(x,y))$ is independent of x and y. The generative process starts from the top of the pyramid, choosing values for the $\vec{V}_M(x,y)$ at all points. Once these are generated the values at the next level, $\vec{V}_{M-1}(x,y)$, are generated. The process continues until all of the wavelet coefficients are generated. Finally the image is computed using the inverse wavelet transform.

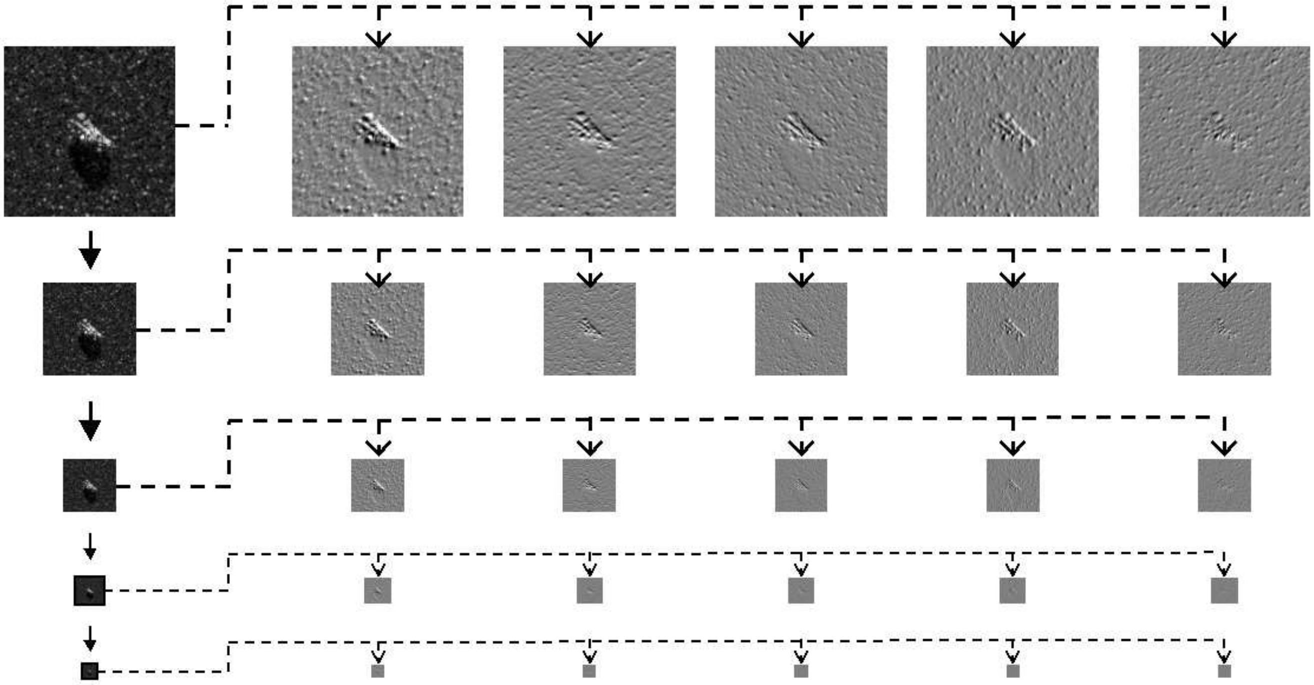

Figure 1. The Wavelet decomposition of a SAR target chip. In the upper left is the original SAR amplitudes.

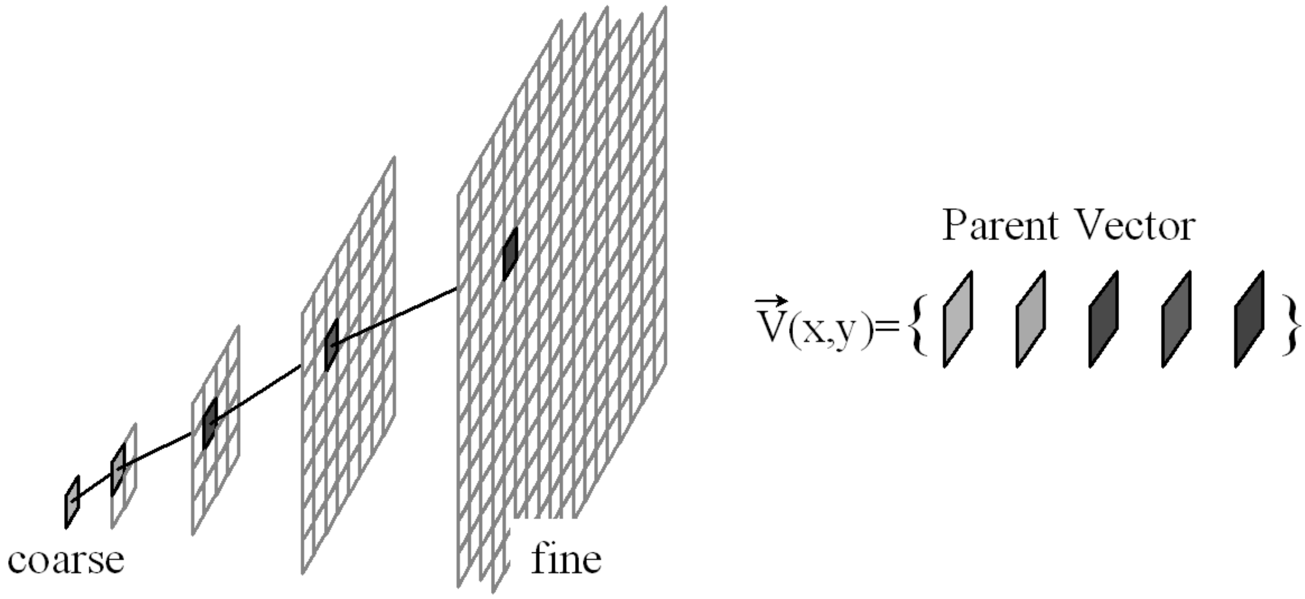

Figure 2. The parent vector. Each cell in these arrays represents the collection feature responses at that location and resolution.

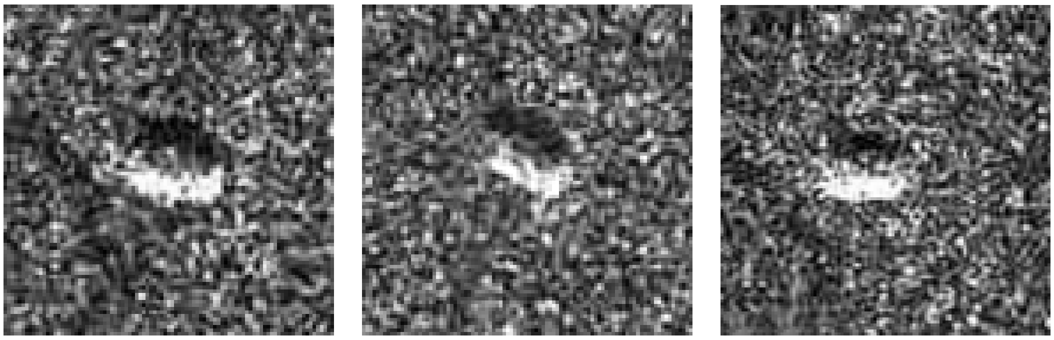

Figure 3. Target chips synthesized by sampling directly from the probabilistic distribution generated by the flexible histogram model generated from several example vehicles.

It is important to note that this probabilistic model is not made up of a collection of independent chains, one for each $\vec{V}(x,y)$. Parent vectors for neighboring cells have substantial overlap as coefficients in the higher pyramid levels (which are lower resolution) are shared by neighboring cells at lower pyramid levels. Thus, the generation of nearby cells will be strongly dependent.

2.2. Estimating the Conditional Distributions

The conditional distributions in equation (2) must be estimated from observations. We choose to do this directly from the data in a non-parametric fashion. Given a sample of parent vectors $\left\{\vec{S}(x,y)\right\}$ from an example image we estimate the conditional distribution as a ratio of Parzen window density estimators:

$$p(\vec{V}_l(x,y)|\vec{V}_{l+1}^M(x,y)) = \frac{p(\vec{V}_l^M(x,y))}{p(\vec{V}_{l+1}^M(x,y))} \approx \frac{\sum_{x',y'} R(\vec{V}_l^M(x,y),\vec{S}_l^M(x',y'))}{\sum_{x',y'} R(\vec{V}_{l+1}^M(x,y),\vec{S}_{l+1}^M(x',y'))} \tag{3}$$

where $\vec{V}_l^k(x,y)$ is a subset of the parent vector $\vec{V}(x,y)$ that contains information from level l to level k, and $R(\cdot)$ is a function of two vectors that returns maximal values when the vectors are similar and smaller values when the vectors are dissimilar. We have explored various $R(\cdot)$ functions. In the results presented the $R(\cdot)$ function returns a fixed constant if all of the coefficients of the vectors are within some threshold θ and zero otherwise.

The cross-scale conditional distributions defined in equation (3) can be used to define the distribution of parent vectors in a target image. This directed conditional structure is very useful when one wishes to sample from the distribution – no computationally expensive operations such as Gibbs sampling are necessary (this is in distinct contrast to Markov Random Fields[9] or in the FRAME method[10]). Given this simple definition for $R(\cdot)$ sampling from $p(\vec{V}_l(x,y)|\vec{V}_{l+1}^M(x,y))$ is very straightforward: find all x',y' such that $R(\vec{S}_{l+1}^M(x',y'),\vec{S}_{l+1}^M(x,y))$ is non-zero and pick from among them to set $\vec{V}_l(x,y) = \vec{S}_l(x',y')$.

In previous work we have demonstrated that images generated by sampling from the above distribution are of very high quality.[1,11] By synthesizing new images from models built from SAR imagery, we can visualize the ability of this model to capture the complex structural patterns inherent in SAR. An image synthesized by sampling from this distribution is shown in Figure 3.

3. TARGET CLASSIFICATION

Given a particular distribution over parent vectors, it might at first seem as though classification of a target chip could be best accomplished by computing the likelihood that that chip was generated by each of the multi-scale target models. If target chip is viewed as a single sample from a vector valued random process, this is the optimal Bayesian classification scheme. However, each target chip is actually a collection of *many samples* from the ergodic process characterized by the distribution. The distinction between these two views is subtle, and is crucial for obtaining a powerful classification measure.

Here is a simple example to illustrate this: consider a process which creates random binary images by repeatedly flipping a single coin. Each cell is colored white with an independent probability of 0.75. Suppose you were asked to decide which of two images is more likely to have been generated by this process: the first has 75 white cells and 25 black ones; the second image has 100 white cells (and 0 black). Intuitively, it seems more likely that the first image was generated by this process. But if we treat each image as single sample from a joint distribution, the probabilities indicate otherwise. The probability of generating the first image is much lower than that of the all white image (roughly 3×10^{-25} compared to roughly 3×10^{-13}). Why does this approach fail to pick out the correct image? It does not take into account that the overwhelming majority of samples which are generated by this process will have about 75 white cells. While it is true that the first image is more likely than the second, it is much less *typical*. Formally, typical images are those whose entropy is approximately the same as the entropy of the class to which they belong. The fact that most images are typical is known as the Asymptotic Equipartition Property.[12]

A better way to decide which of these two images was generated by the above process is to measure which is more typical. This is done using the Kullback-Liebler (KL) divergence or *cross-entropy*†. The cross-entropy is a measure of the difference between two distributions:

$$\begin{aligned}
D(p||q) &= \int p(\vec{V}(x,y)) \log \frac{p(\vec{V}(x,y))}{q(\vec{V}(x,y))}\, dx\, dy && (4)\\
&= \int p(\vec{V}(x,y)) \log p(\vec{V}(x,y)) - \int p(\vec{V}(x,y)) \log q(\vec{V}(x,y))\, dx\, dy && (5)\\
&= -H(p) - \int p(\vec{V}(x,y)) \log q(\vec{V}(x,y))\, dx\, dy && (6)
\end{aligned}$$

It can be viewed as the difference between two expected log likelihoods: the log likelihood of samples of $p(x)$ under the distribution $p(x)$, and the log likelihood of samples of $p(x)$ under $q(x)$. For the first image we estimate the probability of white cells to be $p_1 = 0.75$. For the second image set we estimate $p_2 = 1.0$. The true probability of a white cell is $q = 0.75$. For the first image, the cross-entropy is $D(p_1||q) = 0.28$, while the second is $D(p_2||q) = 0.0$, a perfect fit.

We have implemented a cross-entropy discrimination measure, and have tested it both on optical images and SAR data.[2] Here we present a discrete version which is far more efficient computationally.

4. FLEXIBLE HISTOGRAMS: EFFICIENT DIVERGENCE APPROXIMATION

We approximate the distribution in equation (2) by the Parzen density estimator:

$$p(\vec{V}(x,y)) = \sum_{x',y'} R\left(\vec{V}(x,y), \vec{V}(x',y')\right) \quad (7)$$

Where $R(\cdot)$ is the boxcar Parzen density kernel defined by:

$$R(\cdot) = \Theta\left(T - \left\|\left[\vec{V}(x,y) - \vec{V}(x',y')\right](D)^{-1}\right\|_\infty\right) \quad (8)$$

where each component $d_{i,i}$ of diagonal matrix D scales the corresponding feature response; and $\Theta(\cdot) = 1$ if its argument is greater than 0, and is 0 otherwise; Z is a normalization factor. The L_∞ norm ($||\cdot||_\infty$) is equal to the largest difference along any dimension. The value of $p(\vec{V})$ is proportional to the count of parent vectors which fall within a $N \times M$ dimensional hypercube centered at $\vec{V}(I,x,y)$. We can define a histogram whose bins are given by $B(x,y) = p\left(\vec{V}(x,y)\right)$. We call this representation a flexible histogram because the centers of the bins are determined by the the parent vectors that appear in the training data. In this way, when building models for different targets the bins used in the histogramming process are specialized. The process of construction is outlined schematically in Figure 4.

†Cross entropy is not symmetric and is therefore not a metric.

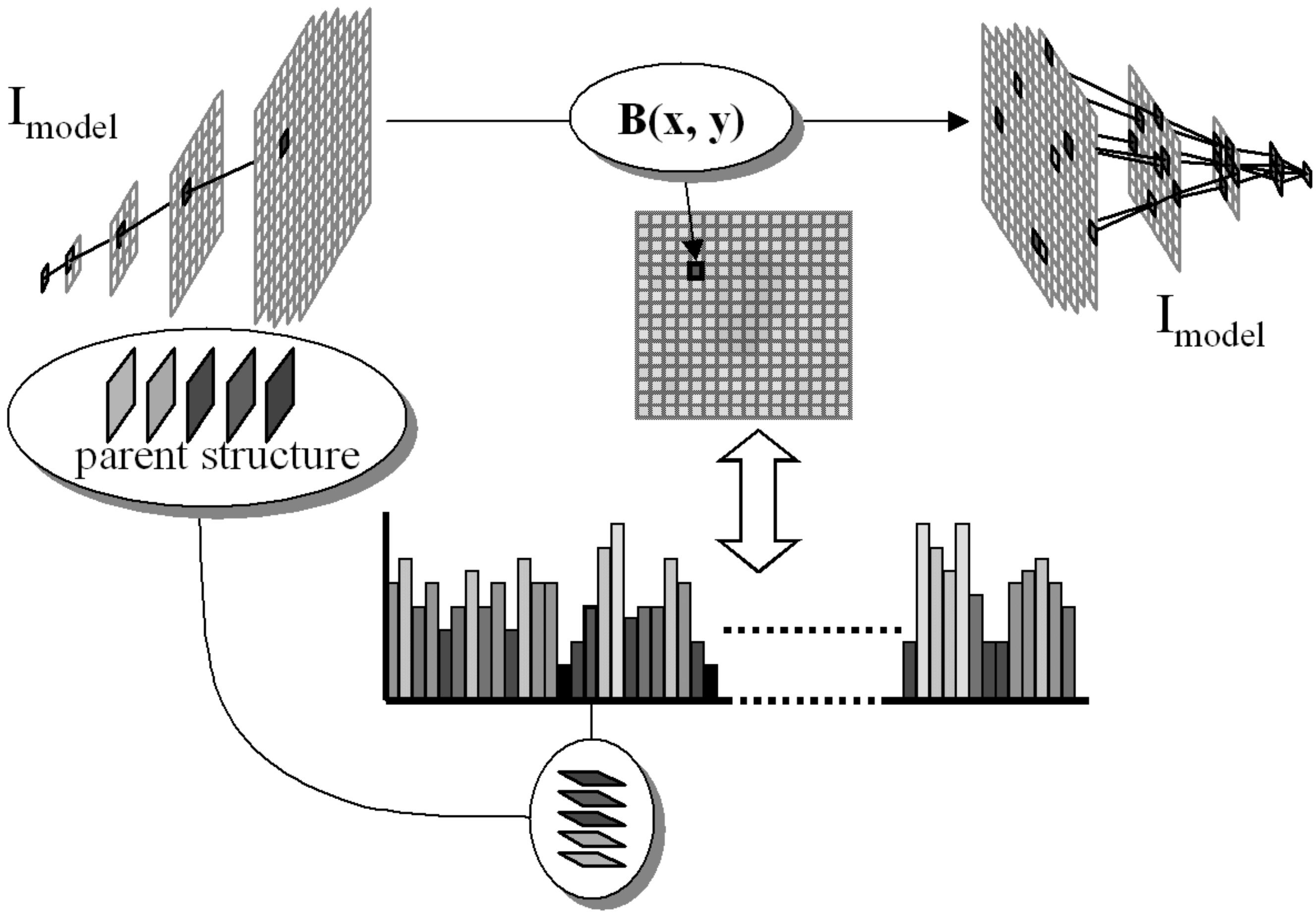

Figure 4. The flexible histogram accumulates the frequency of parent vectors which are within an $N \times M$ dimensional hypercube centered at each parent vector in the model image, to provide a measure of multiresolution redundancy.

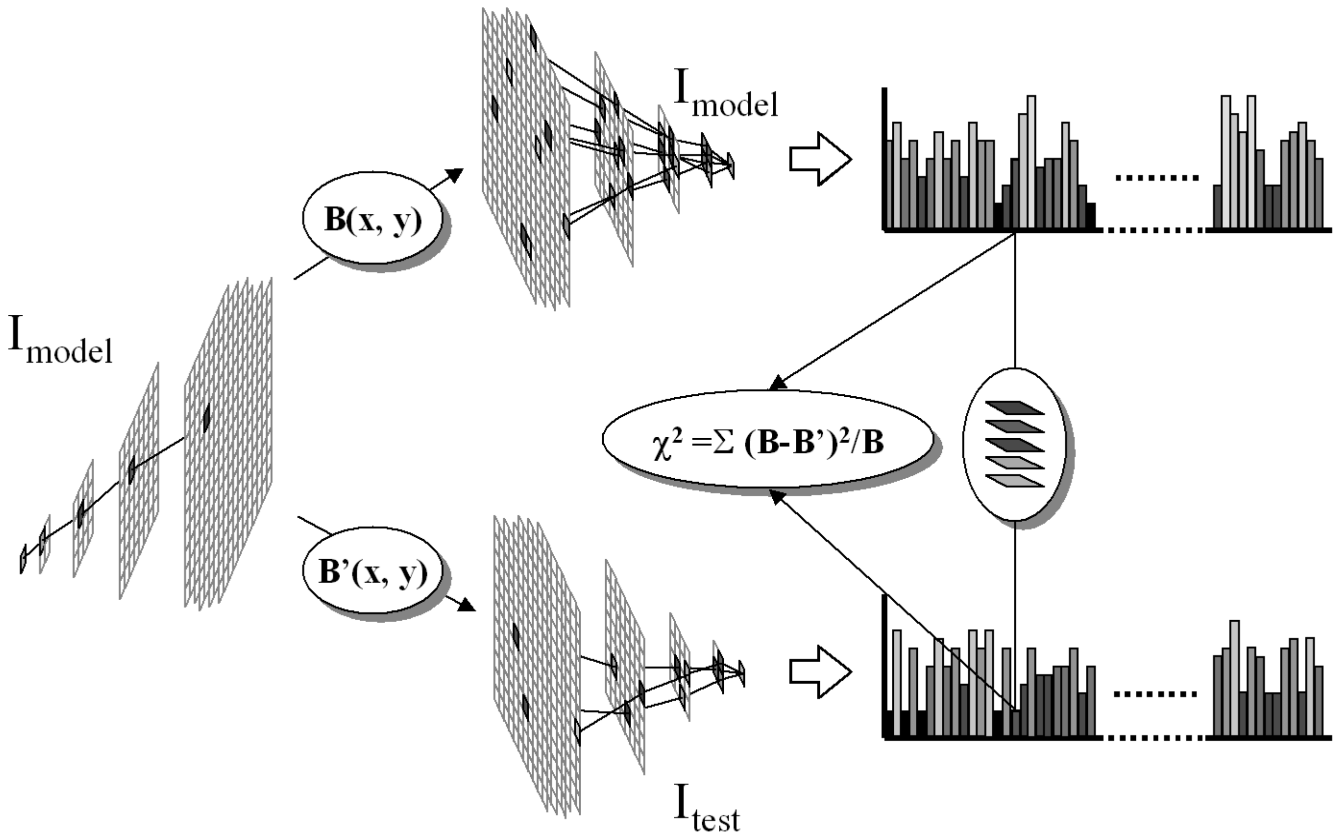

Figure 5. By comparing flexible histograms with the χ^2 measure we obtain a measure of the multiresolution similarity between the model image I_{model} and the test image I_{test}.

By using the boxcar Parzen density kernel in equation (8), a significant computational advantage is obtained. Because the parent vectors for neighboring cells share coefficients at lower resolutions, the computation of $p\left(\vec{V}(x,y)\right)$ is not independent for regions neighboring (x,y). Thus, if a parent vector fails to fall into the histogram bin because some feature at a particular resolution is beyond the threshold distance, then all parent vectors which share that coefficient can be eliminated from consideration. In practice this reduces the complexity of the algorithm from $O\left(N^2\right)$ in the number of cells to an average case complexity of $\Omega(N)$.

The cross-entropy between two distributions can be effectively approximated by a chi-squared (χ^2) test on the bin counts in the flexible histogram (the χ^2 measure is in fact a two term Taylor series expansion of the cross-entropy[12]). Using the χ^2 measure in this way can be viewed as a comparison of Parzen density estimates:

$$\chi^2 = \frac{\left(B(x,y) - B'(x,y)\right)^2}{B(x,y)} \propto \sum_{x,y} \left\{ \frac{\sum_{x',y'} R\left[\vec{V}(I,x,y), \vec{V}(I,x',y')\right]}{\sum_{x',y'} R\left[\vec{V}(I,x,y), \vec{V}(I',x',y')\right]} \right\}^2 \tag{9}$$

where $R(\cdot)$ is the Parzen kernel in equation (8). This process is outlined in Figure 5.

A classification decision is made by comparing the likelihood measure of a target image to a threshold η_{model}. Using standard χ^2 tables a value for η_{model} for any desired level of certainty can be determined. However, we do not assume that the flexible histogram completely describes the target class distribution; choosing a fixed χ^2 likelihood could eliminate true-positives which fall below the chosen threshold, yet which lie far above the false-positives and should therefore be detected. By choosing η_{model} empirically, we maximize the percentages of true-positives while guaranteeing an expected level of false-positives.

5. INCORPORATING MULTIPLE MODEL IMAGES

Given access to multiple example images of the target vehicle, it is reasonable to assume that that a discrimination system should be able to perform better by taking advantage of the additional information. With slight modification, we can extend the single model flexible histogram technique to incorporate information from multiple examples of the target.

Consider now the situation where we have a set model images $\{I_{M0}, I_{M1}, \ldots, I_{Mn}\}$. We begin with the initial flexible histogram generated by accumulating the frequency of similar parent vectors in I_{M0} measured with respect to itself; information from the additional images can be incorporated into this histogram by adding the frequencies for each of the parent vectors in I_{M1} through I_{Mn} which are within the bins defined by I_{M0}. The additional parent vectors thus have the effect of refining the the frequency-counts in the histogram bins, but does not affect the number of bins present. This improves the target model by incorporating the relative frequencies of parent vectors in the additional model images, which will increase the accuracy of the histogram approximation to the distribution. And the flexible histogram becomes a more accurate estimate of the redundancy of the entire *target class*.

Using each of the n images as a base image for determining the bins in a flexible histogram, we can build n such estimators of class membership. Thus we can build a set of histograms $\{B_1, B_2, \ldots, B_n\}$ where histogram B_i is a measurement of the redundancy of each of $\{I_{M0}, I_{M1}, \ldots, I_{Mn}\}$ measured with respect to the parent vectors in I_{Mi}.

6. EXPERIMENTAL SETUP

We now discuss experiments testing the performance of the flexible histogram method for detection and classification of SAR images of vehicles. The images used in this experiment are drawn from a library of SAR images of vehicles measured at X-band. The images are of approximately 1 foot by 1 foot resolution. Our experiments are similar to those described in Velten *et al.*[3] which describes a template based approach to classification and reports results on a database which has high intersection with the one we use here. The primary difference being that the confusion class in these experiments contains a greater number of vehicle types.

In one sense, getting meaningful detection or classification results is problematic. In our experiments we are only considering images of SAR vehicles. Presumably, in a full ATR system, large regions of clutter would be considered and some simpler test (e.g. a CFAR detector) would pass along regions of interest (ROIs) to a classifier. As observed by Velten *et al.*,[3] the set of images we are using represent a more difficult set of ROIs than clutter or other man-made objects and so these experiments are a more stringent test than would be encountered in a practical ATR system. Furthermore, our primary goal is to provide a comparison between two approaches and for that purpose the data set we are using (with its restricted imagery) is perfectly valid. It is from the comparison standpoint that we would like these results to be viewed primarily.

6.1. Template Based Approach Used for Comparison

For comparison purposes, we have implemented the algorithm described in Velten *et al.*.[3] The algorithm described there compares a thresholded and normalized template, M, to a thresholded normalized image chip (or ROI), S. As described, the template, M, is computed as an average over images within a designed aspect range. The detection statistic is the minimum L_1 difference between the template and the ROI computed over all template aspect ranges for a give vehicle type. In our implementation, we compute 36 templates for each vehicle model. Templates are computed using anywhere from three to nine images (from a training vehicle) depending on the number of aspect views that fall within the each 10 degrees of aspect. The total number of images used to compute each vehicle template model was 233 images.

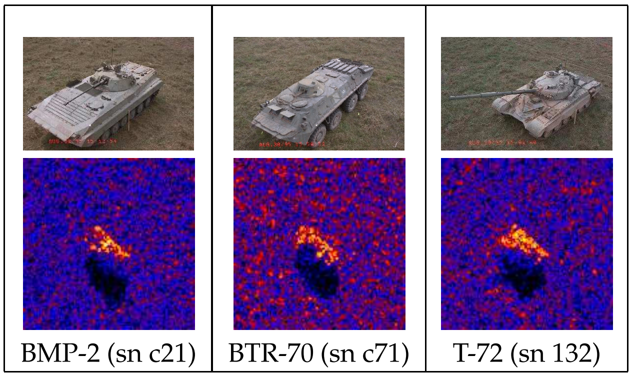

Figure 6. Training Vehicles for Recognition Class.

6.2. Flexible Histogram Approach Used for Comparison

The results described here are based upon models for the target vehicle developed from 10 training images of each vehicle. The images were chosen uniformly in aspect, every 36^o. From these images 10 flexible histograms were generated for each vehicle using the extended procedure described in section 5. The detection statistic is the minimum χ^2 difference between the 10 pairs of flexible histograms generated for the training and testing images (measured with respect to each of the ten training images).

The choice of relatively small number, 10, images to generate each vehicle model is far less than the 233 images used in the template based approach, and was made to reflect a more reasonable estimation of the information available in real applications.

6.3. Description of Training and Test Set

As in Velten *et al.*[3] we consider three vehicle types, BMP-2, BTR-70, and T-72, in our classifier. Figure 6 shows example SAR images of the vehicles used to train the class models for both the template based approach and the flexible histogram approach.

Images of different vehicles (same type of vehicle, different serial number) are used as the test set for the recognition class. Examples are shown in figure Figure 7. The images of training vehicles were collected at a depression angle of 17^o, while performance was tested on images of vehicles collected at 15^o. As stated, testing vehicles have different serial numbers than the vehicles used for training, except in the case of the BTR-70 (only one serial number in the data set). Another set of vehicles are used as confusers and are shown in figure Figure 8.

In this experiment we have a total of 777 independent testing vehicles for the recognition class (exclusive of 196 BTR-70 images which were collected from the same vehicle at different depression angles) and 1643 confusion images, for a total of 2616 images tested in this experiment.

6.4. Detection Results

Figure 9 compares the detection performance of both methods over the same image set. The detection statistic from a single vehicle model is used as a discriminant versus the rest of the test set. Figure 9 (a) (flexible histogram) and Figure 9 (d) (template method), for example, use the BMP-2 statistic as a discriminant against the confusion vehicles of figure Figure 8 as well as the remaining vehicles in the recognition class (i.e. T-72 and BTR-70). The results of figure Figure 9 do not include detection statistic from the training vehicle (except for the BTR70, for reasons stated above). As is readily observed, the flexible histogram approach compares favorably to the template based approach for all three vehicle types.

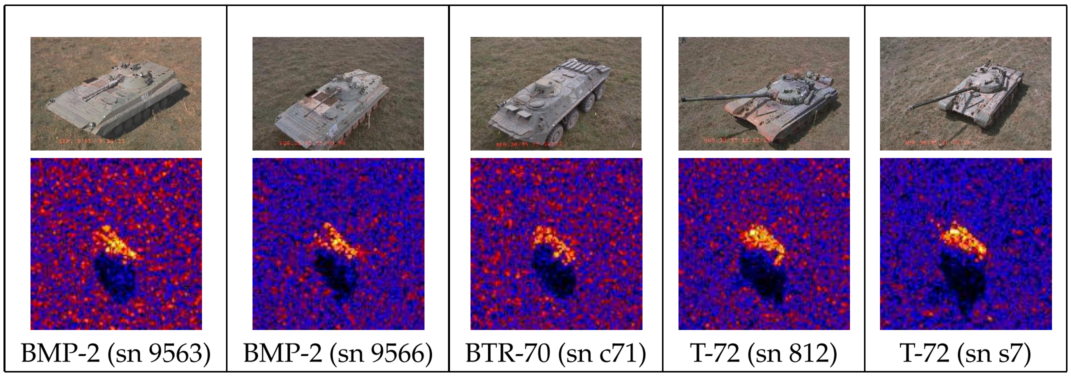

Figure 7. Testing Vehicles for Recognition Class.

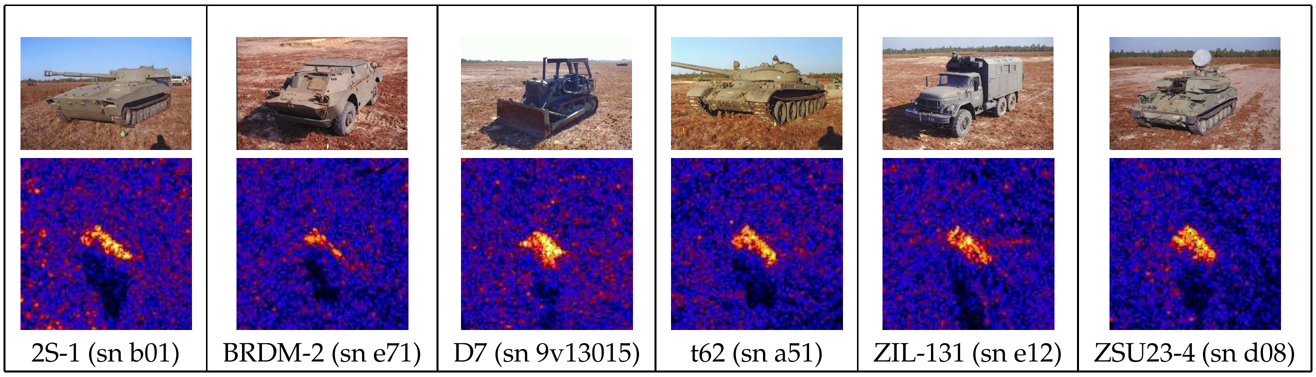

Figure 8. Testing Vehicles for Confusion Class.

6.5. Classification Results

As stated, in this set of experiments, we lack a true *detection* statistic and so for the purposes of our experiment we use the following for both methods:

1. Given an ROI, compute each model statistic.
2. Choose the most likely (max or min statistic) over all of the vehicle models as the class label.
3. Compare the designated statistic against a threshold.
4. If it passes the threshold, label the ROI with the winning class, otherwise designate the ROI as "not classified".

In the steps described above, the threshold is model dependent and is set such that 90% of test vehicles of that model type will pass. The classification results (as a percentage) of this experiment are shown in table Table 1. Numbers in parentheses are the corresponding classification rate for the template based approach. As the detection rate is the same for both methods it is not surprising that the classification performance on recognition vehicles is nominally the same. We do observe that the "not classified" rate is less than 10% for some cases, in particular the T-72. This is due to the confusion among the models (e.g. some of the rejected T-72s do pass the BMP-2 test).

It is in the performance on the confusion class that we note the largest difference. This is not surprising in light of the ROC curves from the detection results.

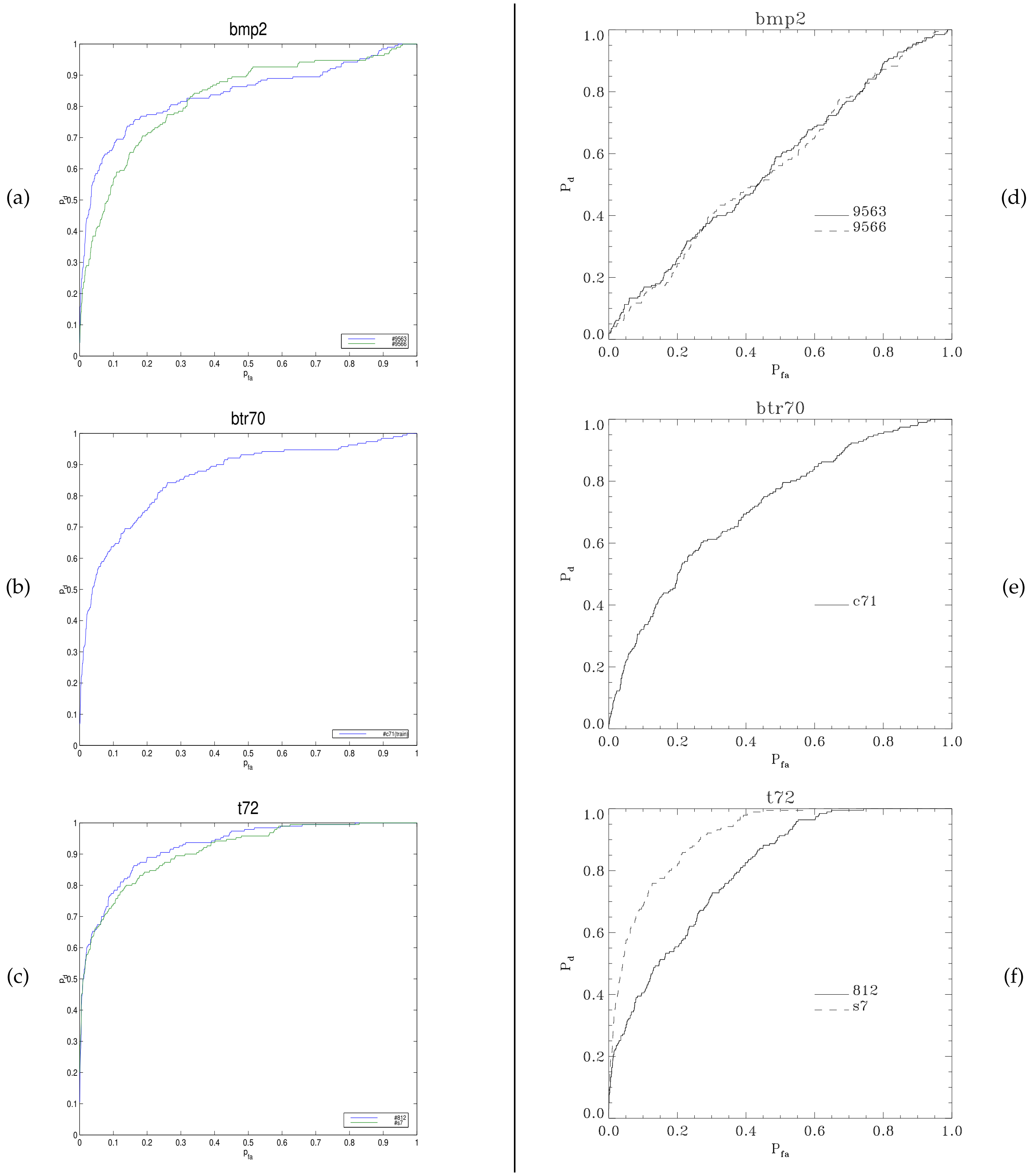

Figure 9. ROC curves for models generated by models for BMP2 (a,d), BTR70 (b,e) and T72 (c,f) vehicle types using the flexible histogram model (a-c) and the template matching method (d-f).

	BMP2-C21 @ 17^o		BTR-C71 @ 17^o		T72-132 @ 17^o		not classified	
BMP2-9563 @ 15^o	0.837	(0.903)	0.053	(0.005)	0.011	(0.005)	0.099	(0.087)
BMP2-9566 @ 15^o	0.858	(0.847)	0.063	(0.015)	0.005	(0.015)	0.074	(0.122)
BTR70-C71 @ 15^o	0.042	(0)	0.853	(0.898)	0.005	(0)	0.100	(0.102)
T72-812 @ 15^o	0.021	(0.041)	0.053	(0)	0.858	(0.795)	0.068	(0.164)
T72S7 @ 15^o	0.021	(0.047)	0.126	(0.010)	0.779	(0.921)	0.074	(0.021)
BRDM2 @ 15^o	0.174	(0.285)	0.174	(0.332)	0.132	(0.029)	0.520	(0.354)
D7 @ 15^o	0.542	(0.529)	0.137	(0.153)	0	(0.223)	0.321	(0.095)
T62 @ 15^o	0.626	(0.524)	0.121	(0.117)	0.068	(0.264)	0.185	(0.095)
ZIL131 @ 15^o	0.379	(0.274)	0.168	(0.310)	0	(0.314)	0.453	(0.102)
ZSU-23-4 @ 15^o	0.474	(0.394)	0.089	(0.095)	0.021	(0.453)	0.416	(0.058)
2S1 @ 15^o	0.616	(0.493)	0.179	(0.234)	0.079	(0.069)	0.126	(0.204)

Table 1. Confusion matrix for the flexible histogram approach (template-based results in parentheses) at 90% detection rate.

7. DISCUSSION

The flexible histogram multiresolution target discrimination approach described here makes explicit the requirement that to be considered similar, images must contain similar distributions of *joint* feature responses over multiple resolutions. SAR classification rates shown here, though preliminary, compare favorably to the Wright Patterson baseline classifier. Further analysis, including experiments on larger data sets, and on data which includes images with targets and clutter in close proximity, are required to further refine the system and obtain a better estimate of its overall potential.

Future research suggested by this work includes specializing the flexible histogram technique to the particular characteristics of SAR imagery. In the work presented here, we only considered images generated from the magnitude of the SAR signal. In future work we plan to examine using the full complex signal to increase model specificity. Additionally, the particular wavelet decomposition used here is generic, and in future work we will be considering different decompositions which may be able to better match the images generated from SAR.

REFERENCES

1. J. S. De Bonet, “Multiresolution sampling procedure for analysis and synthesis of texture images,” in *Computer Graphics*, ACM SIGGRAPH, 1997.
2. J. S. D. Bonet and P. Viola, “Texture recognition using a non-parametric multi-scale statistical model,” in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, 1998.
3. V. Velten, T. Ross, J. M. S. Worrell, and M. Bryant, “Standard SAR ATRevaluation experiments using the MSTAR public release data set,” in *Algrorithms for Synthetic Aperture Radar Imagery V*, E. G. Zelnio, ed., *Proc. SPIE* **3370**, pp. ?–?, 1998.
4. R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, 1973.
5. M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Nikoukhah, and A. S. Willsky, “Modeling and estimation of multiresolution stochastic processes,” *IEEE Transactions on Information Theory* **38**(2), pp. 766–784, 1992.
6. W. Irving, A. Willsky, and L. Novak, “A multiresolution approach to discriminating targets from clutter in SAR imagery,” *Proc. SPIE* **2487**, 1995.
7. N. S. Subotic, L. M. Collins, M. F. Reiley, and J. D. Gorman, “A multiresolution generalized likelihood ratio detection approach to target screening in synthetic aperture radar data,” *Proc. SPIE* **2487**, 1995.
8. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multiscale transforms,” *IEEE Transactions on Information Theory* **38**, pp. 587–607, March 1992.
9. S. Geman and D. Geman, “Stochastic relaxation, gibbs distributions, and the bayesian restoration of images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, pp. 721–741, 1984.

10. S. C. Zhu, Y. Wu, and D. Mumford, “Filters random fields and maximum entropy (FRAME): To a unified theory for texture modeling,” *To appear in Int'l Journal of Computer Vision* .
11. J. S. D. Bonet and P. Viola, “A non-parametric multi-scale statistical model for natural images,” in *Advances in Neural Information Processing*, vol. 10, 1997.
12. T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.