# MINACE filter classification algorithms for ATR using MSTAR data

Rohit Patnaik and David Casasent

Dept. of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

## ABSTRACT

A synthetic aperture radar (SAR) automatic target recognition (ATR) system based on the minimum noise and correlation energy (MINACE) distortion-invariant filter (DIF) is presented. A set of MINACE filters covering different aspect ranges is synthesized for each object using a training set of images of that object and a validation set of confuser and clutter images. No prior DIF work addressed confuser rejection. We also address use of fewer DIFs per object than prior work did. The selection of the MINACE filter parameter *c* for each filter is automated using training and validation sets. The system is evaluated using images from the Moving and Stationary Target Acquisition and Recognition (MSTAR) public database. The classification scores ( $P_c$ ) and the number of false alarm scores for confusers and clutter ( $P_{FA}$  and  $P_{CFA}$  respectively) are presented for the benchmark three-class MSTAR database with object variants and two confusers. The pose of the input test image is not assumed to be known, thus the problem addressed is more realistic than in prior work, since pose estimation of SAR objects has a large margin of error. Results for both confuser and clutter rejection are presented.

**Keywords:** Automatic target recognition (ATR), distortion-invariant filters, minimum noise and correlation energy (MINACE) filter, synthetic aperture radar (SAR) object classification

## **1. INTRODUCTION**

Our concern is automatic target recognition (ATR) using synthetic aperture radar (SAR) imagery. We present our initial SAR ATR module designed using images from the Moving and Stationary Target Acquisition and Recognition (MSTAR) public dataset<sup>1</sup>. The database consists of X-band SAR images with 1ft. x 1 ft. resolution. We use a subset of the MSTAR database for three-class object classification and we address confuser and clutter rejection. We consider recognition of several versions (variants) of three objects: BMP2 (an armored personnel carrier), BTR70 (an armored personnel carrier), and T72 (a tank), and rejection of two objects: D7 (a bulldozer), and ZIL131 (a truck). For each target, images were captured at two different depression angles (15° and 17°) over a full 0°-360° range of aspect view (pose). About 200-300 different aspect view images of each object are available. The BMP2 and T72 have variants present in the database. These are vehicles with different serial numbers. A benchmark experiment was proposed<sup>2</sup> to evaluate MSTAR classification algorithms in which the BMP2, BTR70, and T72 are used as object classes. All aspect views of the three targets at a 17° depression angle are used for training. The test set consists of all aspect views of the three targets (and their variants) to be recognized at a different 15° depression angle, and all aspect views of the two confusers (D7 and ZILI31) to be rejected at a different 15° depression angle. Note that we use a new set of confusers D7 and ZIL131 than did prior work. Most prior work used the 2S1 (a rocket launcher) and the D7 as the standard confusers to be rejected; Reference 3 also presented results using the standard confusers. The confusers we chose to reject are less like military vehicles than are the original ones, as noted earlier<sup>3</sup>. No images of the variants are present in the training set. Table 1 lists the number of images available for each of the targets and their variants and confusers at the two depression angles. A '-' entry indicates that the images of the variant are not used in the training stage. We use the BMP2-9563 and the T72-132 as the references for the two objects in the training stage as prior work<sup>3</sup> did. Thus, the  $17^{\circ}$ images of the BMP2-9566, BMP2-c21, and the T72-812 and T72-s7 are not used in the training stage.

The object classification system should recognize objects with a  $2^{\circ}$  difference in depression angle and also recognize variants (and handle a  $2^{\circ}$  difference in depression angle) without having trained on such images. We also consider rejection of confuser and clutter images. The MSTAR clutter database consists of 100 clutter scenes of size 1478x1784 pixels. These images are mainly trees, etc. In earlier work<sup>4</sup>, a screener was applied to each clutter scene to generate a total of 1616 64x64 pixel clutter regions of interest. To evaluate clutter rejection performance, we use these 1616 clutter chips as the clutter to be rejected. All test inputs are evaluated versus the filters for the three targets. The highest match-

score is used to evaluate the performance of the system. If the filter for the correct target produces the highest matchscore, it contributes to  $P_C$  if the match-score is  $\geq$  some threshold, otherwise it is rejected. If the highest match-score for a confuser or clutter image is  $\geq$  the threshold, it contributes to  $P_{FA}$  (*confuser false alarms*) or  $P_{CFA}$  (*clutter false alarms*) respectively. When variants are not considered,  $P_C$  is a percentage out of 587. When recognition of variants is required,  $P_C$  is a percentage out of 1365.  $P_{FA}$  is a percentage out of 548 and  $P_{CFA}$  is a percentage out of 1616. One can vary the match-score threshold to obtain receiver operating characteristic (ROC) curves for  $P_C$  vs.  $P_{FA}$  and then choose a suitable operating point. The equal error rate (EER) is an accepted comparison operating point on the ROC for which  $(1 - P_C = P_{FA})$ . We use this operating point to compare our results to prior work. For evaluating clutter rejection performance, one also considers  $P_D$ , the percentage of targets that are detected (regardless of whether classification is correct).

**Table 1.** Number of aspect views of each target available at a 17° and 15° depression angle, '-' indicates images of the variants are available but are not used in the training stage.

	BMP2			DTD 70		T72		D7	<b>7</b> II 121
	9563	9566	c21	DIK/U	132	812	s7		ZILIJI
17°	233	-	-	233	232	-	-	299	299
15°	195	196	196	196	196	195	191	274	274

Most prior DIF SAR work divided the full 360° of aspect angular range into 12-16 bins of smaller sizes. In all prior DIF work, the aspect angle of the input was assumed known. However, pose estimation of SAR objects is known to be poor with  $\pm 20^{\circ}$  range of pose estimation errors<sup>5</sup>. We wish to use fewer MINACE filters per object than did prior work and we do not assume that the aspect angle of the test input is known. If the pose of an input test image is assumed known, performance results have been shown to improve<sup>6</sup>. Reference 6 notes that if the input pose is assumed known, the EER improves by 5% (from 23%-18%), and the  $P_{FA}$  (at  $P_D = 90\%$ ) drops from 45% to 25%. Our goals are also to automate the selection of the MINACE parameter c. We use a validation set of only a few confuser and clutter images to select c. The extended maximum average correlation height (EMACH) filter has been used for MSTAR classification<sup>4</sup>. However, the selection of the EMACH filter parameter  $\beta$  seems to require lots of test set data. No prior DIF work has considered confuser rejection. We also note that no prior three-class MSTAR paper has considered both confuser and clutter rejection. In this paper, we present a SAR ATR system for the above database based on the minimum noise and correlation energy (MINACE) distortion-invariant filter (DIF)<sup>7</sup>. For each target, a set of MINACE filters is synthesized covering different angular ranges. To synthesize the MINACE filters for an object, a training set of images of that object and a validation set of confuser and clutter images is used to select the MINACE filter parameter c. The MINACE filter for each object is a combination of images of that object only; it does not contain any confuser or clutter images. Thus, no false-class training is done.

We now discuss the image preprocessing we use. The original target chips range from 128x128 pixels to 192x193 pixels in size. We use the central 64x64 pixels of each of these target chips; it contains the target, its shadow and background clutter. Not all images are centered. However, since DIF correlation-based approaches are shift-invariant, we do not recenter the images. All training and test images are 64x64 pixels. SAR images have a large dynamic range. We log-scale the images as  $log_{10}(I+1)$ , where *I* represents the intensity of a particular pixel, to reduce the dynamic range. Most prior work has used log-scaled images. The log-scaled images are energy-normalized to have unit energy before synthesizing the MINACE filters and before performing recognition correlations. All target and clutter chips are normalized this way. Energy normalization of images is needed because of the large range of energy differences present between different aspect view images of the same object. However, *this makes all targets more similar (thus recognition is more difficult) and it makes clutter rejection more difficult*. The rest of this paper is organized as follows. In Sect. 2, we discuss prior work using the MSTAR database. Section 3 describes the MINACE filter theory. Section 4 presents our MINACE filter-synthesis algorithm. In Sect. 5, we present results for three-class object classification and confuser and clutter rejection.

### 2. PRIOR MSTAR THREE-CLASS WORK

We now discuss prior MSTAR three-class work. Our main concern is methods that have been applied to three-class object classification (with variants) while considering rejection of confusers or clutter. We note that no prior MSTAR work has addressed both confuser and clutter rejection. Additionally, we do not consider comparisons to prior work that

assumed the pose of an input image was known, as such work is not fair. In prior MSTAR classification results, an operating point corresponding to a detection rate  $P_D = 90\%$  is usually used, and results are presented using the conditional classification rate,  $P_{CC}$ , the number of detected objects that are correctly classified. The  $P_C$  we use is  $P_{CC}P_D$ . When we discuss  $P_C$  scores for prior work, we thus reduce  $P_{CC}$  by a factor of  $P_D$ . A discussion of prior MSTAR work follows. All results are presented using images at a 17° depression angle for training and images at a 15° depression angle for testing.

A distance-based mean square error (MSE) template-matching method<sup>2</sup> was applied to the three-class database. For each object, 36 templates were synthesized, each template covered 10° in aspect range (aspect bin) and was formed by averaging over all training images in the aspect bin. They considered three-class object classification (including variants) and rejection of two confusers (two tanks not present in the MSTAR public database). There was no discussion of log-scaling or energy normalization of target chips. They achieved a poor EER of 30%, that is,  $P_C = 70\%$  at  $P_{FA} = 30\%$ . They did not assume that the pose of an input object was known. Thus, the results presented are fair.

Support vector machines (SVMs) have been used for MSTAR classification<sup>6,8,9</sup>. Reference 6 applied a new support vector representation and discrimination machine (SVRDM) to the three-class database and considered recognition of variants and rejection of the two standard confusers. All images were log-scaled but were not energy normalized. The pose of test inputs were not assumed known. They achieved an EER of 23% and at  $P_D = 90\%$ , they achieved  $P_C = 85.4\%$  and  $P_{FA} = 45.3\%$ . They also achieved  $P_{CFA} = 0\%$  at  $P_D = 90\%$  using a different set of clutter chips than the ones we use here. In Ref. 8, a polynomial SVM was applied to the three-class database and recognition of variants was considered. However, confuser rejection was not considered and a  $P_C = 93\%$  was reported. A Gaussian SVM<sup>9</sup> was also used on the three-class database and recognition of variants and rejection of the two standard confusers, the 2S1 and the D7 were considered. A pose estimation algorithm<sup>9</sup> was noted and claimed to produce an average error of less than 8°. However, for classification results, they assumed the pose of an input image was known. Thus, the results presented are artificially better. They obtained a  $P_{FA} = 34.3\%$  at  $P_C = 84.9\%$  ( $P_D = 90\%$ ).

The performance of several other classifiers were evaluated in Refs. 9 and 10. All classifiers were applied to the three-class database. Recognition of variants and rejection of the 2S1 and D7 confusers were considered and the aspect angle of a test input was assumed known. Performance results were presented at  $P_D = 90\%$ . In Ref. 9, a template matching method (using 12 average templates per object) gave a  $P_C = 80.4\%$  and a poor confuser rejection score of  $P_{FA} = 46.5\%$ . The perceptron method achieved  $P_C = 82.9\%$  but a very poor  $P_{FA} = 72.8\%$ . An optimal hyperplane classifier gave  $P_C = 84.6\%$  and  $P_{FA} = 61.5\%$ . All these  $P_{FA}$  scores are very poor. In Ref. 10, a template matching method (using 12 templates per object) gave a  $P_C = 80.4\%$  and a  $P_{FA} = 47\%$ . A multi-resolution principal component analysis (PCA) method gave  $P_C = 79.7\%$  and  $P_{FA} = 40\%$ . And, a quadratic mutual information classifier gave  $P_C = 84\%$  and  $P_{FA} = 46\%$ . Again, all the confuser rejection  $P_{FA}$  results are very poor. Among all the classifiers in Refs. 9 and 10, the Gaussian SVM obtained the best scores of  $P_C = 84.9\%$  and  $P_{FA} = 34.3\%$ . However, the pose of the input test images were assumed known.

A feature space trajectory (FST)<sup>3</sup> classifier using shift-invariant magnitude Fourier transform features was applied to the three-class MSTAR database. Target chips were log-scaled and energy normalized. The aspect angles of input test images were not assumed known. An EER score of 17%, i.e.,  $P_C = 83\%$  and  $P_{FA} = 17\%$ , was obtained (recognition of variants and rejection of the two standard confusers were considered). This is the best EER score in prior work when aspect angle was not assumed known. A continuous prototype version of the FST was also proposed<sup>11</sup>, which produces a smoother manifold than the FST did. A low pass filter was applied to obtain a smoother trajectory. They obtained scores of  $P_C = 87.7\%$  and  $P_{FA} = 20.8\%$  at  $P_D = 90\%$ . However, these results are suspect, since they assumed the pose of the test inputs were known.

We now address prior DIF work. No prior MSTAR DIF work considered confuser rejection. Thus, we can only note the classification  $P_C$  scores when no rejection threshold is used. A combination of the maximum average correlation height (MACH) filter and the distance classifier correlation filter (DCCF)<sup>12</sup> was applied to the three-class database and recognition of variants was considered. Sixteen aspect bins were used for each target, thus 48 MACH filters were used. A pose-estimation algorithm based on morphological processing of the input images was proposed but does not seem to have been employed. The pose of the input images were assumed known, although Ref. 12 implies it was not, since the performance results using the DCCF are almost identical in Refs. 12 and 13, and Ref. 13 clearly states that the test input poses were assumed known. An input image is correlated versus MACH filters covering the correct aspect bin for each of the targets and if the peak-to-sidelobe ratio (PSR) scores are above an acceptable threshold, DCCFs are applied to the input image to classify it. They obtained  $P_C = 97\%$  (with variants) and confuser rejection was not addressed. That work<sup>12</sup> presented clutter rejection scores using the original 100 MSTAR clutter scenes and scores of  $P_D = 96\%$  and  $P_{CFA} = 0.083/\text{km}^2$  for a total of 24 km<sup>2</sup> of data (this is a total of two clutter false alarms) were shown. We cannot directly compare our clutter rejection performance to this work since we use a different set of clutter chips. We note that their future work<sup>12</sup> shows poorer clutter rejection results.

The polynomial distance classifier correlation filter (PDCCF)<sup>13</sup> was applied to the three-class database but confuser or clutter rejection was not addressed. They obtained an excellent  $P_C = 99.1\%$  (with variants). However, test set performance was used to determine which aspect bin filters for which object classes had to be extended to the PDCCF case and to determine which filter powers (parameters) to use and how many PDCCF filters to use per aspect bin per class. Additionally, the pose of the input test images were assumed known. Thus, the results are suspect. In Ref. 4, the extended maximum average correlation height (EMACH) filter was applied to the ten-class MSTAR database. Sixteen EMACH filters were used per object (a total of 48 filters for the three-class case). Test images were correlated versus all filters. Thus, input pose was not assumed known. We computed the  $P_C$  score (with variants) for the three-class case to be at least 84.4%. That work<sup>4</sup> presented object detection-clutter rejection performance. However, clutter and variant data were used in the training stage, thus results are suspect. For ten-class data (using EMACH filters) without variants, they achieved  $P_{CFA} = 2.5\%$  at  $P_D = 95\%$  and  $P_{CFA} = 1.25\%$  at  $P_D = 90\%$ . The corresponding  $P_{CFA}$  scores when variants were included were 4% and 2.5% respectively.

From this review of prior work, considering only cases when the aspect view of the test input is not assumed known and cases where confuser rejection was considered, we can expect  $P_C$  scores of  $\approx 70\%$  and  $P_{FA}$  scores of  $\approx 30\%$ . Thus, high scores are not expected. Clutter rejection with  $P_{CFA} = 2.5\%$  can be expected (in detection). We want to use fewer filters per object than prior work did and wish to automate the selection of the MINACE filter parameter c.

## **3. MINACE FILTER THEORY**

This section describes the version of the original MINACE filter<sup>7</sup> that we use. Vectors (matrices) are denoted as lower (upper) case bold letters. All data are in the Fourier Transform (FT) domain. The 2-D FT of the filter is lexicographically ordered into a column vector **h**. The 2-D FT of each training set image included in the filter is lexicographically ordered into a column vector  $\mathbf{x}_i$  of the data matrix  $\mathbf{X}$ . The filter is required to give a specified correlation peak value for each training set image included in the filter; these values (usually one) are specified by the elements of a column vector **u**. These *peak constraints* are described by

$$\mathbf{X}^{\mathrm{H}}\mathbf{h} = \mathbf{u} = \begin{bmatrix} 1 & 1...1 \end{bmatrix}^{\mathrm{T}},\tag{1}$$

where  $()^{H}$  denotes the complex conjugate (Hermitian) transpose. To improve performance, the filter **h** is also required to minimize a combination of correlation plane energy due to training images and correlation plane energy due to distorted versions of the objects to be recognized. We use zero-mean white Gaussian noise to model the expected distortion power spectrum. We choose the energy function to be minimized as

$$E = \mathbf{h}^{\mathrm{H}} \mathbf{T} \mathbf{h} \,, \tag{2}$$

where  $\mathbf{T}$  is a diagonal matrix whose diagonal entries are the spectral envelope of the training images and noise at each frequency. That is,

$$\mathbf{T}(k,k) = \max[\mathbf{S}(k,k), c\mathbf{N}(k,k)], \tag{3}$$

$$\mathbf{S}(k,k) = \max[\mathbf{S}_1(k,k), \mathbf{S}_2(k,k), \dots, \mathbf{S}_{N_T}(k,k)], \qquad (4)$$

where  $S_i$  is a diagonal matrix whose diagonal entries are the elements of the lexicographically ordered 2-D power spectrum ( $|FT|^2$ ) of training image *i*,  $N_T$  is the total number of training images, N is the identity matrix, the maximum value in S is normalized to one, and c ( $0 \le c \le 1$ ) controls the variance of the noise. The Lagrange multiplier solution that minimizes the expression in Eq. (2) subject to the constraints in Eq. (1) is

$$\mathbf{h} = \mathbf{T}^{-1} \mathbf{X} (\mathbf{X}^{\mathrm{H}} \mathbf{T}^{-1} \mathbf{X}) \mathbf{u} \,. \tag{5}$$

The filter parameter c and the images to be included in the filter are selected using a set of training images and a set of validation images (Sect. 4.2). S is based on all the images in the training set, not just the ones included in the filter. Thus, S and T (for a fixed c) do not change as new training images are included in the filter. This is computationally

attractive as **S** for a filter needs to be computed only once while **T** has to be computed once every time *c* is changed. The MINACE filter can be shown to be a linear combination of training images that have been preprocessed by  $T^{-1/2}$ . A lower value of *c* makes the filter minimize correlation plane energy due to training images more and makes  $T^{-1/2}$  emphasize higher spatial frequencies. This makes the filter more discriminative to false objects but it makes recognition of distorted versions of an object more difficult. A higher value of *c* makes the filter emphasize correlation plane energy due to noise (that models distortions) more and makes  $T^{-1/2}$  emphasize lower spatial frequencies. This improves the distortion-tolerance performance of the filter but it also makes the rejection ability of the filter worse. Thus, *c* trades-off recognition versus rejection performance. Since **T** in Eq. (3) is based on the *max*() operator, as *c* is varied, **T** does not change for all spatial frequencies, e.g. as *c* is increased, the value of **T** at higher spatial frequencies does not change.

## 4. MINACE FILTER SYNTHESIS

## 4.1. Issues in MINACE SAR Recognition Filters

We now note several issues in the use of MINACE filters for SAR classification.

#### 1. Do not assume that the aspect angle of an input test image is known

Pose estimation of SAR objects is known to be poor with a  $\pm 20^{\circ}$  range of errors<sup>5</sup>. Thus, we do not assume that the pose an input test image is known. Most prior work assumed pose was known. If the aspect angle of a test image were known, classification performance improves considerably since the outputs of fewer filters need be considered for classification.

#### 2. The number of filters needed per object must be determined

Prior MSTAR DIF work used sixteen filters per object<sup>4,12</sup> (a total of 48 filters for the three-class database). There was no discussion of why sixteen filters per object were needed. We wish to automate the selection of the number of filters needed per object based on the performance of the filters on the training and validation sets (the validation set for a filter consists of a few confuser and clutter images that should be rejected by the filter). Currently, we start by using two filters per object; each filter covers 180° of aspect range. If the filter-synthesis algorithm does not produce an acceptable design for either filter, we resynthesize the filters for that object using three filters per object, and so on.

#### 3. Guidelines are needed for the number of training images N to be included in a filter

If most or all of the available training set images for a filter have to be included in a filter before it can recognize all training set images (that is, before it produces correlation peak values  $\geq$  some specified threshold for all of the training set), the filter is expected to have poor recognition performance on the test set. However, for the MSTAR database, we found that a maximum limit on the number of images included in the filter does not seem to be an issue (we present data in Sect. 4.2). Thus, we do not presently limit the maximum value of *N*.

#### 4. How *c* affects good clutter rejection performance

From initial tests, we found that *a higher value of c seems to be better for clutter rejection* (we show data supporting this in Sect. 4.2). This is different from other databases (e.g. IR). For other databases, a lower value of *c* is better for clutter rejection. This is expected, since use of a lower *c* value emphasizes higher spatial frequencies and this should make clutter and targets more different; clutter in SAR is specular and *this and our use of energy normalized imagery seems to be the reason for the different trends*. We note that clutter rejection performance improves when more filters per object are used. We note that only two prior MSTAR DIF papers<sup>4,12</sup> considered clutter rejection.

#### 5. How c affects good confuser rejection performance

From initial tests, we found that a lower value of c is better for confuser rejection (data is shown in Sect. 4.2). This is expected since a lower value of c emphasizes higher spatial frequencies and should make the filter more discriminative against confuser images and thus confuser rejection performance improves. Thus, for good clutter rejection performance, we need to use a higher value of c, while for good confuser rejection performance we need to use

a lower value of c. Selecting a value for c to meet both of these objectives is thus a challenging problem. If rejecting clutter images is not of concern, we can obtain better confuser rejection performance by using a lower value of c. We note that *no prior MSTAR work considered both clutter and confuser rejection*.

#### 6. An initial value for *c* is needed

We use a starting value of c = 0.003. We then increased and decreased it and analyzed training and validation set results.

#### 7. Specifying a minimum training set peak constraint

We add images to the filter until the filter produces correlation peak values  $\geq$  the minimum training set peak constraint for all training set images. We chose this value to be 0.9.

8. Selection of the maximum confuser and clutter validation peak values allowed and the number of confuser and clutter false alarms allowed during filter synthesis

From initial tests, we chose the thresholds to be 0.7. We do not allow any clutter false alarms during filter synthesis. Clutter typically does not have structure and clutter errors indicate that the MINACE filter is not using the structure information in recognizing true objects and discriminating them against other objects. We allow 10% confuser false alarms during training, that is, we allow two confuser validation peaks out of 20 to be above 0.7. In prior work, poor confuser rejection performance (usually around 30%) was obtained, and we do not expect excellent confuser rejection results. We initially assumed that confuser false alarms would be due to images in the same aspect range as that of the filter. However, from initial tests, we found that confuser false alarms also occur due to confuser images outside the aspect range of the filter. This further supports the fact that, if we assume that the pose of an input test image is known, the confuser rejection performance improves and misleading results occur.

### 4.2. Initial Data Trends

This section presents initial data trends and discusses our choice for the parameters of the MINACE algorithm. To obtain initial results, we chose to recognize the BMP2-9563 object and reject two (new) confusers (D7 bulldozer, and ZIL131 truck) and twenty clutter chips. We formed MINACE filters to recognize the 90°-270° aspect range (180° aspect range) synthesized with different values of *c*. The training set for each filter consists of all available images at a 17° depression angle in the aspect range of the filter. We first synthesize the filter with the image closest to an aspect view of 180° (back view of the vehicle). This filter was correlated versus the rest of the images in the training set and the training set image producing the lowest correlation peak value was noted. That image was then included in the filter, a new filter was synthesized and the process was repeated until all images had been included in the filter. The performance of each of the filters as more images were added to it was evaluated for  $P_C$  using all 108 test set images of the BMP2-9563 object at a 15° depression angle and in the aspect range covered by the filter.  $P_{FA}$  performance was measured using ten images (out of 274) of each of the confusers at a 15° depression angle; and  $P_{CFA}$  performance was measured using twenty clutter chips. We used approximately every 27<sup>th</sup> confuser image beginning with the first image. In the MINACE algorithm, the *validation set* for each filter consists of the same 20 confuser and 20 clutter images used here. No variant data was considered in these initial trend tests.

If the correlation peak for a BMP2 test image is  $\geq$  a threshold, it contributes to  $P_C$ , if the correlation peak for a confuser image is  $\geq$  the threshold, it contributes to  $P_{FA}$ , and if the correlation peak for a clutter chip is  $\geq$  the threshold, it contributes to  $P_{CFA}$ . To see trends, we plot these correlation peak values as the number of images N included in the filter is increased. Figures 1 and 2 show test results for the filters with two different values of c (0.0003 and 0.003). There are a total of 111 training set image of the BMP2. Thus, the horizontal axis in all figures is N and ranges from 1 to 111. In each figure, we also show (as a function of N) curves of the minimum correlation peak for the full training set (curve  $\Box$ ), the minimum true test peak (curve  $\circ$ ), and the correlation peak required to achieve  $P_C = 95\%$  on the test set (curve \*).

Figure 1a shows results for all 108 test set images at a 15° depression angle; the three curves noted are also shown. We see that test set peaks increase as more images (N) are added to the filter and then beyond  $N \approx 70$  the increase in peak height is negligible. We considered requiring the filter to have some minimum number of images included, but found this was not necessary. In Fig. 2a, we see the same trend but the curve tends to flatten out at a lower N level with

a larger value of *c*. The flattening out of the test peaks with *N* seems to simply indicate that the training and test set images are different due to the different depression angles. The minimum true test peak (curve  $\circ$ ) is noticeably below the threshold for  $P_C = 95\%$  (curve \*). From this, we see that there are several outliers in the test set and thus we consider the 95% \* curve in our analysis.



Figure 1. Correlation peak values for the test and validation set images for the  $90^{\circ}-270^{\circ}$  aspect range filter with a lower c = 0.0003.

We next consider performance versus clutter; the correlation peaks for each of the 20 clutter chips are shown in Figs. 1b and 2b. If we use the \* curve as our threshold and operate in the flat region of the curve, there is only one clutter false alarm in Fig. 1b and none with a larger c value in Fig. 2b. Thus, a higher value of c is better for clutter rejection, as we noted earlier.

We next consider confuser rejection. Figs. 1c and 2c, and 1e and 2e, show the correlation peaks for the five confusers in each class in the 180° angular range of the filter. The confuser peaks are much lower in Figs. 1c and 1e with a smaller value of *c*. Thus, a lower *c* is best to improve confuser rejection. Figs. 1d and 2d, and 1f and 2f, show the correlation peak values for the five confusers outside the aspect range of the filter. These peaks are also large and are often larger than those in the angular range. Thus, *we must check all confusers at all aspects versus each filter*. Thus, if the input object's pose is assumed known, better results are expected. Using the \* curve as our threshold, we find several confuser false alarms. This is expected, since prior work typically had  $P_{FA} > 30\%$ .

We find the \* threshold (for  $P_c = 95\%$ ) to be larger ( $\approx 0.8$ ) for a higher value of *c* (it is  $\approx 0.7$  for the lower value of *c*) in Fig. 1. This is expected, since a larger *c* emphasizes lower spatial frequencies and thus makes training and test data more similar. In MINACE filter synthesis, we use c = 0.003 for the initial value (as in Fig. 2) as clutter rejection is better. In synthesis, we check the correlation peak values for the validation set clutter and confusers and require their values to be < 0.7, which is below our expected true correlation peak threshold of 0.8 (the \* curve in Fig. 2). For different objects and for filters with a smaller aspect angle range, *different thresholds were not used but should improve performance*. We note that when recognition of variants is considered, we expect to have to reduce the threshold below 0.8, hopefully not too much below 0.7 (to limit  $P_{FA}$ ).



Figure 2. Correlation peak values for the test and validation set images for the  $90^{\circ}-270^{\circ}$  aspect range filter with a larger c = 0.003.

## 4.3. MINACE Synthesis Algorithm

Our current MINACE filter-synthesis algorithm follows.

1. Initialize *c* to 0.003. Set the correlation peak threshold for the training images *true\_th* to 0.9 and the correlation peak threshold for the confuser and clutter images in the validation set *false\_th* to 0.7.

- 2. Synthesize a filter using one of the images from the training set with this value of c. We choose the image closest to the aspect view in the center of the aspect range of the filter (e.g. for a filter covering 90°-180° in aspect range, we choose the image with aspect view closest to 135°) as the first image.
- 3. Correlate this filter with the remaining images in the training set. If any of the correlation peaks are below  $true\_th$ , then add the image with the lowest correlation peak to the filter and synthesize a new filter with the new and the first image. Continuing this process ensures that the designed filter will recognize all training set images with correlation peaks  $\geq$  some minimum (in this case, 0.9). For training set images included in filter synthesis, the filter (by definition) gives correlation peaks of 1.0 to satisfy the correlation peak constraints.
- 4. Repeat step (3) and continue including more images from the training set in the filter, until the filter gives correlation peaks  $\geq true\_th$  with all images in the training set. We note that this step terminates if all available training set images are included in the filter.
- 5. Correlate this filter with the confuser and clutter images in the validation set. Note the number of confuser and clutter validation peaks that are  $\geq 0.7$ . We presently allow at most two confuser false alarms and zero clutter false alarms from the validation set. If there are more than two confuser false alarms, we decrease *c* by 0.0005, otherwise if there are any clutter false alarms, we increase *c* by 0.0005. In both cases, we repeat steps (2)-(4) after changing the value of *c*. If no filter (or value of *c*) can be found with the required validation set performance, we terminate filter synthesis.

If a filter cannot be synthesized, we use more filters per object (we first try three filters, then four per object, etc.).

## 5. THREE-CLASS TEST RESULTS

We now present performance results for the MINACE filters for the BMP2, BTR70 and T72. Filters could not be synthesized for any of the three objects using only two or three filters per object. We first present performance results for the filters for the BMP2 and the T72 (each used four filters per object). For the BMP2 object, the test set consists of 195 images of the BMP-9563, 196 images of each of the two variants (9566 and c21), 274 images of each of the two confusers (D7 and ZIL131) all at a 15° depression angle, and 1616 clutter images. A test image is correlated with all filters for the object (aspect range is not assumed known). *The highest correlation peak value is used to compute performance scores*. If the correlation peak for BMP2-9563 test images is  $\geq$  a threshold, it contributes to  $P_{obj}$ . If the correlation peak for an image of the variants is  $\geq$  the threshold, it contributes to  $P_{var}$ .  $P_{obj}$  is a percentage out of 195,  $P_{var}$  is a percentage out of 392.  $P_C$  is the performance on the test set objects (including variants); it is a percentage out of 548 and  $P_{CFA}$  is a percentage out of 1616. Table 2 shows these scores as the correlation peak threshold Th is varied. We also show the individual  $P_{obj}$  and both  $P_{var}$  variant scores (percentage out of 196) and the individual  $P_{FA}$  scores for each confuser (percentage out of 274).

Table 2.	Performance	scores for	r one-class	(BMP2)	object	classification	with	variants	and	with	confuser	and	clutter	rejection	using
four filters	s of the object.														

Th	D	P <sub>var</sub>			D		D		
11	Г <sub>оbj</sub>	9566	c21	Total	r <sub>C</sub>	D7	ZIL131	total	I CFA
0.774	90.26	72.96	83.16	78.06	82.11	4.38	10.95	7.66	0.06
0.764	92.31	73.47	84.69	79.08	83.48	4.74	12.77	8.76	0.12
0.762	92.31	74.49	85.71	80.10	84.16	4.74	13.50	9.12	0.12
0.754	93.33	77.04	87.76	82.40	86.03	6.57	16.42	11.50	0.25
0.744	93.85	78.57	91.84	85.20	88.07	9.12	21.17	15.15	0.25
0.738	94.36	79.59	92.35	85.97	88.76	9.49	22.99	16.24	0.31
0.733	94.36	81.63	94.39	88.01	90.12	9.85	25.55	17.70	0.43

From Table 2, we see that the ZILI31 object produces many more false alarms than the D7 object. The ZIL131 truck looks more like a target than the D7 bulldozer. We note that the same trend occurs for filters for all objects. Also note that the recognition performance for the c21 variant is significantly better than the recognition performance for the 9566 variant. The clutter rejection performance is very good (< 0.5% for  $P_C = 90.12\%$ ). We can achieve  $P_C > 90\%$  (with

variants) at  $P_{FA} = 17.7\%$  and  $P_{CFA}$  (clutter) = 0.43%. These scores are all better than prior work, but only one object is considered. For the other BTR70 object filters, we obtained better results:  $P_C = 95.41\%$  (this object has no variants, this is probably why results are better) at  $P_{FA} = 4.74\%$  and  $P_{CFA} = 0.25\%$ . One can produce receiver operating characteristic (ROC) curves for the data and then select a threshold or an operating point to provide the performance desired, better  $P_C$ , better  $P_{FA}$ , etc.

We now evaluate the set of four MINACE filters for the T72 object. The test set consists of 196 images of the T72-132 and 195 and 191 images of the two variants, 812 and s7, respectively. The number of confuser and clutter images is the same as before.  $P_{obj}$  is a percentage out of 196,  $P_{var}$  is a percentage out of 386, and  $P_C$  is a percentage out of 582. Table 3 shows the performance scores when four filters per object are used as *Th* is varied. The T72 filters have very poor confuser rejection performance. For example, for  $P_{obj} = 95.39\%$ ,  $P_{FA} = 33.94\%$ , while for the BMP2, for  $P_{obj} = 94.36\%$ ,  $P_{FA} = 16.24\%$ . If the confuser rejection performance of the filters for one of the objects is poor, the confuser rejection performance for the three-class object case will also be poor. We also note from Table 3 that the recognition of the T72-812 variant is very difficult. For example, for  $P_{obj} = 94.39\%$ , the recognition rate for the T72-812 variant is only 62.56\%, that is 30% less than for the object. We thus used six filters per object for the T72 object (using the same MINACE synthesis rules as before). The  $P_{FA}$  score dropped from 33.94% to 20.07% at a  $P_{obj}$  of 94.39%. For  $P_C = 80.07\%$ , the  $P_{FA}$  score dropped from 33.58% to 26.46%. Thus, we use six filters per object for the T72 object in our three-class tests.

 Table 3. Performance scores for one-class (T72) object classification with variants and with confuser and clutter rejection using four filters of the object.

$T_{h}$	D	$P_{var}$			D		D		
11	P <sub>obj</sub>	812	s7	Total	$P_C$	D7	ZIL131	Total	P CFA
0.740	80.10	31.79	61.26	46.37	57.73	13.87	16.06	14.96	0.43
0.734	82.14	34.36	64.4	49.22	60.31	14.60	18.25	16.42	0.43
0.724	84.18	40.00	69.11	54.40	64.43	16.79	22.63	19.71	0.43
0.717	86.22	42.05	70.68	56.22	66.32	18.98	24.45	21.72	0.43
0.706	88.27	47.69	75.39	61.40	70.45	21.17	28.10	24.64	0.50
0.699	90.31	50.26	77.49	63.73	72.68	22.26	31.39	26.82	0.50
0.690	92.35	54.36	81.68	67.88	76.12	25.18	35.77	30.47	0.62
0.680	93.88	62.05	84.29	73.06	80.07	29.20	37.96	33.58	0.80
0.679	94.39	62.56	84.29	73.32	80.41	29.56	38.32	33.94	0.80
0.673	94.90	67.18	84.29	75.65	82.13	34.31	41.61	37.96	0.80
0.668	95.92	70.77	85.34	77.98	84.02	36.13	43.43	39.78	0.87

We now present the performance of the full system for three-class object classification (with variants) and with confuser and clutter rejection using four filters each for the BMP2 and BTR70, and six filters for the T72. An input test image is correlated versus all 14 filters and the highest correlation peak is used to determine performance scores.  $P_C$  is a percentage out of 587 when variants are not considered; it is a percentage out of 1365 when variants are included. We obtained an EER of 27%. We note that using six filters per object for all three objects, the  $P_C$  score at the EER point increases by 3%, that is  $P_C = 76\%$  for a  $P_{FA}$  of 24%. Since this is a significant increase, we compare our results to prior work when six MINACE filters per object are used (a total of 18 filters). Table 4 presents three-class object classification scores (with and without variants) using six filters each for the BMP2, BTR70, and T72.

We now compare our results to prior work. The only fair prior work apart from our earlier work<sup>3</sup> (EER = 17%) is the MSE template-matching method<sup>2</sup> which achieved a poor EER of 30%. By fair, we mean not assuming the pose of the inputs images is known, not including variants in the training set, and evaluating confuser rejection performance. Reference 6 notes that if the input pose is assumed known, the EER improves by 5% (from 23%-18%) and the  $P_{FA}$  (at  $P_D$  = 90%) drops from 45% to 25%.

Since we are considering DIFs, we consider prior DIF work. The EMACH filter work<sup>4</sup> used sixteen filters per object and achieved  $P_C \ge 84.4\%$  (with variants) for the three-class database (we extracted this score from ten-class data) but confuser rejection was not considered. And, clutter and variant data was used in the training stage. By comparison, we obtain better  $P_C = 90.26\%$  (with variants) than prior work when no rejection threshold is used. Since the PDCCF

work<sup>13</sup> used test data in the training stage and assumed the test input pose was known (Sect. 2), their  $P_C = 99.1\%$  score is not a valid score for comparison. We obtain an EER of 24% which is 6% better than the MSE template-matching method<sup>2</sup> and we use a total of 18 templates versus the 108 templates used in that work or the 48 filters used in other DIF work. The various other classifiers (other than the Gaussian SVM) evaluated in Refs. 9 and 10 achieved  $P_C$  scores ranging from 79.7%-84.6% and  $P_{FA}$  scores ranging from 40%-72.8%. We achieve similar  $P_C$  scores ranging from 80%-85.1% and better  $P_{FA}$  scores ranging from 35.4%-50.2% without assuming the aspect angle of the test images to be known, while they<sup>9,10</sup> did. Our clutter rejection performance is better than prior work<sup>4</sup> which achieved  $P_{CFA} = 4\%$  at  $P_D$ = 95% and  $P_{CFA} = 2.5\%$  at  $P_D = 90\%$ . Our related  $P_{CFA}$  scores are 1.7% and 1.2% respectively.

The scores we have presented here are from initial work. Our main objectives were to use few filters per object and automate the selection of the MINACE filter parameter *c*. Presently, we use the same filter-synthesis rules for each object. In future work, we will examine the use of different filter-synthesis rules for different objects and the use of different test thresholds peaks for different filters. These modifications and the use of a peak-to-correlation plane energy ratio (PCER) match-score instead of the correlation peak are expected to improve performance results in future work.

	P	С	P	D				
Th	w/o	w/	w/o	w/	D7	7H 131	Total	$P_{CFA}$
	variants	variants	variants	variants	D7	ZILIJI	Total	
0.775	90.80	75.02	93.19	79.34	13.50	28.83	21.17	0.87
0.771	91.14	75.60	93.53	80.00	14.60	31.39	22.99	0.93
0.769	91.31	76.04	93.70	80.73	15.33	32.48	23.91	0.93
0.763	91.99	77.07	94.38	82.05	17.15	35.04	26.09	0.99
0.758	92.16	78.02	94.55	83.15	18.25	38.32	28.28	0.99
0.755	92.67	78.39	95.06	83.66	20.80	39.05	29.93	0.99
0.753	92.84	79.05	95.23	84.32	22.63	40.88	31.75	0.99
0.749	93.02	79.63	95.57	85.13	26.64	43.43	35.04	1.05
0.747	93.36	80.00	95.91	85.49	26.64	44.16	35.40	1.05
0.745	93.53	80.59	96.08	86.23	27.74	45.62	36.68	1.05
0.742	93.70	81.03	96.42	86.74	30.29	47.81	39.05	1.05
0.740	94.04	81.68	96.93	87.69	32.12	48.54	40.33	1.11
0.738	94.04	82.05	97.10	88.35	32.85	49.64	41.24	1.11
0.734	94.72	83.00	97.79	89.38	35.40	51.82	43.61	1.18
0.727	95.57	84.03	98.81	90.99	37.59	55.11	46.35	1.18
0.723	95.91	84.54	99.15	91.72	40.15	56.57	48.36	1.24
0.718	96.25	85.05	99.66	92.60	41.97	58.39	50.18	1.30
0.707	96.42	86.01	99.83	94.14	47.45	62.41	54.93	1.42
0.702	96.42	86.74	99.83	95.02	50.36	64.23	57.30	1.67

**Table 4.** Performance scores for three-class (BMP2, BTR70, and T72) object classification and detection (with and without variants) and with confuser and clutter rejection using six MINACE filters per object.

## ACKNOWLEDGMENTS

The support of this work by Raytheon Missile Systems is gratefully acknowledged.

## REFERENCES

1. Moving and Stationary Target Acquisition and Recognition (MSTAR) Public Dataset, website: https://www.sdms.afrl.af.mil/datasets/mstar/

- T. D. Ross, S. W. Worrell, V. J. Velten, J. C. Mossing, and M. L. Bryant, "Standard SAR ATR evaluation experiments using the MSTAR public release data set," in *Algorithms for Synthetic Aperture Radar Imagery V*, E. G. Zelnio, ed., *Proc. SPIE* 3370, pp. 566-573, September 1998.
- 3. L. D. Ramamoorthy and D. P. Casasent, "Classification and rejection of MSTAR data," in *Optical Pattern Recognition XV*, D. P. Casasent, T. Chao, eds., *Proc. SPIE* **5437**, pp 265-276, April 2004.
- 4. R. Singh and B. V. Kumar, "Performance of the extended maximum average correlation height (EMACH) filter and the polynomial distance classifier correlation filter (PDCCF) for multiclass SAR detection and classification," in *Algorithms for Synthetic Aperture Radar Imagery IX*, E. G. Zelnio, ed., *Proc. SPIE* **4727**, pp. 265-276, August 2002.
- 5. Q. H. Pham, A. Ezekiel, M. T. Campbell, and M. J. Smith, "New end-to-end SAR ATR system," in *Algorithms for Synthetic Aperture Radar Imagery VI*, E. G. Zelnio, ed., *Proc. SPIE* **3721**, pp. 292-301, August 1999.
- 6. C. Yuan and D. P. Casasent "A new SVM for distorted SAR object classification," in *Optical Pattern Recognition XVI*, D. P. Casasent, T. Chao, eds., *Proc. SPIE* **5816**, paper 2, March 2005.
- 7. G. Ravichandran and D. P. Casasent, "Minimum noise and correlation energy optical correlation filter," *Applied Optics* **31**(11), pp. 1823-1833, April 1992.
- 8. M. L. Bryant and F. D. Garber, "SVM classifier applied to the MSTAR public data set," in *Algorithms for Synthetic Aperture Radar Imagery VI*, E. G. Zelnio, ed., *Proc. SPIE* **3721**, pp. 355-360, August 1999.
- 9. Q. Zhao and J. C. Principe, "Support vector machines for SAR automatic target recognition," *IEEE Trans. on* Aerospace and Electronic Systems 37(2), pp. 643-654, April 2001.
- 10. Q. Zhao, J. C. Principe, V. L. Brennan, D. Xu, and Z. Wang, "Synthetic aperture radar automatic target recognition with three strategies of learning and representation," *Optical Engineering* **39**(5), pp. 1230-1244, May 2000.
- 11. M. L. Bryant, "Target signature manifold methods applied to MSTAR dataset: preliminary results," in *Algorithms for Synthetic Aperture Radar Imagery VIII*, E. G. Zelnio, ed., *Proc. SPIE* **4382**, pp. 389-394, August 2001.
- 12. A. Mahalanobis, D. W. Carlson, and B. V. Kumar, "Evaluation of MACH and DCCF correlation filters for SAR ATR using the MSTAR public database," in *Algorithms for Synthetic Aperture Radar Imagery V*, E. G. Zelnio, ed., *Proc. SPIE* **3370**, pp. 460-468, September 1998.
- 13. M. Alkanhal and B. V. K. Vijaya Kumar, "Polynomial distance classifier correlation filter for pattern recognition," *Applied Optics* **42**(23), pp. 4688-4708, August 2003.