Use of Non-Ideal Training Data in SAR ATR for Targeting

S. J. Lycett, J. Denton and D. Blacknell
QinetiQ
R105
Malvern Technology Centre
St. Andrews Road
Great Malvern, Worcs, WR14 3PS
United Kingdom

ABSTRACT

This paper investigates a template-matching approach to automatic target recognition (ATR) for SAR for the case in which positive identification of a single known target type is required. ATR schemes are necessarily supported by databases of training imagery and of particular interest in this study was the impact on performance of using a database of simulated imagery to match against real in situ SAR imagery. Assessments have been made using 30cm resolution, X-band, spotlight imagery to provide the in situ test data and training databases firstly based on different examples from the same in situ data to provide an idealised baseline and secondly using simulated imagery. The impact on performance for both single channel and polarimetric data is reported and discussed.

1. INTRODUCTION

A particularly important automatic target recognition (ATR) task for synthetic aperture radar (SAR) targeting applications is that of single target identification which requires the positive identification of a single known target type within a relatively confined area. A database of previously generated training images will be available which may contain previously collected *in-situ* data, images of the target on a turntable in a controlled environment (ISAR) or images generated through the use of signature prediction tools. In this paper, a template-matching approach to the single target identification problem has been considered rather than a feature-based approach as discussed in Section 2. The robustness of template-matching to variations in aspect angle, measurement day and use of different vehicles of the same type has been studied using 30cm resolution, X-band, ISAR data of a military vehicle as discussed in Section 3. This natural within-class variation has been used to quantify confidence bounds in the matching process. The performance of the resulting ATR scheme has then been assessed in the ideal case in which independent training and test sets derived from the same 30cm resolution, X-band, spotlight imagery data source are used. However, in an operational system, the training database will be non-ideal in that it will have been obtained under different circumstances to the operational imagery against which the ATR is to be used. Performance assessments using a training database of simulated target signatures and a test set of spotlight SAR imagery have thus been performed to investigate the impact on performance of using such non-ideal training data. The results are presented in Section 4 and conclusions are drawn in Section 5.

Several data-sets have been used to support this study consisting of in situ images of targets in spotlight SAR imagery, turntable ISAR imagery of a target and simulated target images. These are described in more detail in the Appendix.

Paper presented at the RTO SET Symposium on "Target Identification and Recognition Using RF Systems", held in Oslo, Norway, 11-13 October 2004, and published in RTO-MP-SET-080.

2. TARGET CLASSIFICATION SCHEMES

Target classification is achieved using a database containing previously generated imagery of targets from a variety of different classes of interest. This is called the training set. The target image under test is compared to the images in the database and the closest match is established in some manner. If a sufficiently close match is found, then the target is declared to belong to the same class as the closest match. However, if the match is insufficiently close, then the target is declared as being unrecognised. The fundamental characteristic of different classification algorithms is how they establish the closeness of the test image to the target classes represented in the training set. There are many possible ways in which this can be achieved but they basically divide into two different types of approach, namely template-matching and feature-based classification.

In template-matching, the test image is compared with every image in the training database on a pixel-by-pixel basis to find the closest match. This comparison will normally be achieved by some form of correlation operation with the classification being declared on the basis of the highest correlation value found. One of the main drawbacks with template matching is that searching through each image in the training database can be extremely computationally expensive since the training database must contain representative images of every target in every possible configuration. This can require a very large number of images given that SAR images of the same vehicle at different aspect angles can look very different, together with the fact that a particular vehicle can assume many different configurations depending on the articulation of moving parts, such as the turret of a main battle tank, and the attachment of objects such as oil drums. Nevertheless, the template-matching concept provides a basic, robust approach to classification.

The alternative feature-based approach to classification alleviates the problem of searching through every image in the training database by characterising the targets of a particular class in terms of particular properties termed ‘features’. Features can be based on obvious physical characteristics of the target such as length and width or may be more abstract such as statistical measures of the variation in pixel brightness across the target. The idea is that a single set of features can be used to characterise many of the target images from a particular class. This means that, rather than comparing a particular test image with a large number of training images, the feature values measured for the test image can be compared with the single set of feature values which characterise a whole set of training examples. Feature-based approaches are thus potentially much more efficient than temple-matching approaches. However, the problem comes in defining an appropriate set of features and defining the method of comparison with features measured for the training examples.

For the targeting application considered in this study, the classification problem is to detect and recognise one particular target type within a SAR scene. For example, the mission may be to find a missile launcher based on collateral information that one or more are likely to be present in a particular area. In this case, the training database will require examples of only one target type and so the use of a template-matching approach may be appropriate. For this reason, the classification scheme developed in this study has taken the template-matching approach.

3. WITHIN-CLASS SIMILARITY

3.1 Introduction

SAR imagery can be presented in a number of forms. The initial SAR image produced from the raw data consists of complex-valued pixels which constitute the complex SAR image. The complex image can then be used to form the amplitude image, by taking the modulus of the complex values, or the intensity image, by taking the square of the modulus of the complex values. Either of these can be displayed as greyscale images to the operator although, because of the large dynamic range of SAR imagery, it is found that the

amplitude image is the more suitable of the two for display. For template-matching, it is necessary to establish which form of SAR image it is most appropriate to use. The determining factors on which this decision should be made are the degree to which targets from the same class are assessed as being similar, i.e. the similarity within class, and the degree to which targets from different classes are assessed as being different, i.e. the classification performance. An additional consideration is that a different form of SAR image may be even more appropriate for classification based on template matching. For example, if an appropriate threshold is applied to the target images to give a binary output such that pixels on the target are assigned a value 1 and pixels in the background are assigned a value 0 then a true template of the target results. Whether this form of image provides better template-matching performance needs also to be considered.

3.2 Correlation measures

The standard measure of correlation is given by the correlation coefficient

$$\rho = \frac{\frac{1}{N}\sum_i x_i y_i - \left(\frac{1}{N}\sum_i x_i\right)\left(\frac{1}{N}\sum_i y_i\right)}{\sqrt{\left(\frac{1}{N}\sum_i x_i^2\right) - \left(\frac{1}{N}\sum_i x_i\right)^2}\sqrt{\left(\frac{1}{N}\sum_i y_i^2\right) - \left(\frac{1}{N}\sum_i y_i\right)^2}} \tag{3.1}$$

where x and y are pixel values in the images to be correlated, N is the number of pixels and i enumerates the pixels in the image. The correlation coefficient can take values between –1 and 1 where a value of 1 indicates perfect correlation, a value of 0 indicates that the variables (i.e. images in this case) are uncorrelated and a value of –1 indicates perfect anti-correlation (e.g. one image will be the negative of the other image). This measure is the best estimate of the correlation assuming that both variables are subject to additive, Gaussian noise. For SAR images, the noise process is multiplicative speckle [1] and so the correlation coefficient given above will not provide the most accurate estimate. However, the calculation of this measure can be implemented extremely efficiently using Fourier transforms and this advantage far outweighs the disadvantage of some loss of accuracy. In this section, the variation of the correlation coefficient over example images for a particular target class is investigated. The objective is to establish that the correlation coefficient can be used to determine that the images are of the same target.

3.3 Self-correlation

The ISAR turntable imagery used for this investigation provides images of the target at every degree over 360° of aspect angle. Each of these images can be correlated with the same full set of 360 images to produce 360×360 values of the correlation coefficient. In the following results, the HH polarimetric channel was used.

The self-correlation properties using SAR images in the complex, intensity, amplitude and binary forms have been investigated. Figure 3.1 shows the averaged correlation coefficient variation is shown as a function of angular difference. In these graphs, the fitting procedure has smoothed the spike at 0° angular difference and the fact that all the graphs rise to a value of 1.0 at this point is not clear. However, the important point is that the value of the correlation coefficient drops off at different rates for different image forms. This drop-off rate is most for the complex image and least for the amplitude image. In terms of classification, it is desirable that the procedure should be robust to small changes in aspect angle otherwise the training database would need to be populated using images at a very fine sampling of the aspect angle. From this point of view, it is thus clear that the amplitude image is to be preferred. The “de-correlation” length for the amplitude image, measured in this case by a drop in the correlation coefficient to 0.8, is about 5° for the amplitude image which is an important consideration when forming the training database.

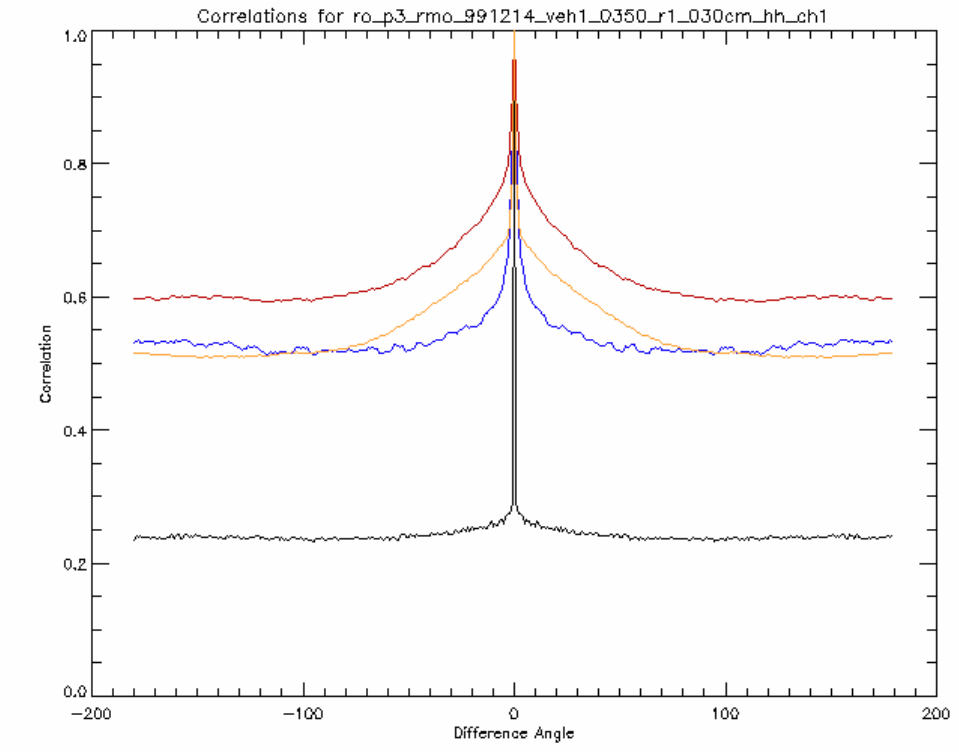


Figure 3.1: Plots of average correlation coefficient as a function of angular difference for the complex image (black), the intensity image (blue), the amplitude image (red) and the binary image (yellow).

3.4 Cross-correlation between different days

It is now necessary to investigate the robustness of the correlation coefficient to variations in the imaging circumstances. Firstly, if the same target is imaged on two different days, a robust classification scheme should still recognise it as the same target.

It is to be expected that, due to noise variations in imaging between the two days, the maximum correlation for an image at a particular aspect angle on one day will not necessarily be obtained using the image at exactly the same aspect angle on the other day. However, it should be the case that the angular difference between the best matching images will be small. Given the de-correlation length of 5° observed earlier, it is thus reasonable to search for the maximum correlation value within ± 5° of the matching angle.

Figure 3.2 shows graphs of the maximum correlation values found within this interval as a function of aspect angle. In other words, for images at each aspect angle on the first day, the maximum correlation with images on the second day within an angular difference of ± 5° was found and plotted on the graph. It can be see that the amplitude image produces the highest maximum values whilst the complex image produces the lowest maximum values.

The mean and standard deviation of these maximum values as they vary over aspect angle have been calculated and are presented in Table 3.1. This quantifies the observation that correlation applied to the amplitude image provides the greatest robustness to image variations between different days.

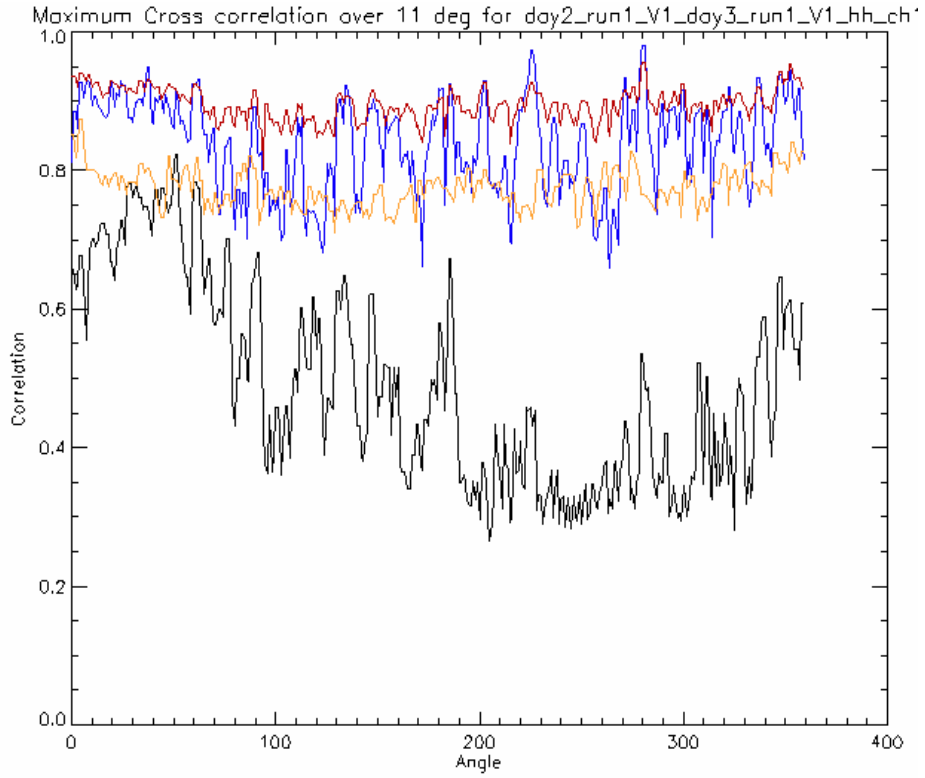


Figure 3.2: Plots of the maximum correlation within an angular difference of ± 5° from the corresponding aspect angle for target images obtained on different days for the complex image (black), the intensity image (blue), the amplitude image (red) and the binary image (yellow).

Image form	Mean	Standard deviation
Complex	0.49	0.15
Intensity	0.84	0.07
Amplitude	0.90	0.02
Binary	0.78	0.03

Table 3.1: Mean and standard deviation of maximum correlation values over aspect angle for plots shown in Figure 3.2.

3.5 Cross-correlation between different examples of same vehicle

A similar exercise can now be undertaken to investigate robustness to image variations which result when the correlation measure is applied to images of two different vehicles of the same type, i.e. it is only the registration numbers that are different.

Figure 3.3 shows graphs of the maximum correlation values found within the angular difference interval of ± 5°. Once again it can be seen that amplitude image correlation provides the highest maximum values whilst complex image correlation provides the lowest maximum values.

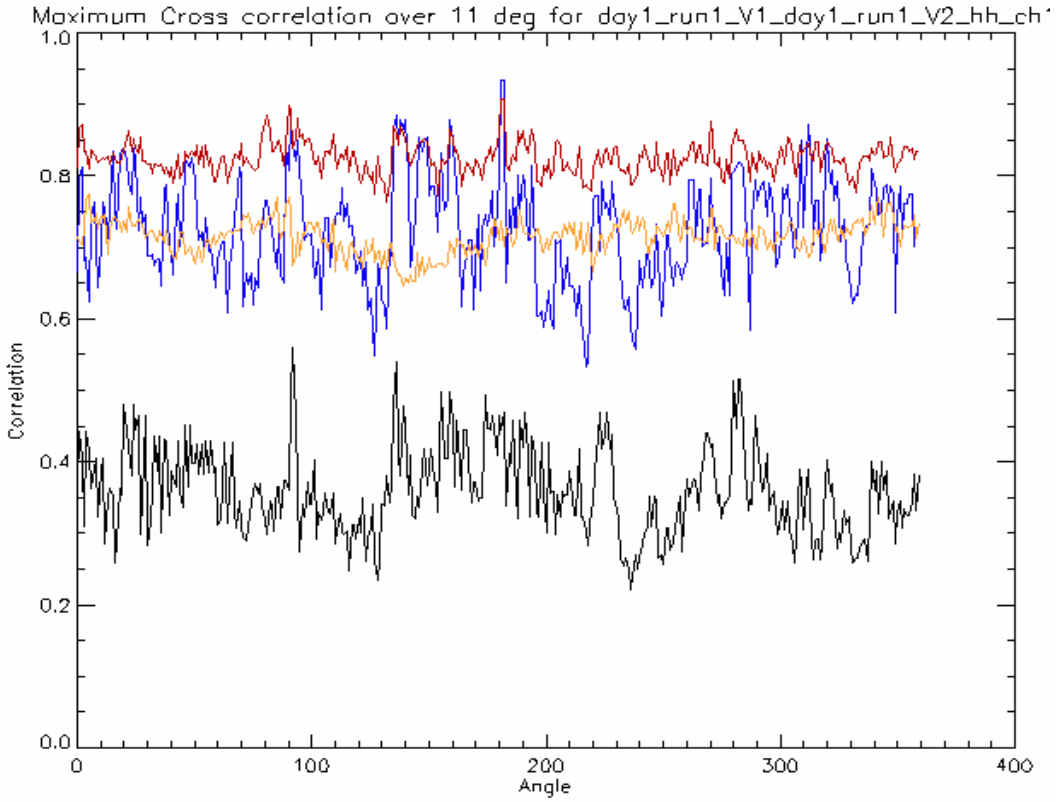


Figure 3.3: Plots of the maximum correlation within an angular difference of ± 5° from the corresponding aspect angle for target images obtained using a different vehicle of the same type for the complex image (black), the intensity image (blue), the amplitude image (red) and the binary image (yellow).

These observations are quantified in Table 3.2 using the mean and standard deviation of the maximum correlation value as before. It is clear that amplitude image correlation provides the greatest robustness to variations between different examples of the same target type. It should be noted that the mean maximum correlation coefficient in this case is significantly lower than the value obtained for variation between different days. Essentially the experiment for variation between different days involves taking a single example of the vehicle type, driving it onto the turntable, imaging it, driving it off, driving it on again and imaging it a second time. However, the experiment for different examples of the same vehicle type involves exactly this process together with the replacement of the vehicle with another of the same type between imaging runs. Thus it is to be expected that the loss in correlation observed for different days will be incurred together with an additional loss associated with use of different examples of the vehicle. These results are thus consistent with expectation.

Image form	Mean	Standard deviation
Complex	0.36	0.06
Intensity	0.72	0.07
Amplitude	0.83	0.02
Binary	0.71	0.02

Table 3.2: Mean and standard deviation of maximum correlation values over aspect angle for plots shown in Figure 3.3.

4. PERFORMANCE RESULTS

4.1 Introduction

The aim of this section is to assess the performance of the template-matching approach to target recognition. On the basis of the discussions in section 3, the amplitude image has been used in the experiments. Furthermore, since the trials data which has been used to support these experiments is fully polarimetric, the amplitude of the polarimetric span image has been used.

Whilst it has been argued earlier that template-matching is not appropriate for many-class problems due to the computational effort required, it is valid to explore the performance experimentally when computation time is not a critical issue. In particular, it is important to introduce the concept of an unrecognised class and also to investigate the impact on performance of the presence of targets not represented in the training database (termed "confusers"). These issues will be addressed in this section under the discussion of classification results.

The main problem to be addressed by this study was identified as that of single target identification which is of particular relevance for targeting applications. This can be viewed as an extreme example of the classification problem with confusers when everything other than the single target of interest can be considered to be a confuser. The results for the single target identification performance assessment are also presented in this section and follow naturally from the discussion of more general classification.

4.2 Classification results

For a classifier to be effective, it must not only be able to classify targets of interest, but it must also be able to declare targets as being unrecognised if there is insufficient confidence that the target belongs to one of the known classes. The correlation coefficient provides an effective means of assessing confidence since it provides a normalised measure lying between 0 and 1 which quantifies the goodness of the match between a test image and a training image. Thus it is simply necessary to define a threshold on the correlation coefficient values such that positive classification declarations are only made when the coefficient exceeds this threshold and otherwise the target is declared to be unrecognised.

Previous discussions regarding similarity within class in section 3 have shown that the maximum correlation coefficient generally exceeds 0.8 for targets from the same class which have been imaged under different circumstances. This value thus provides a suitable threshold value to be considered. In the following results, a threshold values of 0.8 will be used together with 0.0 to illustrate the performance when there is no unrecognised class.

	A	B	C	D	E	F	G	H	I	U
A	79	1	0	5	1	1	14	0	0	0
B	0	96	0	2	2	0	0	0	0	0
C	4	8	74	3	5	2	5	0	1	0
D	8	1	0	85	0	4	3	0	0	0
E	5	1	0	5	86	0	3	0	1	0
F	6	10	0	5	8	65	5	0	0	0
G	0	0	0	2	3	1	95	0	0	0
H	3	16	0	7	2	0	5	59	8	0
I	1	13	0	1	5	0	3	0	78	0

Table 4.1: Confusion matrix for 9 class problem with correlation threshold defining unrecognised class set to 0.0.

	A	B	C	D	E	F	G	H	I	U
A	68	1	0	0	1	1	12	0	0	17
B	0	94	0	1	2	0	0	0	0	4
C	0	0	0	0	0	0	0	0	0	100
D	8	0	0	77	0	4	2	0	0	9
E	4	0	0	0	58	0	1	0	0	38
F	0	0	0	4	2	14	0	0	0	81
G	0	0	0	2	3	1	93	0	0	2
H	0	0	0	0	0	0	0	5	0	95
I	0	0	0	0	0	0	0	0	0	100

Table 4.2: Confusion matrix for 9 class problem with correlation threshold defining unrecognised class set to 0.8.

Tables 4.1 and 4.2 show confusion matrices including an unrecognised class using the correlation coefficient thresholds of 0.0 and 0.8 respectively. The entries show the percentage of test examples classified as a particular target class. When there is no unrecognised class, the average correct classification rate is close to 80%. However, it is interesting to note that., when the unrecognised class is introduced, a number of images fail to produce correlation values exceeding 0.8 despite being represented in the training data. Indeed, almost all test examples of target types C, H and I are declared as being unrecognised as well as significant numbers of target types E and F. It should be noted that target types C, H and I are the decoy targets in the scene whilst E is a tractor with trailer and F is a water tower, i.e. none are true military vehicles. In fact, the only vehicles for which the classifier is confident , i.e. A, B, D and G, are precisely the military vehicles in the scene. This appears to suggest that the characteristics of military vehicles appear to make them more easily distinguishable than non-military objects.

	A	B	C	D	E	F	G	H	I	U
A	70	1	0	0	0	0	12	0	0	17
B	0	95	0	1	0	0	0	0	0	4
C	0	0	0	0	0	0	0	0	0	100
D	9	1	0	79	0	0	2	0	0	9
E	4	1	0	0	0	0	3	0	0	93
F	4	0	0	7	0	0	4	0	0	85
G	1	1	0	2	0	0	95	0	0	2
H	0	0	0	0	0	0	0	0	0	100
I	0	0	0	0	0	0	0	0	0	100

Table 4.3: Confusion matrix for 4 class problem with 5 confusers and correlation threshold defining unrecognised class set to 0.8.

In an operational ATR system, it is likely that the training database will contain comprehensive examples of military vehicles but that other objects will not necessarily be represented. Thus it is important to understand the performance of a classifier when the training set is restricted to military vehicles but confusers are present in the test set. Table 4.3 shows the classification results when only target types A, B, D and G are represented in the training database but an unrecognised class is included with a threshold of 0.8. It can be seen that the classifier rejects all the decoys (C, H and I) as being unrecognised as well as most of the tractor with trailer images (E) and water tower images (F). Confusers which are classified as military targets are false alarms and these represent 4.4% of the total in this experiment. For the military vehicles, 8% are declared as being unrecognised so that 92% can be considered to have been detected as being military vehicles. These concepts of detection and false alarm rates within a classifier will be seen to be of great relevance in the single target identification case. Finally, the correct classification rate for the four military vehicles is 85% in this case.

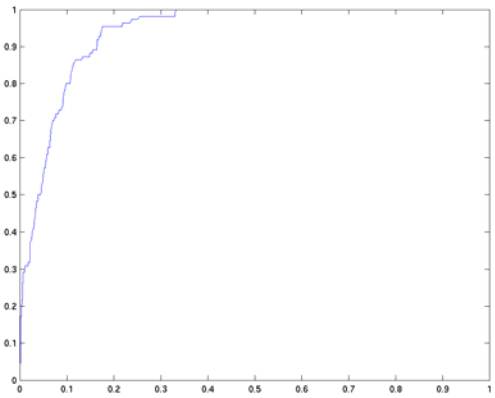
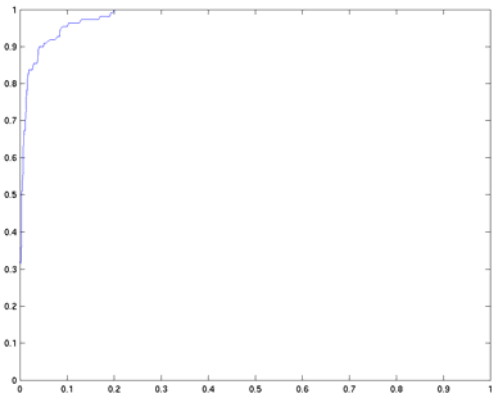
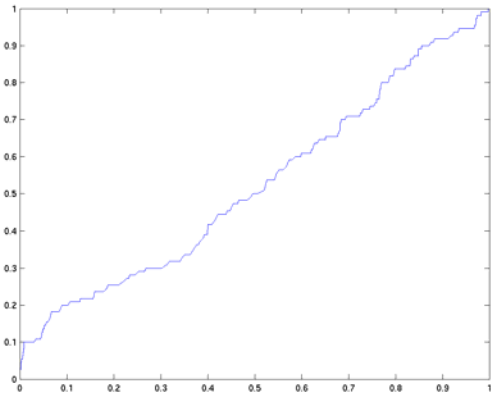
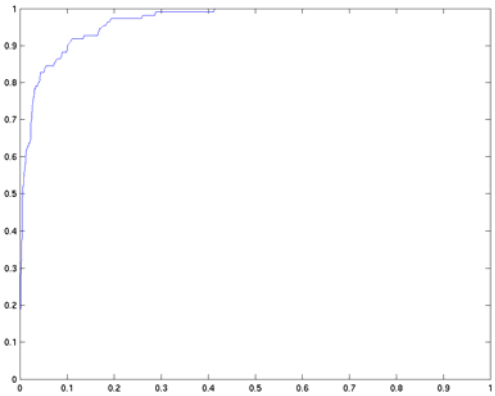
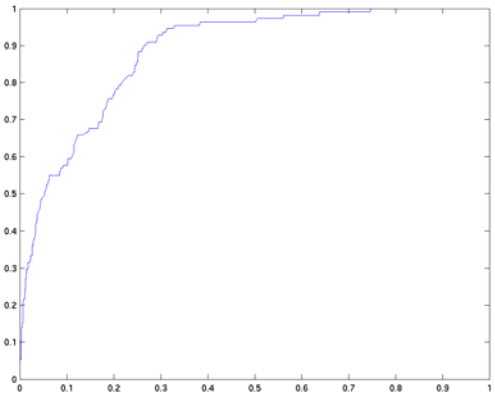
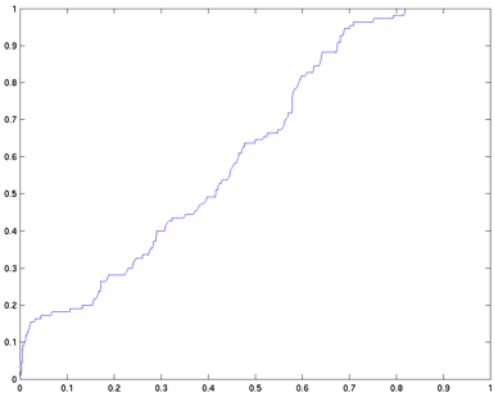
4.3 Single target identification results

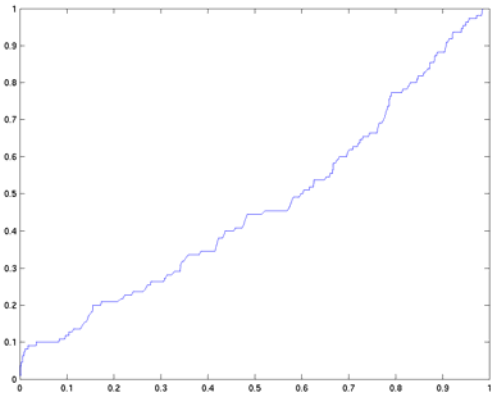
Single target identification is essentially a classification problem in which the training database contains only one class of military target. A suitable confidence threshold must be set, i.e. a threshold on the correlation coefficient value in this case, and then a test target image is declared as either being the target of interest or unrecognised. The concepts of detection and false alarm rates are then entirely appropriate to this classification problem. Detections occur when an example of the target class is correctly recognised whilst false alarms occur when a confuser is incorrectly declared to be a member of the target class. It is thus possible to plot the probability of detection against the probability of false alarm as the correlation threshold is varied to produce the so-called receiver operating characteristic (ROC) curves most usually associated with target detection assessments.

ROC curves have been produced for each of the nine targets in the *in situ* trial set. In each case, the training database contains only images of the particular class of interest. All the images in the test set are classified as either belonging to this class or as being unrecognised. The correlation threshold is varied from 0 to 1 and the resulting variation of probability of detection against probability of false alarm is plotted. Figure 4.1 shows the resulting curves for Targets A to I excluding Target G whilst the first graph in Figure 4.2 shows the ROC curve for Target G. It is interesting to first consider the decoy targets C, H and I. The curves for these essentially follow the line of equal probabilities for detection and false alarm. This indicates that no useful performance is being achieved since, for any given threshold, the same proportions of detections and false alarms are being obtained. In other words, the classification is essentially random. At the other end of the scale, the military vehicles (A, B, D and G) are showing a level of classification performance in that relatively high probabilities of detection can be achieved for relatively low probabilities of false alarm. The ROC curves for the tractor with trailer (E) water tower (F) lie between these extremes. These results are consistent with the observation made in section 4.2 that the military vehicles appear to be more easily distinguishable than the non-military objects.

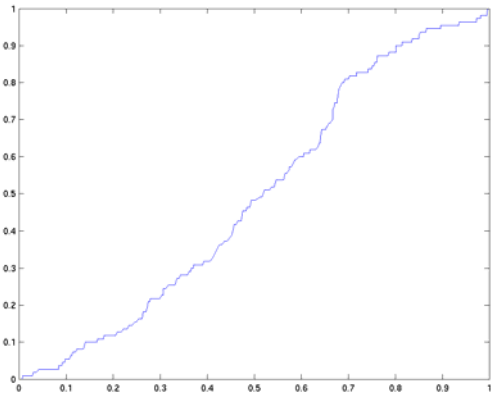
	Probability of detection	Probability of false alarm
Target A	0.80	0.11
Target B	0.95	0.09
Target D	0.89	0.10
Target G	0.98	0.14

Table 4.5: Summary of probabilities of detection and false alarm for military vehicles using a correlation threshold of 0.8.

	
Target A	Target B
	
Target C	Target D
	
Target E	Target F

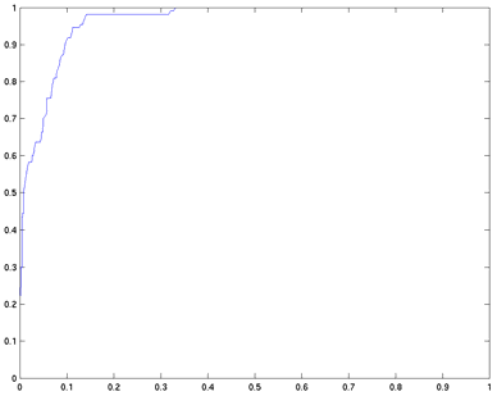


Target H

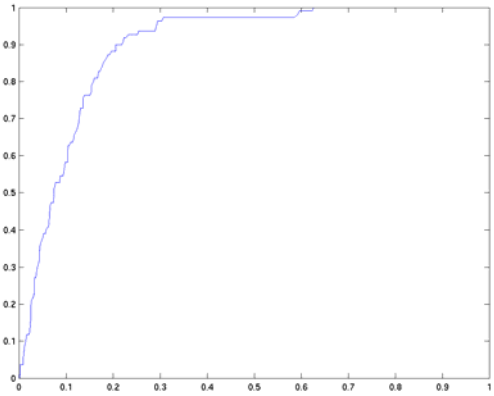


Target I

Figure 4.1: ROC curves showing probability of detection versus false alarm as correlation threshold is varied for Targets A to I excluding Target G.



Target G



Target G using simulated training images

Figure 4.2: ROC curves showing probability of detection versus false alarm as correlation threshold for Target G using training database from in situ trial data(left) and from simulated signature predictions (right).

The performance against the military vehicles will now be examined more closely. Operationally, a single correlation threshold needs to be set and the results of section 3 suggest that a value of 0.8 is a suitable choice since within-class variations still produce values exceeding this. Using this threshold, the probabilities of detection and false alarm shown in Table 4.5 are obtained.

It can be seen that the probabilities of false alarm are all around 0.1 (although that for Target G is a little higher than the rest) whilst the probabilities of detection vary from 0.8 to 0.98 reflecting the varying degrees to which each target is distinguishable from the others. From an operational perspective, this means that to achieve a detection probability of between 80% to 100% a false declaration will be made 10% of the time. It should be noted that, from the results of section 4.2, it can be concluded that most of the false declarations will arise from other military vehicles since the non-military objects were easily classified as being unrecognised.

An important operational factor for classification algorithms is the question of how the training database of imagery is populated. In these experiments, the test and training images were all obtained form exactly the same experimental trial. However, in practice, the training imagery may have to be obtained using turntable ISAR imaging or through the use of signature predictions using CAD (computer aided design) models of the vehicles of interest [2]. The impact of this disparity between the training data source and the operational data needs to be investigated.

A limited assessment of this issue has been made under the current study. For Target G, a training set consisting of signature predictions was generated at the appropriate resolution and geometry. The single target identification performance was then assessed as before using all the *in situ* trial test images and resulted in the ROC curve shown in the right of Figure 4.2. It is very encouraging that a significant level of classification performance has been achieved in this case. However, it is clear that the performance is also significantly worse than when test and training data are from the same data source as in the ROC curve on the left hand side of Figure 4.2. Using a correlation threshold of 0.8, the probability of false alarm in this case is 0.09 whilst the probability of detection is 0.55. Thus, in comparison to the results above for the same data sources, the false detections have been maintained at around 10% whilst the correct declarations have fallen substantially to 55%.

The performance results presented in this section have demonstrated that single target identification can be achieved using a template-matching approach but that the achievable performances are limited by the practical consideration that the training database will be populated using a different data source than that which will be used operationally. The conclusion is that improved performance needs to be achieved using algorithms which are, in particular, robust to the use of data from different sources, e.g. ISAR turntable imagery or signature predictions.

5. CONCLUSIONS

The results of various performance assessments have been reported using a set of four military vehicles and five non-military objects acting as “confusers”. It has been found that detection rates for a specific military vehicle within this target set ranges from 80% to 98% depending on the vehicle type with a false classification rate of about 10%. These results were obtained using training and test imagery from the same experimental data source. However, in practice the training database will be populated beforehand using, for example, ISAR turntable measurements or signature predictions. A limited experiment using training data for one vehicle generated from a signature prediction model revealed a drop in performance to 55% for a false alarm rate of about 10%.

The performance results which have been obtained have demonstrated that single target identification can be achieved using a template-matching approach but that the achievable performances are limited by the practical consideration that the training database will be populated using a different data source than that which will be used operationally. The conclusion is that improved performance needs to be achieved using algorithms which are, in particular, robust to the use of data from different sources, e.g. ISAR turntable imagery or signature predictions. A possible avenue of future research is thus the use of a classification approach which exploits understanding of the physics of the radar interaction with the target [2].

ACKNOWLEDGEMENTS

This work was supported by the UK MOD under the Applied Research Programme contract number A/CTA/N03504 entitled “All Weather Target Engagement”.

REFERENCES

[1] [1] C J Oliver and S Quegan, “Understanding Synthetic Aperture Radar Images”, Artech Housee, 1999.

[2] [2] D B Andre, “Generation of Simulated radar ground-target database, its validation & ATR performance”, RTO Symposium on Target Identification and Recognition using RF sensors, Oslo, 11-13 October, 2004.

APPENDIX

The data sets used in this study all consisted of fully polarimetric, 30cm resolution X-band imagery. The first data set was obtained at a Regular Army Assistance Trial (RAAT) on Salisbury plain in which spotlight imagery of targets in open ground were obtained every degree over 360° at an incidence angle of 5°. An example image is shown in Figure A.1 with labels showing the targets of interest in this study. The second data set consisted of ISAR turntable images of Target G formed every degree over 360° at an incidence angle of 35°. The images at every 10° are shown in Figure A.2. The third data-set consisted of images generated from signature predictions for Target G at 5° for direct comparison with the RAAT trial imagery. These simulated images at every 10° are shown in Figure A.3.

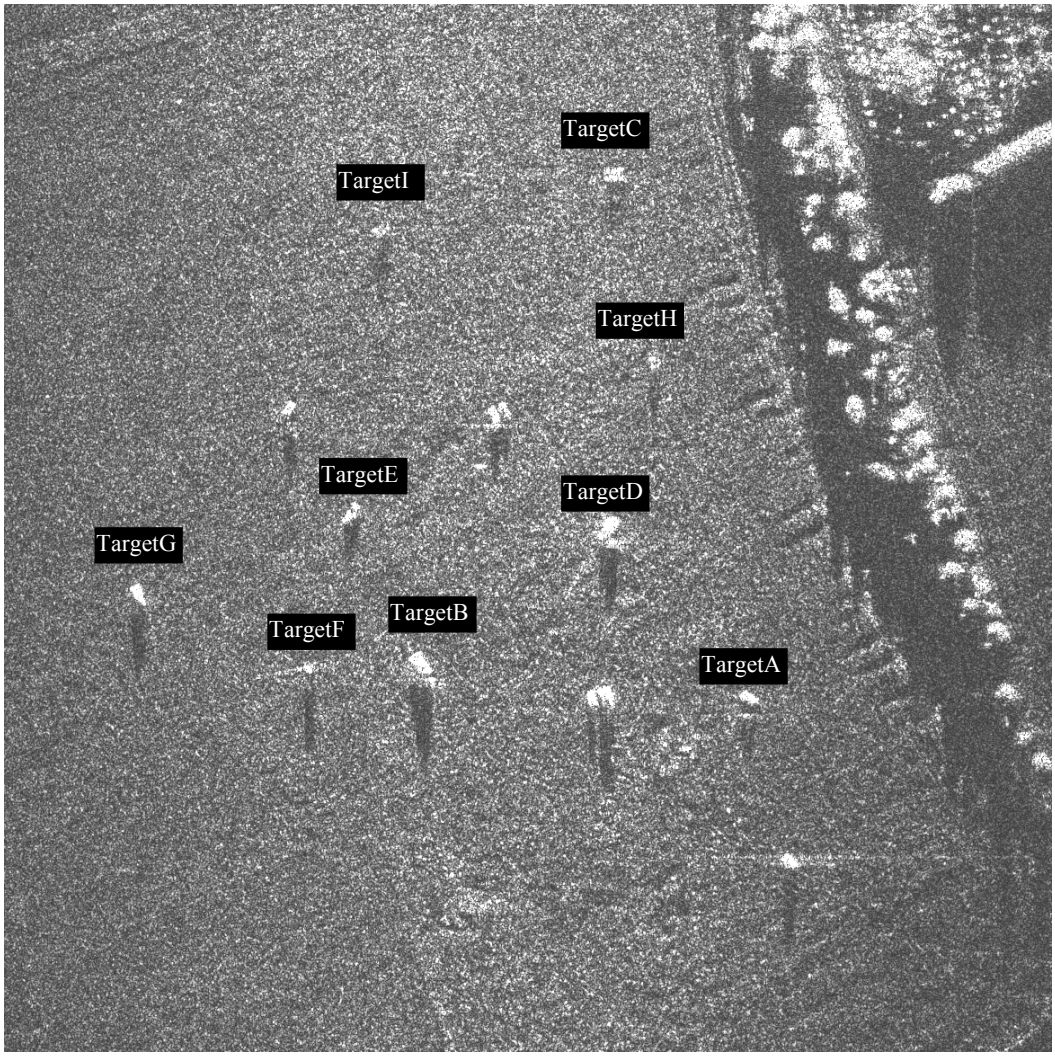


Figure A.1: An example image from the RAAT trial data-set.

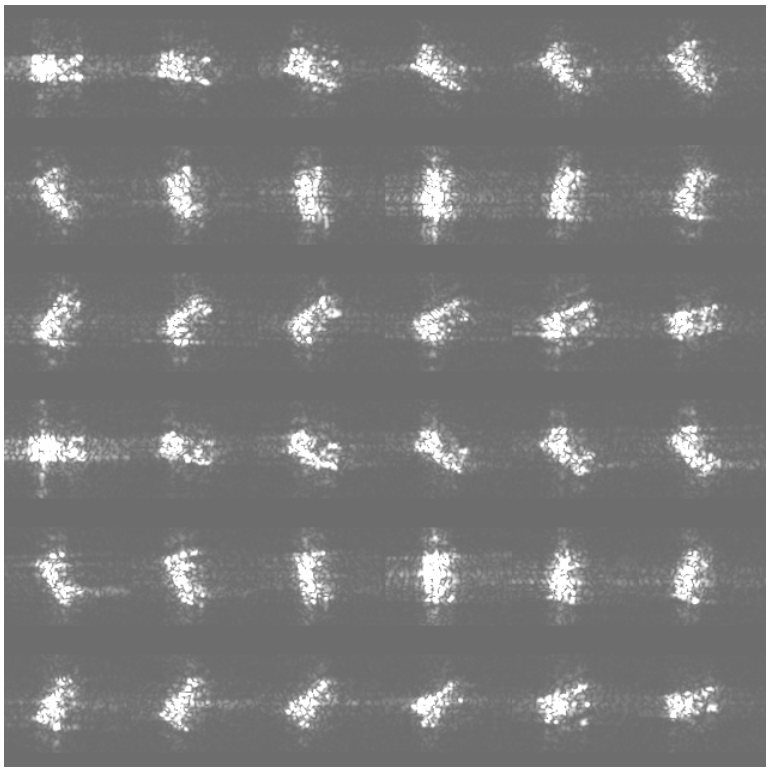


Figure A.2: Images of Target G at 10° intervals from the ISAR turntable data-set

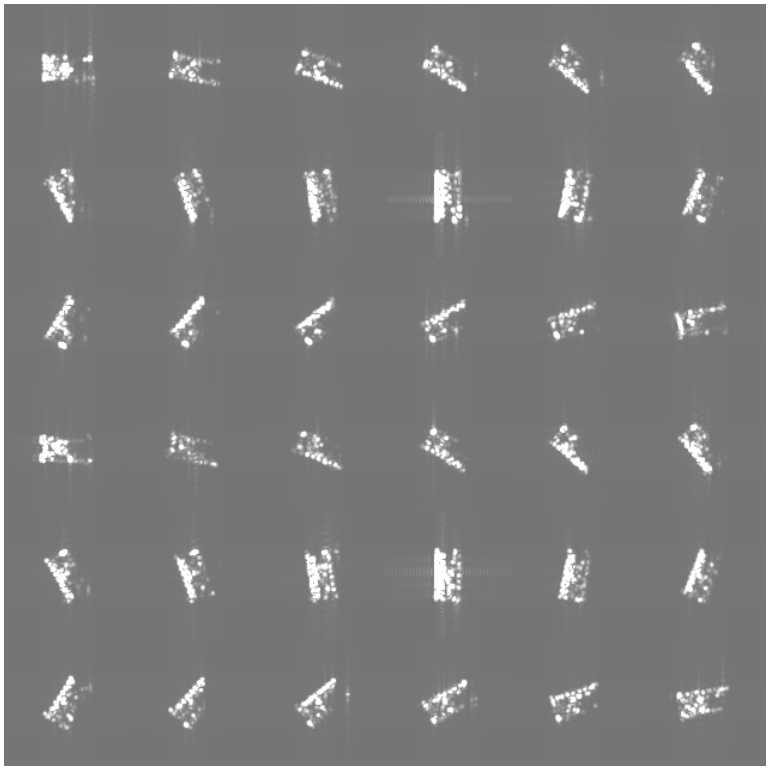


Figure A.3: Images of Target G at 10° intervals from the simulated data-set.