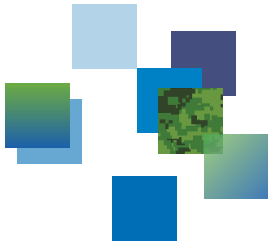
Defence Research and Development Canada
Recherche et développement pour la défense Canada

DRDC | RDDC



Numerical experiments of training Cascade detectors with synthetic sonar imagery

John A. Fawcett
DRDC – Atlantic Research Centre

Defence Research and Development Canada

Scientific Report
DRDC-RDDC-2015-R132
July 2015

Canada

Numerical experiments of training Cascade detectors with synthetic sonar imagery

John A. Fawcett
DRDC – Atlantic Research Centre

Defence Research and Development Canada
Scientific Report
DRDC-RDDC-2015-R132
July 2015

Abstract

Most automated classification and some detection schemes for sidescan or synthetic aperture sonar images are trained methods. The training procedure relies upon the availability of sonar images with a sufficient number of images containing the possible targets of interest. In this report we consider the use of artificially-generated target images to train Haar- and Local Binary Pattern (LBP)-cascade detectors. The advantage of this technique is that one can generate thousands of target images covering a variety of geometries and possible target types. In this report, it is shown that in a straightforward fashion, one can train the detectors with a resulting performance comparable to a method trained with real sonar imagery. Various permutations of the training and testing data, in terms of the target and clutter images, from two different sea trials are considered illustrating the gains and loss of performance from having some knowledge from a new environment. Finally, we examine the possibility of using the synthetic targets injected into real imagery to predict automated target recognition performance for a sonar in a given environment.

Significance for defence and security

It is shown that it is possible to obtain high-performance cascade detectors using synthetic target images for the training. The advantage of this type of approach over using real imagery in the training is that allows the possibility to train a methods for previously unseen (by the sonar of interest) target types or environments. In the latter case, some representative environmental images would be required for training, either from the new environment or from other archival data.

Résumé

La plupart des procédés de classification automatisée et certains procédés de détection pour les images de sonar à balayage latéral ou à ouverture synthétique sont des méthodes expérimentées. La procédure de préparation repose sur l'accessibilité à des images captées à l'aide d'un sonar en nombre suffisant pour contenir des cibles dignes d'intérêt. Dans ce rapport, nous examinons la possibilité d'utiliser des images cibles générées artificiellement pour préparer les détecteurs de Haar et les détecteurs en cascade ayant un profil binaire local. Cette technique présente l'avantage de pouvoir produire des milliers d'images cibles de diverses formes géométriques et de types de cibles variées. Dans ce rapport il est démontré, de façon simple, qu'il est possible de préparer les détecteurs pour qu'ils aient un rendement comparable à une méthode faisant appel à une imagerie réelle par sonar. Diverses permutations des données de préparation et d'essai, en fonction des images cibles et des images qui correspondent à un fouillis d'échos, provenant de deux essais en mer, sont considérées comme des exemples de gains et de pertes de la performance liée à une certaine connaissance d'un nouvel environnement. Enfin, nous étudions la possibilité d'introduire des cibles synthétiques dans l'image réelle pour prédire le résultat de la reconnaissance automatisée des cibles pour un sonar dans un environnement donné.

Importance pour la défense et la sécurité

On a démontré qu'il est possible d'obtenir des détecteurs en cascade à haut rendement à l'aide d'images cibles synthétiques pour la formation. L'avantage de cette approche comparativement à l'utilisation d'une image réelle dans la formation est que celle ci permet d'obtenir une méthode pour des cibles ou des environnements qui n'étaient pas visibles auparavant (par le sonar qui nous intéresse). Dans le dernier cas, des images représentant l'environnement seraient nécessaires pour la formation ; elles pourraient être tirées du nouvel environnement ou d'autres données d'archives.

Table of contents

List of figures

Acknowledgements

The author would like to thank the NATO Centre for Maritime Research and Experimentation (CMRE) for providing the sonar data used in this report.

1 Introduction

In recent years, methods from face- and object-recognition literature have been used for automated sidescan/synthetic aperture sonar (SAS) detection and classification [1–3]. A trained method requires a training set of representative images of the targets as well as non-target imagery. Usually, the number of real sonar images of targets obtained from a trial is fairly small (perhaps, in the hundreds) and there may be circumstances when there are no images of a particular target type. Thus, the use of synthetic (i.e. computer-generated) target images for training could be very useful. However, this depends upon whether a detector trained with synthetic target imagery is, in fact, a good detector for real sonar images with real targets.

There has been other work [4–7] which utilizes synthetic data for the training process. In [4] the authors developed a fairly sophisticated method of inserting artificial targets into background images and then used the resulting images to train a detector/classifier based upon spatial-filter features. They used the trained classifiers and obtained very good results for a test set of real sonar images. In [5], inserted targets were used to investigate the performances of computer detection methods for different seabed types. Ref. 6 considered the insertion of real target snippets into different backgrounds to augment the training set for a Haar cascade detector. The use of inserted synthetic targets is also used in Seabyte's PATT (planning and evaluation software) [7].

In this report, we will utilize sonar data collected with the NATO Centre for Maritime Research and Experimentation (CMRE) MUSCLE system from two different trials, COLOSSUS and MANNEX13. We will investigate a number of different partitionings of the training and testing data, in terms of site and also the use of synthetic or real target imagery. For the "synthetic" cascades, the target mugshots will be constructed by inserting a synthetic target image into real background imagery. The "real" cascades correspond to training with "real" target images. In both cases, the clutter images that are used in the training are real and the same for both types of cascades. We will consider training with imagery from one data set (COLOSSUS) and testing on another data set (MANNEX13). The detection/false alarm results from the "synthetic" and "real" cascades are compared in this case. As well, we also consider cascades trained using the MANNEX13 data: first, just using the clutter files in the training process and then a cascade trained using 1/2 the available MANNEX target images. These numerical experiments indicate the loss of performance when going from real to synthetic targets in the training and also in training with one environment and testing with a new environment. Finally, we will insert synthetic targets into the clutter imagery of the MANNEX13 clutter data to construct a synthetic testing set and we will compare the performance on this synthetic test set to the performance on a real MANNEX13 test set.

2 Detection methods

2.1 Cascade detectors

The main detection method considered in this report is the Haar cascade [2,6,8]. Haar features are a set of simple filter (consisting of rectangles of $\pm$ 1 and possibly rotated) outputs for a fixed window size of image data. For a typical window size, there may be several thousand features. Local Binary Profile (LBP) features are based upon the relative value (bigger or smaller) of the pixel values surrounding a central pixel and this also done with respect to multiple-scaled blocks of data. The openCV software package [9] which we use for the detector implementation allows for the use of Haar, LBP, or Histogram Orientation Gradient (HOG) features when training a cascade. Thus, it is very straightforward to use either the Haar or LBP features (we have not utilized HOG features, thus far in our studies).

For all these features, the basic approach is the same. A set of "positive" windows is prepared—these are a set of fairly-closely cropped images containing targets—they are resized into a set of images of fixed dimensions using an openCV utility [9]. There is also a directory of images (one can use the entire sonar tile in an image format such as TIFF or JPEG) containing no targets (these are negative or clutter images). The training algorithm uses boosting [8] to find a set of the simple features which correctly detects $p\%$ of the targets and for the set of the windows selected from the clutter images, a false alarm rate of $\alpha\%$. This is denoted as a stage. The target images and a set of clutter images, which are not eliminated from the previous stage, are then used to determine a new set of features via boosting for the next stage. This process is continued until the specified number of stages or a desired global false alarm rate is obtained.

This constructed cascade is then utilized by a detection algorithm which finds the windows of image data which successfully pass through all the stages. The detection phase is done very efficiently due to the cascade structure—a large number of data windows are eliminated by the first few stages and thus do not proceed to the later and (usually more features) stages. This detection phase can be done at multiple scalings of the image, so that if an object is simply a scaled version of one used in the training set, it should be detectable. The output of the detector is a set of detection rectangles, specifying the lower left corner of the rectangle and its width and height. There are often many (as many as a thousand in our formulation) rectangles associated with a single detection. The rectangle information is input into a MATLAB routine which provides a smoothed estimate of the number of rectangles (within specified size constraints) as a function of position within the image. This number can be thresholded as a detection statistic.

2.2 Matched-filter detection

As an example of a commonly used, non-trained detector, we consider a matched-filter [10]. We will also show later in this report that the matched-filter output can serve as an useful constraint on the Haar detections. Let us consider a filter consisting of $N_{row} \times N_{colh}$ of ones followed by $N_{row} \times N_{cols}$ of negative ones. This represents a generic region of highlight followed by a shadow region. Sometimes, one inserts columns of zeros between the highlight and shadow regions, representing a transitional region for a target. The number of shadow columns N_{cols} should grow with range to represent the growing length of a shadow with range that a target would cast upon the seabed. For a given sonar altitude a and hypothesized target height h this shadow length SL as a function of target range R can be predicted,

$$SL = \frac{hR}{a-h} \tag{1}$$

or, in terms, of across-track pixel indices $SL(pixels) = SL/\Delta_{ct}$ where Δ_{ct} is the across-track sampling in metres. In this report we take a highlight region of 15 rows and 10 columns and a shadow region of 15 rows and the number of columns grows with range according to Eq.(1). The height of the object of interest is taken to be 0.4 m.

The method of integral images [8] can be used to efficiently compute the output from a simple matched-filter consisting of rectangles of $+1$ and -1 and the shadow length growing as in Eq.(1). The image values $I(i,j)$ themselves are logarithmically transformed into positive and negative values by the transformation,

$$I_L(i,j) = \frac{\log(I(i,j)/\mu)}{\log(2)} \tag{2}$$

where $I(i,j)$ represents an image value, μ is the median image value, and $I_L(i,j)$ is the resultant remapped image. Thus values which are twice the median value get transformed into the value unity and those which are half the median value are mapped into -1. However, the negative values are rescaled and clipped so as to be -1 for values of 1/8 the median, and linearly increasing to zero for values equal to the median value.

One can consider the matched-filter output m_f or one can also compute a mean background value $\bar{m}_f$ of the matched-filter output (the positive values) and normalize the matched-filter outputs by

$$m_f = \frac{m_f}{1 + 4\bar{m}_f} \tag{3}$$

This normalization effectively decreases the matched-filter ouput in regions of high average matched-filter outputs (e.g., in clutter regions).

3 Data collection and preparation

In this report, data from two CMRE trials are used for training and testing. The training data comes from the COLOSSUS trial. This data was collected with a variety of dummy target shapes deployed at 3 different sites and a variety of different types of seabeds. We have utilized this data in the past [11] and computed Haar and LBP-cascades based upon this or portions of this data set. The MANNEX13 data we used comes from one site (there were other sites from this trial that we do not consider) which was mostly benign but had some images containing fields of the sea grass Posidonia. The performance of Haar cascades trained with COLOSSUS data and tested on MANNEX13 data was previously examined in [11] as a function of imagery parameters.

For both data sets, the sonar data are pre-processed in the same manner. The data is converted into decibel units by taking 20Log_{10} of the amplitude data and this data is normalized by computed the mean values along each column (along-track dimension). The image is then two-dimensionally median-filtered and this resulting image is subsampled by a factor of 5 in the across-track dimension and 2 in the along-track dimension. This is done in order to reduce the typical window-size for an extracted target image in order to reduce the number of possible Haar-features computed during the cascade training. This image is then normalized again by the mean of a sliding 150×150 window. Finally, the 0.5% and 99.5% values of the data are found and these values specify the interval $[0, 255]$. The image is then output as a uint8 TIFF image file. Two example resulting images from the COLOSSUS and MANNEX13 trials are shown in Figs. 1 and 2. They illustrate some of the seabed features present at the 2 trials.

For the COLOSSUS data, ground truth positions and manual inspection of the images were used to determine targets in the data. Target mugshots were manually cropped from the images. As well, these basic mugshots were flipped vertically, reshaped in size, and had the cropping window slightly altered in position to increase the number of targets or "positive" images by a factor of 10. The imagery for the MANNEX13 trial was processed in the identical fashion as the COLOSSUS imagery. Most of the MANNEX13 imagery corresponded to the ranges and altitudes of the COLOSSUS data; however, there were some short-range images and also some from a significantly higher altitude.

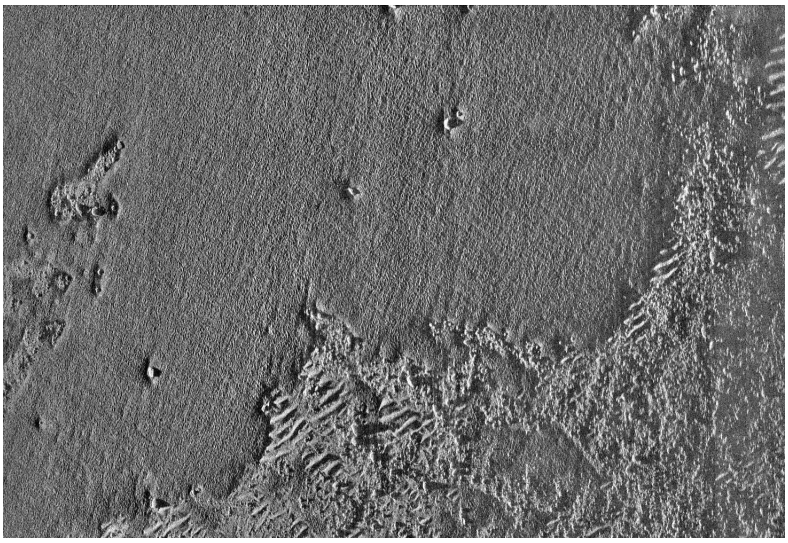


Figure 1: *An example clutter image from the COLOSSUS II set showing some ripples and clutter objects.*

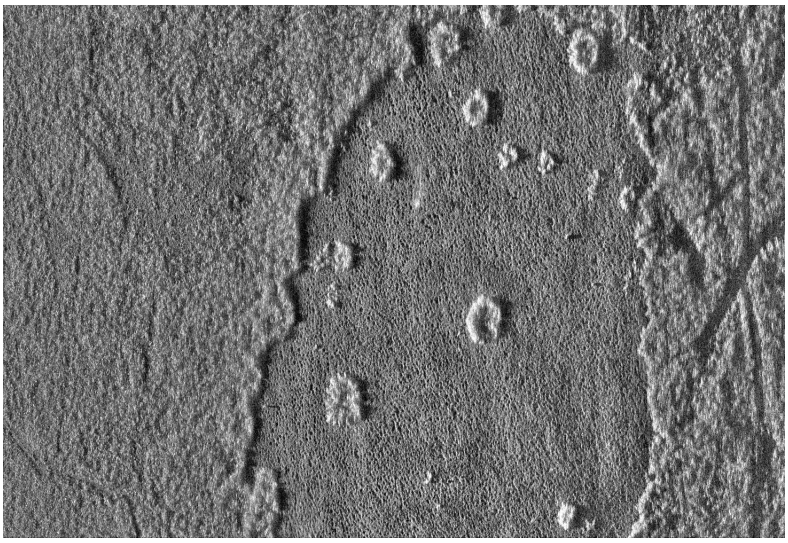


Figure 2: *An example Posidonia clutter image from the MANNEX trial.*

4 Construction of synthetic target images

To construct the synthetic target mugshots, ray-traced templates [1] are used to generate the regions of background, highlight, and shadow on a flat seabed for specified target shape, range, and aspect. A window of a real sonar-clutter image, with the same dimensions as the synthetic mugshot, is also considered. This window may be from a region of a benign background, or a rippled or scoured background, or even a background with rocks. The new mugshot will be a combination of the template and the real background window. First, wherever the template mugshots is denoted as background the new mugshot uses the real imagery. The template's highlight and shadow regions are used in the combination mugshot. However, at the edges of the shadow, a random decision is made whether to use a shadow pixel or use the background level. This creates an irregular boundary for the shadow. In addition, the shadow region is filled in with noise from a Gaussian distribution whose mean and standard deviation is a random factor of the mean level of the clutter image. We also tried using a Rayleigh distribution for the noise but in the numerical experiments obtained somewhat better results with the Gaussian distributions. The highlight regions of the template is also handled in a random fashion. The maximum value of the template highlight is determined. Then all values larger than this value multiplied by a single random scalar are set to a constant random number in the range 180–255. The highlight values less than the threshold are then linearly scaled down. Finally normal random noise, with the background image standard deviation is added to the highlight region. This approach creates a variety of types of highlight regions: ranging from a large portion of the predicted highlight region appearing bright to images where only a small portion of the target highlight appears bright. There are also other random distortions applied to the original templates so that their dimensions and shadow profiles are not perfect. We feel that it is important to add noise and distort the synthetic images. A Haar cascade which is trained with perfect target images and perfectly dark shadows may not be able to detect targets with noisy shadows and imperfect highlight regions.

The noise added to the templates and the scalings are somewhat adhoc and were used after numerical experimentation indicated a good visual match with some real data mugshots. It is difficult to predict theoretically the noise characteristics for the sonar images; for the original sonar imagery, there are distributions (e.g. Rayleigh) describing the expected amplitude of the reverberation. However, we are interested in simulating the TIFF images, which are produced after the logarithm was applied, after median filtering, after normalization and after rebinning the data into [0,255]. There are other complications such as multipath noise in the shadows and issues from the sonar focussing which also cause the sonar statistics to deviate from theoretical distributions. Looking at many real mugshots, there was a wide variety in the distributions of the shadow-pixel values. In the end, we simply went with the distributions

described in the paragraph above.

A directory of images was generated by considering 3 target types: a cylindrical shape, a truncated-cone shape, and a wedge-shape. The templates were generated (we used an existing set) for slant ranges from 25 to 120 m in increments of 5 m and aspect increments of 10° for the cylinder and 5° for the wedge-shape. There is no aspect dependence for the truncated-cone shape. Then for each range and aspect a number of realizations are run (applying the random noise and distortions to the template). In this manner, 17280 synthetic mugshots are created. This is a much larger number of target images than one would normally obtain from trial's data. Of course, the number of real target images can also be increased for training by distortion and by adding realizations of noise. However, synthetically one can generate a fine and equal sampling of target imagery in terms of aspect and range. In Figs. 3 and 4 we show a screen grab of some real mugshot examples from the COLOSSUS trial and some of the simulated mugshot. As can be seen, the simulated images have captured much of the character of the real images. Whether, they will be sufficiently realistic to train a good Haar detector is investigated below.

5 Cascade performances

As previously discussed, we use the openCV software [10] to train, using a set of positive images and a directory of negative images, a Haar cascade. The same routine can also be used to train a LBP-cascade. The basic parameters for the cascade are as follows: (a) window size 24×16 (b) a detection rate of 0.998 per stage and (c) a maximum false alarm rate of 0.4 per stage. We will now consider various partitionings of the training and testing images.

5.1 Train with COLOSSUS backgrounds, test with MANNEX13 images

In this case, the directory of negative images are sonar files (including flipped and slightly resized versions) from all of the sites from the COLOSSUS trial. First, a cascade trained with real target mugshots extracted from the COLOSSUS data set is considered. Second, we consider a cascade trained with synthetic target mugshots. These were generated by inserting the templates into randomly selected windows of the negative images. The templates were modified as described in the previous section. The trained cascades are then applied to a testing set of images from the MANNEX13 data set. The templates were generated for an assumed sonar altitude of 13 m. In fact, for the MANNEX13 data, there were some files at a significantly higher altitude and there were also some sonar images for a shorter range interval.

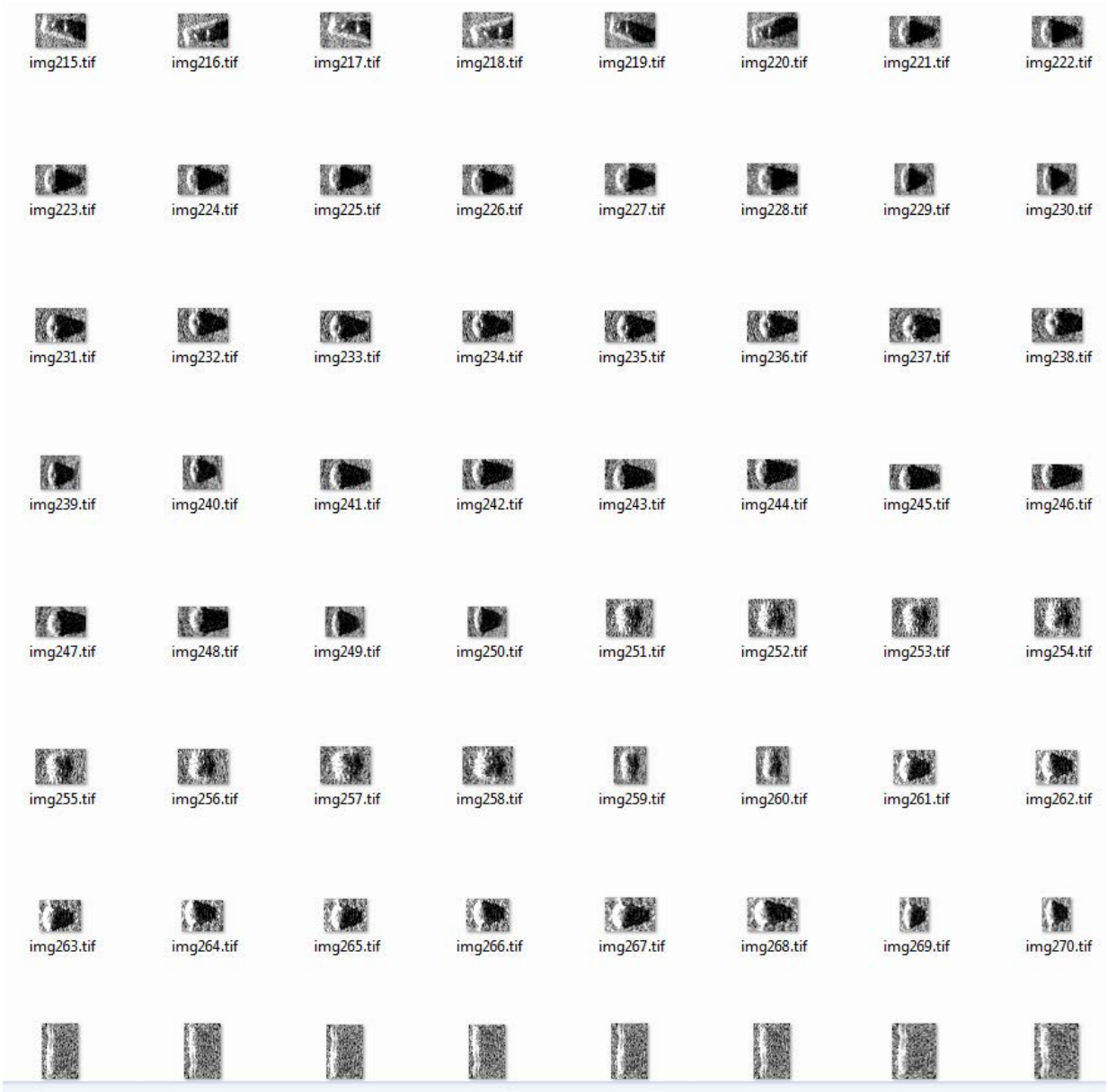


Figure 3: *A screen grab of some of the extracted mugshots from the MANNEX13 data set used for training later in this report.*

Figure 4: *A screen grab of some of the extracted mugshots from the synthetic targets inserted into COLOSSUS backgrounds.*

These files will be included in the analysis of this report. Thus, it should be noted that there are some mismatches between the parameters of the synthetic training and the testing sets. In Fig. 5 we show the resulting ROC curves using the synthetically trained Haar Cascade (red), the real-target trained (blue), the synthetically-trained (red-dashed) when we impose a constraint that a matched-filter ouput should be higher than 0.55 and similarly for the real trained cascade (blue dashed). In Fig. 6 a zoom of the upper left-hand corner of Fig. 5 is shown. The green curve shows the performance of the matched-filter detector by itself. It is clear that all the Haar-based detectors significantly outperform the matched-filter. The real- and synthetically-trained cascades exhibit very similar behaviour. When we also impose the constraint upon the matched-filter output, the synthetically-trained cascade improves in terms of reducing the number of false alarms for the higher detection rates. This is also the case for the real-trained cascade but not as significantly. We did note that the synthetically-trained cascade had false detections in some regions of seemingly innocuous noisy background whereas the real-trained cascade did not. This may be due to including some unrealistic examples of noise in the synthetic training set or it could be that the training learns some of the unrealistic distributions in the shadow and/or highlight regions. It could be that the matched-filter constraint had a particularly beneficial effect in these cases. An example of this type of false detection is shown in Fig. 7. From the ROC curves, it can be seen that even for the lowest level of

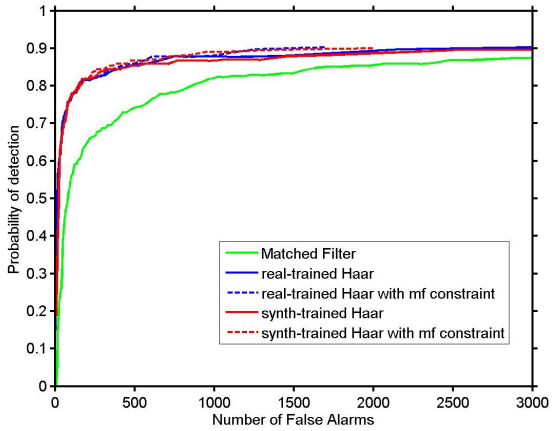


Figure 5: *The ROC curves for (a) the Cascade trained with real COLOSSUS mugshots (blue) (b) the Cascade trained with real COLOSSUS mugshots and a matched-filter constraint (blue-dashed) (c) the Cascade trained with synthetic mugshots (red) (d) the Cascade trained with synthetic mugshots (red-dashed) (e) a matched-filter (green).*

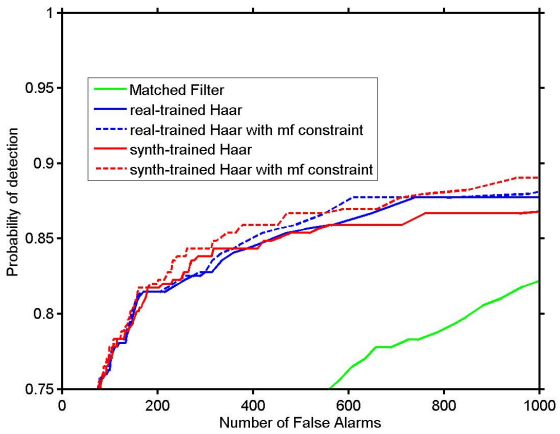


Figure 6: *A zoom of the upper left hand corner of the ROC curve of the previous figure.*

the number of rectangles considered, a number of detections were missed according to the ground truth information that we were using. Many of these detections were very difficult in terms of being barely visible in the background, etc. The synthetically trained cascade was capable of detecting some quite difficult targets. This is shown in Figs. 8 and 9 where a target (a zoom of the target is shown in Fig. 9) with a quite noisy shadow is successfully detected.

The Haar cascades are trained to a maximum number of stages (e.g., 15). In general, the number of false alarms will be reduced as the number of stages is increased at the expense of a somewhat reduced detection rate. However, the optimum number of stages to use with respect to a new data set is unclear. Also, unlike the training procedure, the detection process will use multiple-scales and many rectangles may be associated with a single object. Thus, in terms of the final ROC curve, generated, in this report, by varying the number of rectangles, it is difficult to predict optimal number of stages. The ROC curves shown in shown in this report are for optimal or near-optimal values of the number of stages. In general, it was often better to use less stages than the typical final number specified in the training. Using a lesser number of stages increased the detection rate; however, by increasing the threshold on the required number of rectangles, the false alarm rate can be made similar or even better than using the larger number of stages. The use of the matched-filter constraint can be used to reduce this number of false alarms. One can also train a LBP-cascade with the synthetic data. The resulting ROC curves are shown in Fig. 10 along with the real-trained Haar-cascade curve of Fig. 6. They are quite similar. The real-trained Haar-cascade (green curve) is somewhat better for lower false alarm rates and the LBP curves somewhat better for the high detection rates. In particular, the synthetically-trained LBP cascade performed well with the addition of the matched-filter constraint.

5.2 Training with combinations of MANNEX13/COLOSSUS data and testing with MANNEX13 data

Thus far, we have shown the performance of various cascades trained with one set of data (i.e., COLOSSUS) on a different set of data (MANNEX13). During an AUV survey of a new seabed environment, it should be possible to train a cascade (ignoring the fact that the training process can take some time) with new background images. The new clutter images can be used in conjunction with existing target mugshots from a previous survey or synthetic target images can be injected into the new clutter images. In either case, real target images from the new survey are not used. To investigate cascade performance in this case, we consider our set of MANNEX13 images which contain targets and take one-half the images for training and the other

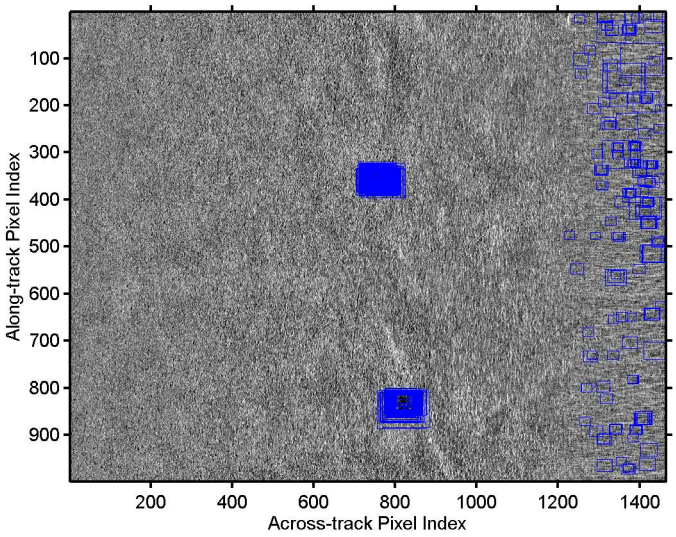


Figure 7: *An example of a image where the synthetic cascade detects noisy false alarms (outer edge of the image).*

half for testing. Target mugshots are extracted from the training set of images. The clutter images are also divided equally into training and testing.

In the ROC curves of Fig 11, we consider a Haar cascade trained using the extracted MANNEX13 mugshots and clutter images. This serves as a benchmark case—a cascade trained with both real targets and background. The other curves correspond to cascades trained with synthetic targets inserted into MANNEX13 clutter images (training set) and the training clutter images also used as the background set. The curves corresponding to using the COLOSSUS targets with MANNEX13 clutter is also shown. The cascade corresponding to training using targets and clutter from the MANNEX site (blue) yielded the best performance. However, the cascades trained with synthetic targets/MANNEX13 background (green) and real COLOSSUS targets/MANNEX13 clutter (red) yield very similar performance until approximately the 80% detection rate. These 2 cascades were better (in terms of false alarm rate) than the cascade trained utilizing only COLOSSUS data (cyan curve) for detection rates in the interval 65-80%, after which the performances were all quite similar. These results seem to indicate that there were some differences in the mugshots of the harder-to-detect target MANNEX images which were not present in the synthetic or COLOSSUS target images (also, recall that some of the MANNEX13 images corresponded to a higher altitude than was present in the COLOSSUS data or the simulation) and thus at the higher detection rates the cascade trained with true MANNEX13 target images is superior.

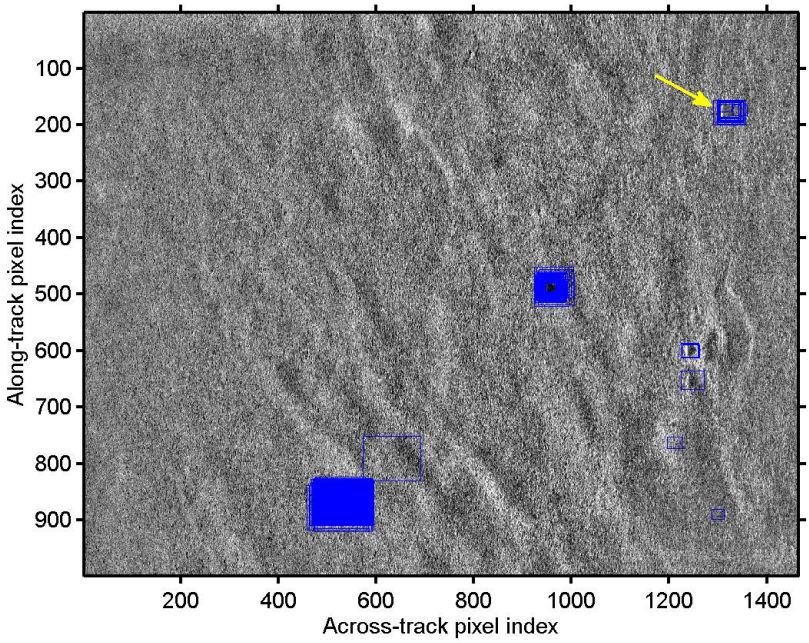


Figure 8: *An example image showing 3 valid detections (the large concentrations of blue rectangles).*

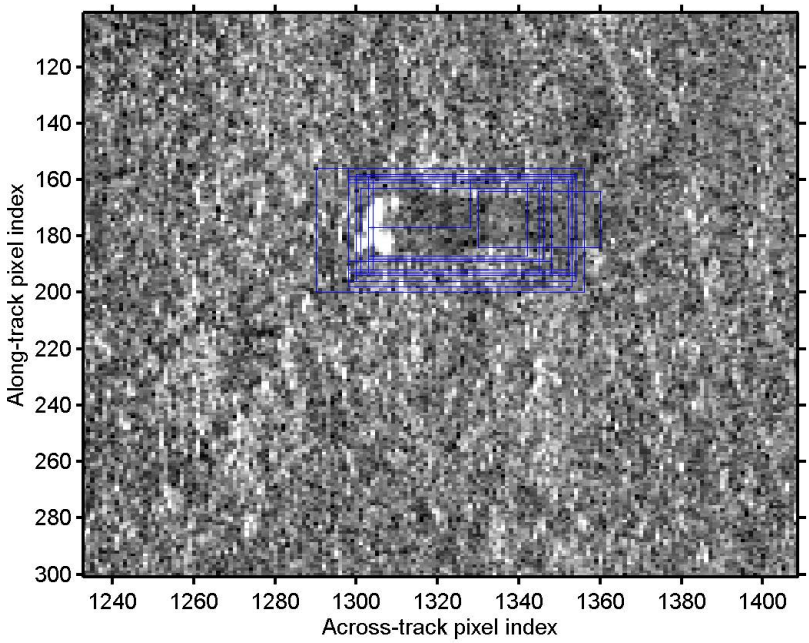


Figure 9: *A zoom of the detection indicated by a yellow arrow in the preceeding figure.*

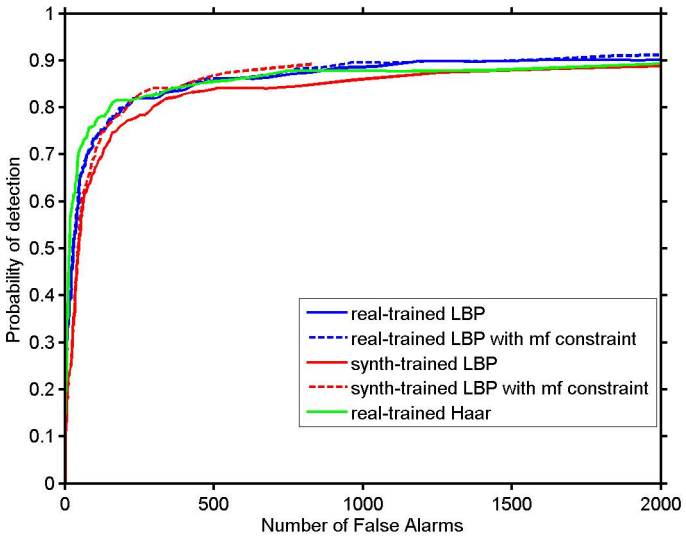


Figure 10: *The ROC curves for (a) the LBP Cascade trained with real COLOSSUS mugshots (blue) (b) the LBP Cascade trained with real COLOSSUS mugshots and a matched-filter constraint (blue-dashed) (c) the LBP Cascade trained with synthetic mugshots (red) (d) the LBP Cascade trained with synthetic mugshots (red-dashed) (e) the real-trained Haar cascade (green).*

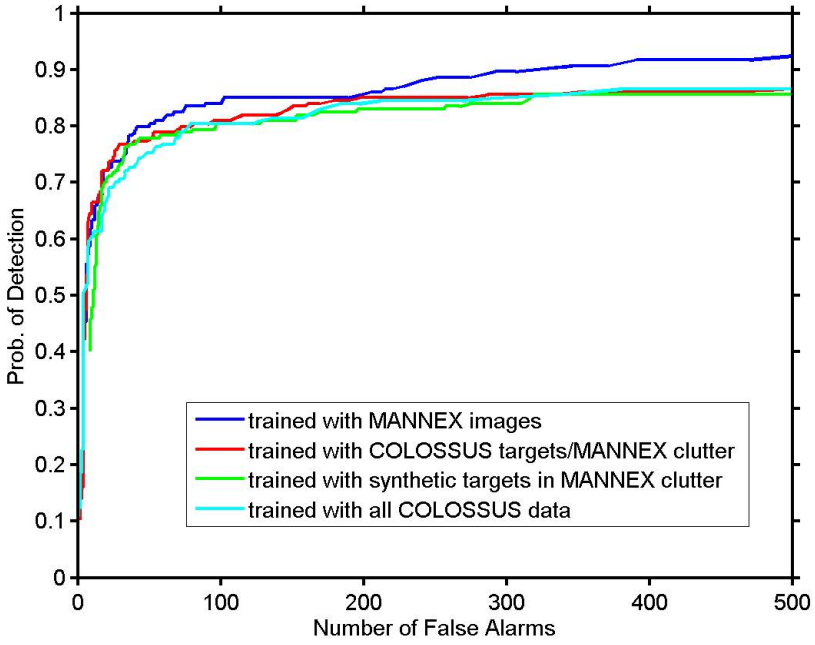


Figure 11: *The ROC curves for various cascades trained with varying amounts of MANNEX13 imagery.*

6 Performance prediction

In the previous section, we considered the performance of various cascades as detectors. It was shown that one could successfully use synthetic target imagery in the training process. In this section, the use of synthetic targets in imagery as a means of predicting detector performance is examined. For example, we can take the MANNEX13 clutter images from the testing set and, using the methodology described earlier, insert synthetic targets (distributed randomly in terms of range, type, and aspect) into these images. An example of such a constructed image is shown in Fig. 12. It was found that using our random rules of noise generation, that there seemed to be too many inserted targets with large shadow/background contrasts. This was particularly evident at larger ranges, because in the target mugshot construction process, we had not built in any rules regarding noise and range.

We changed the target simulation program to generate targets with, on average, a weaker highlight region and also with a higher probability, lower shadow/background ratios. This probability increased with range. An example of a synthetic images generated with this model is shown in Fig. 13. In Fig. 14 we show the resulting ROC curve (blue) for a cascade trained with 1/2 the MANNEX data and then tested with the remaining images. The ROC curves for the Haar-cascades trained with synthetic targets inserted into MANNEX clutter images (from the training set) and then tested on synthetic testing sets (Model 1 - red, Model 2 - cyan) are also shown. As can be seen, the inserted targets from noise Model 1 create a synthetic test set which is overly optimistic and the ROC curve reaches a very high detection rate with low false alarm. The synthetic target Model 2 used weaker highlights and a more noise in the shadows with increasing probability with increasing range. It can be seen that the resulting ROC curve is more realistic and follows the real ROC curve (blue) fairly closely. Finally, the green ROC curve is the curve resulting from using the same cascade (as with the synthetic testing sets) with the real testing set. Here, it can be seen that this cascade follows the blue and cyan curves fairly closely until about the 80% detection rate (a zoom of the curves of Fig. 14 is shown in Fig. 15). For the real testing set, it then starts to fail to detect the "harder-to-detect" targets. The Model 2 for the targets in the synthetic testing set produced a more realistic ROC curve - so, perhaps better detection results would be obtained if that Model was used in producing the target mugshots for training the cascade—we tried this, but, in fact, the resulting performance was poorer than using the less-noisy mugshots for training.

In summary, we were able to replicate the ROC curve for a trained cascade by using a synthetic testing set; however, this required us to modify our initial noise-model for the target insertion. The synthetic curve actually followed the curve for the cascade trained with real MANNEX13 targets better than the one trained with synthetic targets. This may be because the match between the training and testing targets is

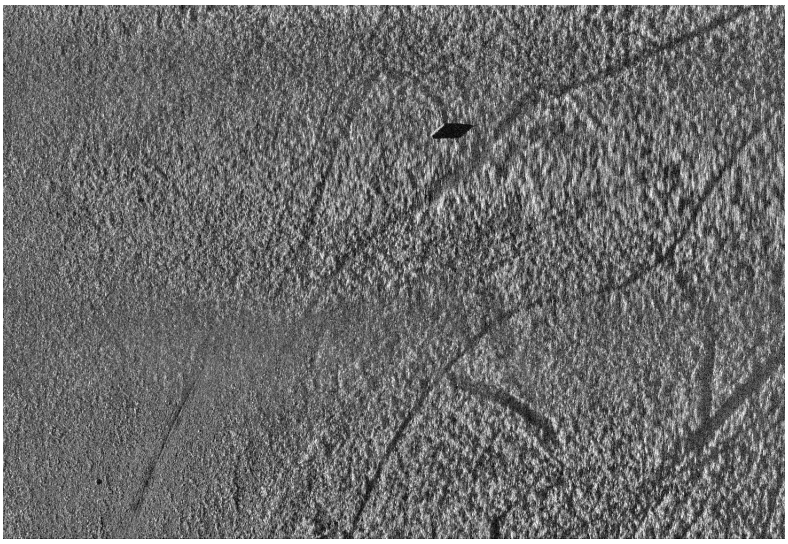


Figure 12: *An example of an inserted synthetic target where the contrast seems somewhat unrealistic—Model 1.*

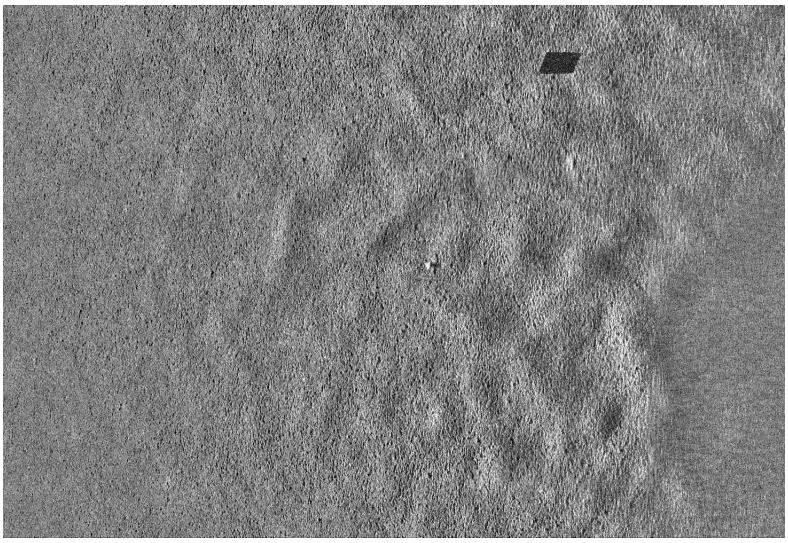


Figure 13: *An example of an inserted synthetic target where the contrast seems more realistic—Model 2.*

closer in this case and more accurately reflects the match between using the same targets from a trial for training and testing.

7 Summary and discussion of results

In this report, we have investigated the use of synthetic target imagery in training Cascade detectors. The synthetic target images were constructed by inserting ray-trace target templates into real image windows. Noise was added to the the templates and various small deformations were applied to make the target appearance more realistic. The clutter images used in the cascade training were real images. We also considered cascades trained with real target mugshots. In the first set of examples, the training images were from the COLOSSUS trial and the testing set was images from the MANNEX13 trial and thus there was a mismatch in the seabed environments. The synthetically-trained cascades yielded a detection performance very similar to that of the real-trained cascades. It should be noted that the synthetic target data set is much larger than the real target mugshot data sets and was finely and uniformly sampled in target aspect and range. From that point of view, it is a little surprising that the synthetic cascade did not outperform the real-cascades. This seems to indicate that despite the smaller number of images, that the real target training sets were sufficiently well-sampled. The synthetic cascades did sometimes give a large number of false detections for some noisy images in the testing set whereas the real-cascades did not. Perhaps, this is due to inaccurately modelling the noise distributions for the shadow and highlight when inserting the templates into the background. It was found that the performance of the cascades (both synthetic and real) could be improved by also insisting that a detection region have a sufficiently high matched-filter output. Cascades using LBP features were also considered and the performance on the testing set was quite similar to the Haar-cascade results. Despite the fact that the synthetic cascades did not outperform the real cascades for the examples of this report, there are scenarios where the synthetic training should be useful: (1) if there is target-type which is not present in archival data (2) if it is a new sonar for which there are few available target images and perhaps also in the case that (3) the environment is very different those represented in archival data (although in this report, the COLOSSUS data appeared to be sufficiently "close" enough to the MANNEX13 imagery to yield very good results.)

In the second set of numerical experiments, we investigated the effects of including some clutter images from the testing (new) site in the training. This was to simulate an AUV survey of a new region, during which a set of new background images can be acquired and used to train a new cascade, either with a set of target images from a previous trial (e.g., COLOSSUS) or with synthetic targets inserted into the new background images. The performance results indicated that simply using the

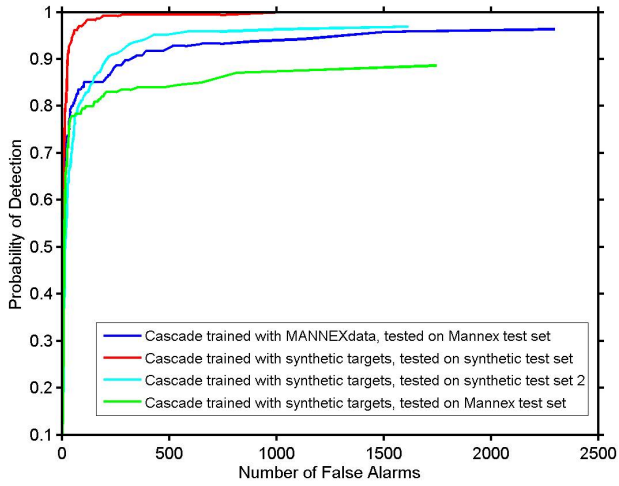


Figure 14: *The ROC curves resulting from using real or synthetic test sets.*

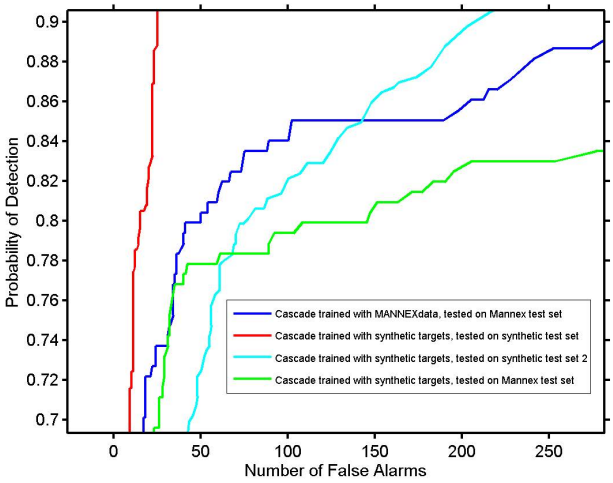


Figure 15: *The low false alarm rate region of the ROC curves in more detail.*

COLOSSUS-trained cascade yielded a very good ROC curve (indicating that there was likely sufficient environmental diversity in the COLOSSUS set to also perform well in the MANNEX13 environment). There was some improvement in the performance of the cascades (real and synthetic) which were trained with MANNEX13 backgrounds for the detection rates of 65–90%. The best cascade for this new environment was, not surprisingly, the one trained with clutter and half the available target images from the MANNEX13 site.

Finally, the use of synthetic targets inserted into imagery as a means of predicting performance was investigated. Using the original template-model to insert targets into background images to form a testing set yielded a ROC curve which was significantly overly-optimistic compared to the performance of a MANNEX13-trained cascade on MANNEX13 data. A second synthetic target model was used with increased shadow noise, reduced highlight and a noise-variation with range. This resulted in a performance estimation which showed good overall agreement with the real Cascade performance. However, it was still overly optimistic at high ($> 80\%$) for the synthetic cascade applied to the real MANNEX13 testing set. It did show quite good agreement for detection rates $\leq 80\%$.

Overall, it is encouraging that the synthetically-trained cascade detectors performed well. This indicates, that when one lacks target images of: (a) a particular target type (b) for a specific sonar or (c) a new environment, synthetically-trained cascades could be successfully used. In general, one could combine the synthetic images with real images. Our synthetic model could certainly be improved to include more realistic shadow and highlight statistics and a more realistic dependence upon the surrounding environment. These improvements could improve the cascade performance and also improve the realism when used to predict detector performance.

References

[1] V. Myers and J. Fawcett, “A template matching procedure for automatic target recognition in synthetic aperture sonar imagery,” IEEE Signal Processing Letters, vol. 17, pp. 683–686, 2010.

[2] J. Sawas, Y. Petillot, and Y. Pailhas, “Cascade of boosted classifiers for rapid detection of underwater objects,” in *Proceedings of ECUA 2010*, Istanbul, Turkey, 2010.

[3] P. Hollesen, W. Connors, and T. Trappenberg, “Comparison of learned versus engineered features for classification of raw sonar images,” in *Canadian AI’11 Proceedings of the 24th Canadian conference on Advances in artificial intelligence*, pp. 174–185, Springer-Verlag, 2011.

[4] E. Coiras, P.Y. Mignotte, Y. Petillot, J. Bell, and K. Lebart, “Supervised target detection and classification by training on augmented reality data", IET Radar Sonar Navig., 2007, Vol. 1, pp. 83–90.

[5] Y. Petillot, S. Reed., E. Coiras, "A framework for evaluating underwater mine detection and classification algorithms using augmented reality", *presented at UDT Conference 2006*, Hamburg, 2006.

[6] C. Barngrover, R. Kastner, and S. Belongie, “Semisynthetic versus real-world sonar training data for the classification of mine-like objects”, IEEE Journal of Oceanic Engineering, Vol.40, No.1, January 2015.

[7] P.Y. Mignotte, J. Vazquez, J. Wood, S. Reed, “PATT: A Performance Analysis and Training Tool for the assessment and adaptive planning of Mine Counter Measure (MCM) Operations planning”, *in Proceedings of Oceans 2009, Biloxi*, 2009.

[8] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” IEEE CVPR, 2001.

[9] Open Source Computer Vision OpenCV 2.4.3 , http://opencv.willowgarage.com/wiki/, (Access date:January 2013).

[10] J. Fawcett and W. Connors, “Adaptive automated detection for synthetic aperture sonar images using seabed classification”, in *Proceedings of the third International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar*, Proceedings of the Institute of Acoustics, Lerici, Italy, 2014.

[11] G. Dobeck, J. Hyland, and L. Smedley, “Automated detection/classification of seamines in sonar imagery,” in *Proceedings of SPIE*, vol. 3079, pp. 90–110, 1997.

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security markings for the title, abstract and indexing annotation must be entered when the document is Classified or Designated.)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

DRDC – Atlantic Research Centre
PO Box 1012, Dartmouth NS B2Y 3Z7, Canada

2a. SECURITY MARKING (Overall security marking of the document, including supplemental markings if applicable.)

UNCLASSIFIED

2b. CONTROLLED GOODS

(NON-CONTROLLED GOODS)
DMC A
REVIEW: GCEC DECEMBER 2012

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Numerical experiments of training Cascade detectors with synthetic sonar imagery

4. AUTHORS (Last name, followed by initials – ranks, titles, etc. not to be used.)

Fawcett, J. A.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total containing information. Include Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
July 2015	32	11

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Scientific Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

DRDC – Atlantic Research Centre
PO Box 1012, Dartmouth NS B2Y 3Z7, Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
01cb	

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC-RDDC-2015-R132	

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

(X) Unlimited distribution
( ) Defence departments and defence contractors; further distribution only as approved
( ) Defence departments and Canadian defence contractors; further distribution only as approved
( ) Government departments and agencies; further distribution only as approved
( ) Defence departments; further distribution only as approved
( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11)) is possible, a wider announcement audience may be selected.)

unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

Most automated classification and some detection schemes for sidescan or synthetic aperture sonar images are trained methods. The training procedure relies upon the availability of sonar images with a sufficient number of images containing the possible targets of interest. In this report we consider the use of artificially-generated target images to train Haar- and Local Binary Pattern (LBP)-cascade detectors. The advantage of this technique is that one can generate thousands of target images covering a variety of geometries and possible target types. In this report, it is shown that in a straightforward fashion, one can train the detectors with a resulting performance comparable to a method trained with real sonar imagery. Various permutations of the training and testing data, in terms of the target and clutter images, from two different sea trials are considered illustrating the gains and loss of performance from having some knowledge from a new environment. Finally, we examine the possibility of using the synthetic targets injected into real imagery to predict automated target recognition performance for a sonar in a given environment.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

automatic target recognition; sonar; sidescan

DRDC | RDDC

SCIENCE, TECHNOLOGY AND KNOWLEDGE FOR CANADA'S DEFENCE AND SECURITY | SCIENCE, TECHNOLOGIE ET SAVOIR POUR LA DÉFENSE ET LA SÉCURITÉ DU CANADA

www.drdc-rddc.gc.ca

DCGER-template-20130914.rtf

Submission-template.csv