

# Composite wavelet filters for enhanced automated target recognition

Jeffrey N. Chiang<sup>a</sup>, Yuhan Zhang<sup>b</sup>, Thomas T. Lu<sup>1c</sup>, and Tien-Hsin Chao<sup>c</sup>  
<sup>a</sup>University of California, Los Angeles, <sup>b</sup>Onescreen Inc, Irvine, <sup>c</sup>Jet Propulsion  
Laboratory/California Institute of Technology, Pasadena, CA 91109

## ABSTRACT

Automated Target Recognition (ATR) systems aim to automate target detection, recognition, and tracking. The current project applies a JPL ATR system to low-resolution sonar and camera videos taken from unmanned vehicles. These sonar images are inherently noisy and difficult to interpret, and pictures taken underwater are unreliable due to murkiness and inconsistent lighting. The ATR system breaks target recognition into three stages: 1) Videos of both sonar and camera footage are broken into frames and preprocessed to enhance images and detect Regions of Interest (ROIs). 2) Features are extracted from these ROIs in preparation for classification. 3) ROIs are classified as true or false positives using a standard Neural Network based on the extracted features. Several preprocessing, feature extraction, and training methods are tested and discussed in this paper.

**Keywords:** Automated Target Recognition, wavelet filter, sonar video image processing.

## I. INTRODUCTION

Automated target recognition (ATR) has been a focus of image processing and artificial intelligence research for some time, with immediate applications in surveillance and autonomous navigation. There are several approaches to ATR, each with limitations in performance and resources. It is very difficult to develop a generalized ATR algorithm due to the complexity of targets and background in various real-world applications. NASA-JPL has designed a multi-stage ATR framework that can be modified and optimized for different inputs [1-2]. First, an image is de-noised and regions of interest (ROIs) are detected using various image filtering techniques [3]. Then a feature extraction is performed on these ROIs, and finally, a neural network is used to classify each ROI as a target or a false positive. The goal of this project was to apply this framework and to optimize it for use with low resolution sonar and camera videos taken by an unmanned underwater vehicle (UUV). A composite wavelet filter and feature extraction techniques are implemented to generate the highest detection rate possible while keeping a low false positive count.

## II. BACKGROUND

### 2.1 ATR Framework

The multi-stage ATR framework developed at NASA-JPL generally breaks image processing into three steps. These steps can be summarized as preprocessing, feature extraction, and false positive reduction, as shown in Figure 1. In this way, speed and accuracy can be balanced. Generally, computationally expensive processing techniques such as filtering and neural networks must be optimized to achieve real-time operations in a computer. The ATR framework aims to identify ROIs that could contain targets in the first step. These ROIs are passed into the neural network for classification--which allows for much faster computation time when compared to analyzing the entire image.

In this framework, preprocessing is defined as any transform performed on the image before features are extracted. Typically, the raw image can be normalized using a histogram filter, in order to enhance contrast. To

---

<sup>1</sup> e-mail: Thomas.T.Lu@jpl.nasa.gov , Tel: (818) 354-9513, Fax: (818) 393-4272

identify targets, a target filter is constructed, either using a simple image editing program or a wavelet function. The image and filter are correlated in the Fourier domain, and a threshold is applied to determine ROIs. This step is responsible for the upper limit of the system's performance--that is, if a target is not detected in this step, it will never be detected by the ATR system. Therefore, it was important to keep the threshold relatively low, to increase the likelihood that the actual target is present in the ROIs detected.

While preprocessing and ROI detection determine percentage of targets actually "caught" by the system, feature extraction and the classification algorithm are essential for the false positives elimination. The goal of feature extraction is twofold: first, to provide essential features that will help the classifier differentiate between true targets and false positives, and second, to reduce the dimensionality of the ROIs to cut computation time. In this project, features were selected in an attempt to quantify our own observations and classifications of the images. As will be shown later, the extracted features are mostly responsible for the average number of false positives per image.

Finally, a reliable classification algorithm must be used to classify ROIs as targets or false positives [4]. Typically a gradient descent back-propagation neural network has been used in the ATR system since it is producible in hardware. These neural networks employ supervised learning algorithms. Thus it was necessary to first train the network on sample data, described in detail in the methods section.

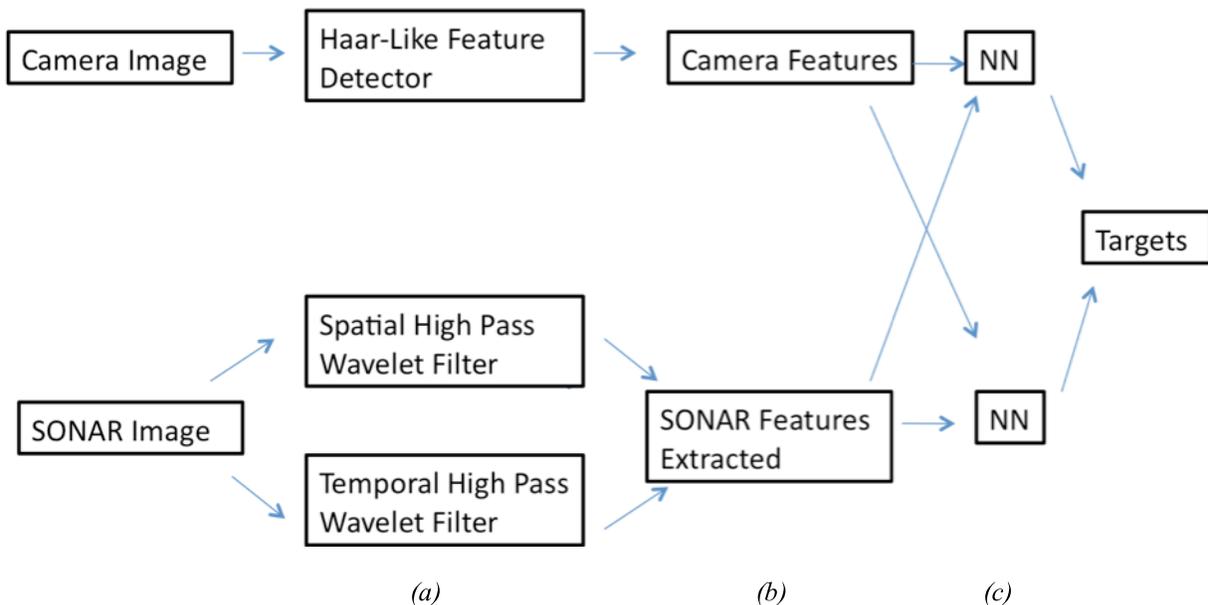


Figure 1. Representation of ATR System. In a) ROI detection occurs using a Haar-Like Feature detector for camera images, and composite filters for sonar images. b) Features are extracted from the detected ROIs, correlated with each other and passed into neural networks for classification (c).

## 2.2 Video Dataset

The sonar images used in this study were extracted from videos taken by UUVs. Because of the nature of sound waves, sonar images are visualized as arcs. The UUV also had a video camera that took continuous video images of the underwater target. The camera images used in this report were generally very low quality. Previous work used Haar-like filters to enhance the camera image quality.

We worked with three video data sets, each with sonar images and camera images. Previous work had been done using Haar-like features to analyze the camera images. However, the sonar images were taken at five different ranges, in which the targets appear very different from each other. Thus we aimed to generate a robust ATR system for the sonar component that was reliable both across ranges (within each video), and across the three videos, which would complement the camera ATR system.

For ease of processing, each sonar frame was converted to a 64 x 640 pixels rectangular image, using a process analogous to converting polar coordinates to Cartesian space.

### III. METHODS

#### 3.1 Preprocessing using the Mexican-Hat Wavelet Filter

In image processing, the Mexican-Hat wavelet is a commonly used filter for edge and blob detection [5]. The function is sometimes approximated by a difference of two Gaussian curves, but in this study the following was used:

$$\varphi(t) = \frac{2}{\sqrt{(3\sigma)\pi^4}} \left(1 - \frac{x^2}{\sigma^2}\right) e^{\frac{-x^2}{2\sigma^2}} \quad (1)$$

where  $\sigma$  is a scale factor.

The Mexican Hat wavelet, as shown in Figure 2 in 1-D and 2-D formats, is an intuitive choice for detecting targets in the provided sonar images. Figure 2(b) shows a fan shaped image that simulates the sonar returns. Targets and edges corresponding to the wavelet are amplified, while all other pixels are suppressed. However, the size, intensity, and shape of targets vary across different ranges within videos. They also vary substantially across videos, making it unfeasible to use one simple filter in the system.

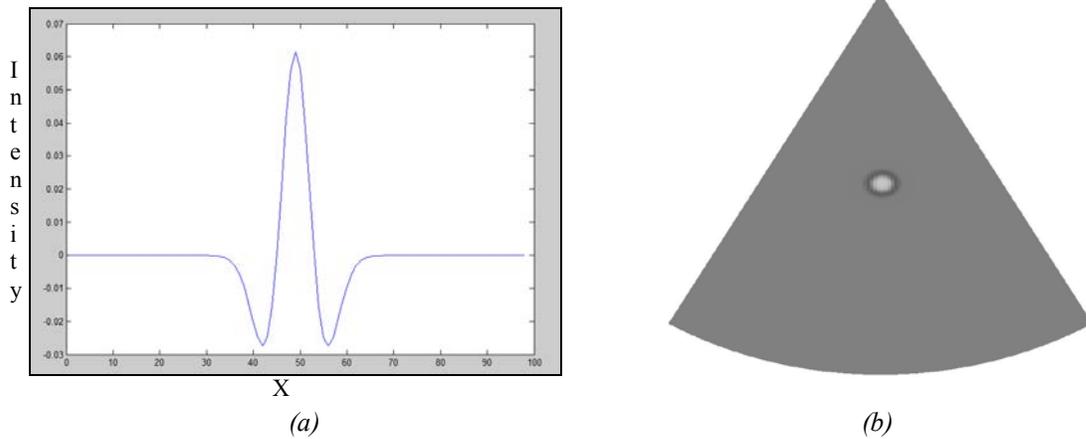


Figure 2. Plots of (a) 1-D Mexican-Hat Filter, and (b) a 2-D Single Mexican-Hat Blob Filter in fan shaped sonar return pattern.

To remedy the problem of variance within ranges, a composite filter made from two Mexican-Hat wavelets was used. Essentially, the filter searched the image for two differently shaped blobs: a small, bright blob corresponding to the target, and a long dark blob immediately below it corresponding to its shadow. This was implemented as a linear combination of two wavelets. Several of these filters were generated, corresponding to a few exemplars arbitrarily picked from the image set. For each range of sonar, five of these composite filters were averaged to create a generalized filter. This totaled to five filters, one for each range of sonar, eliminating the need to generalize across ranges. Figure 3 illustrates the filters used at each range.

Images were filtered temporally and spatially in the Fourier domain. For the temporal filter, filtered images were compared with the average of the last five images, essentially detecting motion. The spatial domain simply searched for ROIs in the current image. The two filtered images were combined using a weighted average to create the decision image, which was used to determine ROIs.

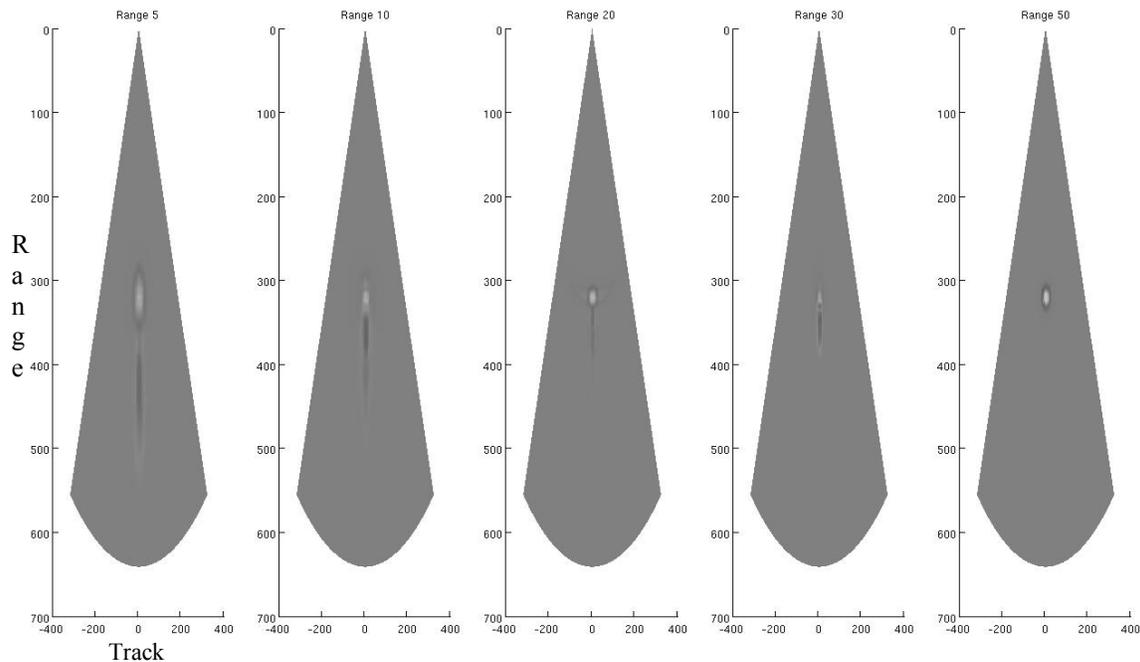


Figure 3. Composite Filters at different ranges in the shape of the sonar return. These are a linear combination of two blob wavelet filters, one corresponding to the target and another corresponding to the shadow.

### 3.2 Feature Extraction

The seven features chosen to classify ROIs were based on previous ATR work and simple intuition and observation.

1. Distance of the target: The sonar data provides the range (distance) information. It appeared that targets tended to appear in the middle ranges of the images.
2. Area of the target: This was found by running a Mexican-Hat filter (designed only for the target) on the raw ROI. The area was represented by the percentage of the ROI covered by the response from the filter. This was chosen because targets are different sizes at various ranges, so an object too big or too small could be dismissed as noise.
3. Ratio of size to distance of target: A close up target would become smaller as it moved further away. Since a neural network deals with linear transformations of its inputs, providing this ratio would emphasize the relationship between area and distance of the target.
4. Peak to side-lobe ratio (PSR): This took the ratio of the highest intensity regions in an ROI to the surrounding area. This was used to represent the intensity and contrast of potential targets.
5. Area of the shadow of the target: A Mexican-Hat filter was used to find the shadow below the ROI, and its area was expressed as a percentage. Since all true targets cast a sound shadow, it was crucial that any ROI with a nonexistent shadow was eliminated as a false positive.
6. Ratio of width to height of shadow: The shadows also appeared to have a distinctive size and shape depending on range. This was a rough attempt to quantify the dimensions of the shadow.

7. *Correlation with camera image*: This was a binary variable for whether the azimuth of the target corresponded with the X coordinate in the camera image. If an ROI corresponded with an object detected in the video camera, it was more likely to be a true positive.

### 3.3 Neural Network Classifier

A three layer feedforward backpropagation neural network was trained to classify ROIs as true or false positives [4,6]. In addition to the seven features outlined above, an additional five binary inputs were included to specify what range the image was taken from. With these dummy variables, the neural network essentially used a different set of parameters for each range--the equivalent of having a neural network for each of the five ranges. With 12 input features, and arbitrarily chosen 11 hidden units were used in the false positive eliminator. The specific neural network used a Levenberg-Marquardt minimization learning algorithm, with random starting weights. Because the starting weights affect the performance of the network, 50 networks with random starting weights were trained and the best one was kept.

The performance of the neural network was measured by entropy: first, all the training data were given to a neural network. The neural network assigned each training data a scalar value (neural network output). If the neural network output for a data point was larger than a threshold, the data point was labeled as a target. The training data were then sorted by their neural network value after classification. All possible threshold values, which divide the training data into two sets (true and false positive), to separate the training data were tested. Then entropy (randomness) for a given threshold is calculated in the two sets using:

$$Entropy = P_A \cdot (-P_{A1} \ln P_{A1} - P_{A0} \ln P_{A0}) + P_B \cdot (-P_{B1} \ln P_{B1} - P_{B0} \ln P_{B0}) \quad (2)$$

where  $P_A$  and  $P_B$  are the probabilities of being in category A and B, respectively;  $P_{A1}$  and  $P_{A0}$  are the probabilities that the target is correctly or falsely identified in category A, respectively;  $P_{B1}$  and  $P_{B0}$  are the probabilities that the target is correctly or falsely identified in category B.

The threshold that gave the smallest entropy was considered as the optimal threshold for the neural network. And the neural network resulting with smallest entropy in the training data was considered to have the best separation performance, and thus kept.

There were several approaches as to how to provide the best training data to the neural network. Two approaches were taken: first, a neural network was "optimized" for each video-- that is, only trained with exemplars from one video. Then, a mixed training set of about 5000 frames was hand-picked from the three videos and used to train another neural network, totaling four different neural networks-- three "optimized" videos, and a combined neural network.

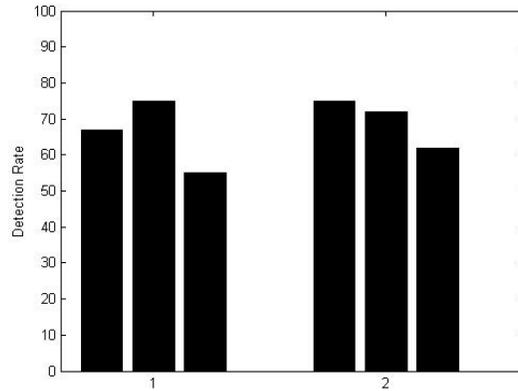
To train the network, a program was written to visualize the feature vectors of positive and negative examples to aid in training. Training sets were split into three fractions: 70% was used for back-propagation, 15% was used for validation, and the last 15% was used to evaluate the neural network's ability to generalize. The best network was selected for the following tests.

## IV. EXPERIMENTAL RESULTS

### 4.1. Detection Rate

After adjusting the filter for each range of object by varying variance, scale and shift, detection was dramatically improved—from an average 20% catch rate using one Mexican-Hat wavelet filter to around 70% using the composite filters. Category 2 in figure 4 shows the maximum detection rate for the three training videos. Although the histogram filter was described earlier as a method for normalizing intensity values, it was also found that this filter enhanced noise, making the filters less effective at detecting ROIs.

To check the robustness of the ROI detection component, the composite wavelet filter was applied to novel videos. Detection rate was comparable to performance in the videos used for training. Figure 4 shows the detection rate for the training videos (Category 2) and new videos (Category 1). Both categories perform similarly. Therefore it suggests that the composite wavelet is robust across different image types.



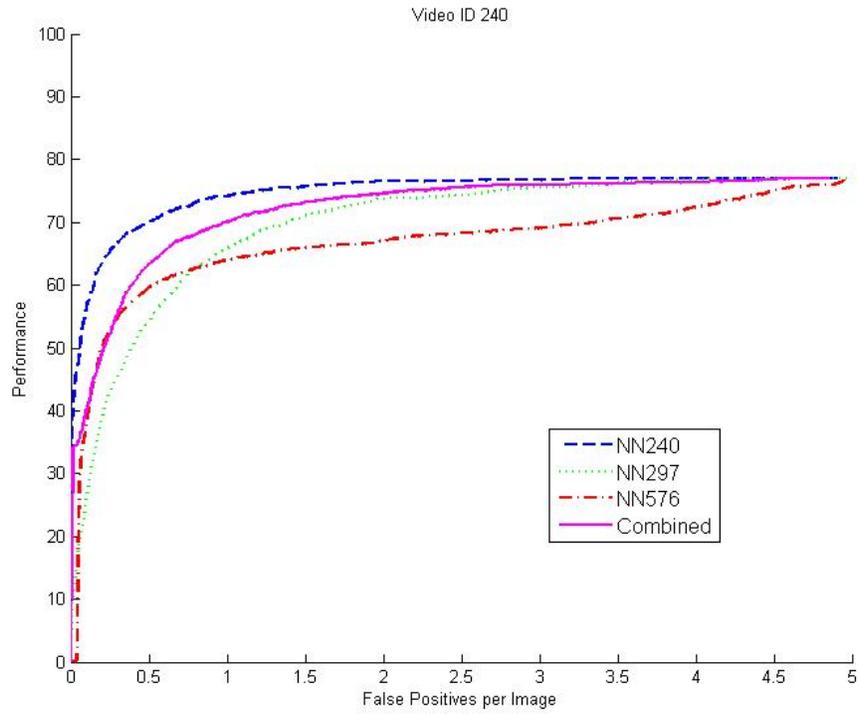
*Figure 4. Detection rate in novel videos versus training videos. Category 1 represents detection rate on three new videos, while category 2 shows detection rate on the three videos used to generate the filter. The similarity suggests that the filter is robust across videos.*

## 4.2. FROC Curve

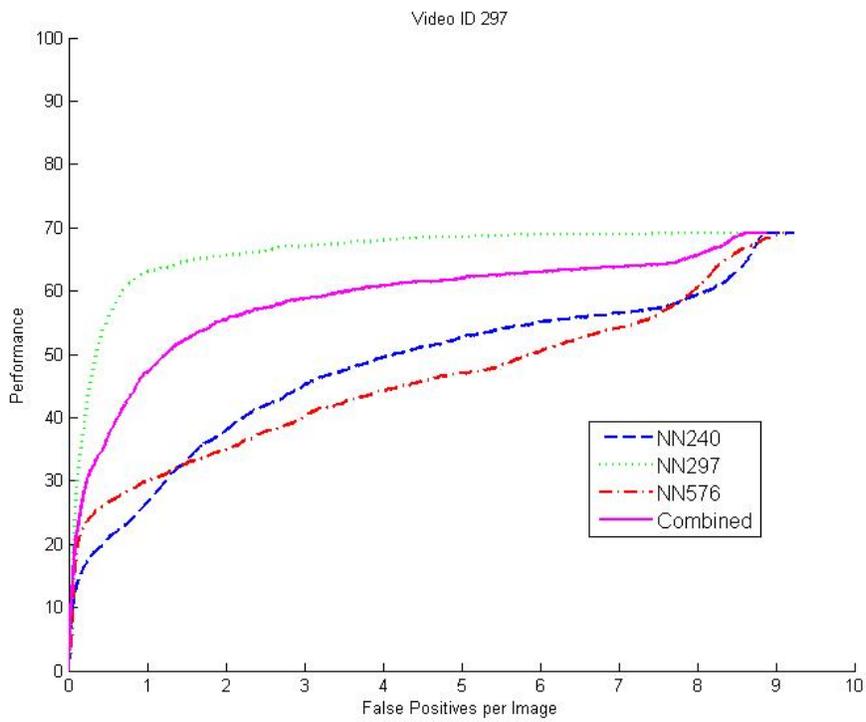
The primary indicator of system performance used was the Free-Response Receiver Operating Characteristic (FROC) curve. The Y axis represents the accuracy of the positive identification, the X axis is the number of false positives per image. This curve illustrates the trade-off between accuracy and false positive rate. A reliable system aims to converge to or approach its maximum detection rate while staying at or under one or two false positives per image.

We used three sets of sonar video images (#240, #297, and #576) to train and test the ATR system. For the different training schemes for neural networks, it was to be expected that the neural network would perform best when tested with the same input images it was optimized for, as shown in the FROC curves in Figure 5. However, generally these networks did not perform well for other video images. The third video, ID number 297, was particularly difficult because the target shapes and intensities were extremely different from the other two videos. In order to increase detection rate in this particular video, detection in the other two videos (576 and 240) was decreased.

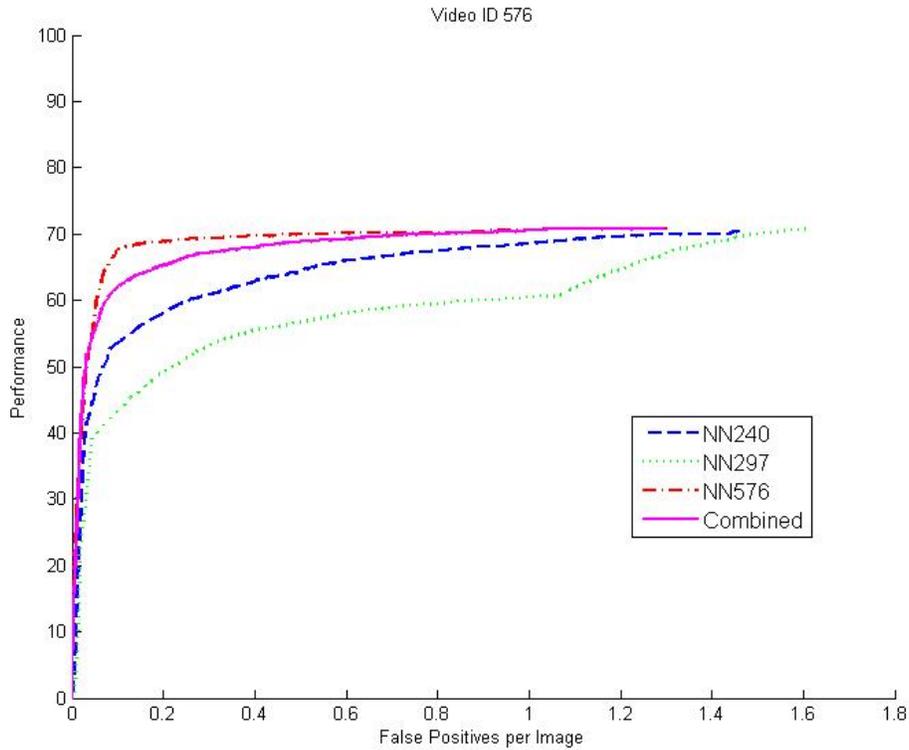
The mixed neural network performed fairly well across all three videos. Although it did not reach the accuracy and false positive count of any optimized neural network, it also did not have a significant performance drop when tested with any other video. Figure 5 compares performance for each optimized neural network and the combined neural network.



(a)



(b)



(c)

Figure 5. FROC Curve results for three different videos: (a) #240, (b) #297, and (c) #576. Each video was run with four different neural networks. As expected the optimized neural network performs best for its own corresponding video, reaching the upper limit of its detection rate while keeping the false positive count to a minimum. The upper limit of performance is determined by the parameters of the filters. The combined neural network also performs well across the three videos, suggesting that a mixed training set makes the network more robust

The combined neural network was also tested on novel videos, which did not contain any frames used to train the combined network. While the network did not perform as well as its optimized counterpart as expected, maximum detection rate was reached around two false positives per image. Figure 6 shows the performance of the combined neural network against the optimized network of two novel videos.

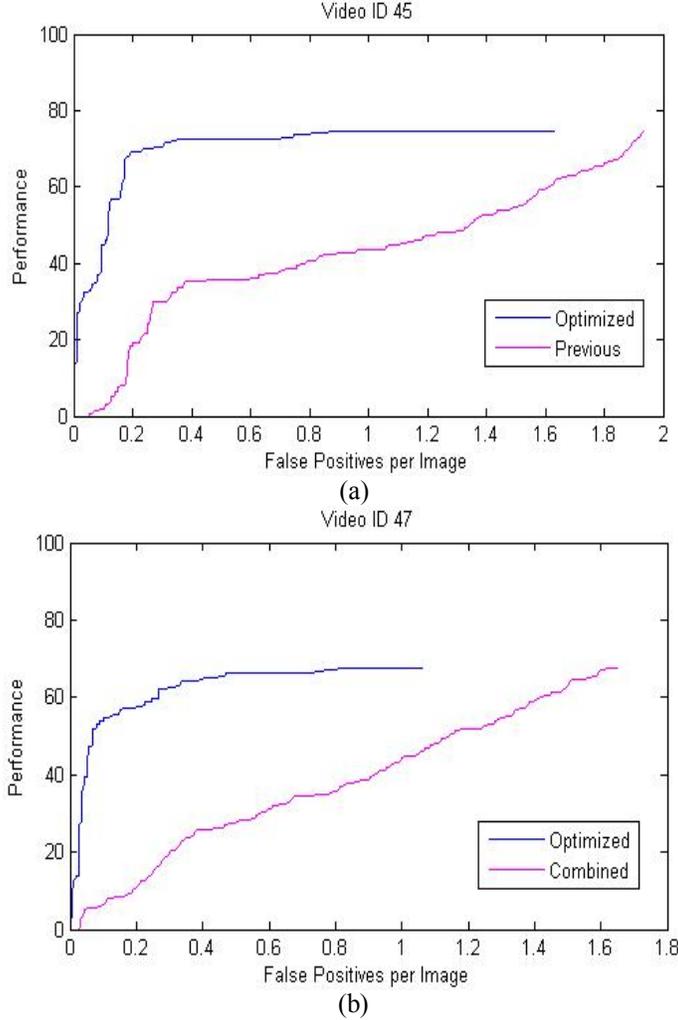


Figure 6. FROC Curve results for two novel videos (a) ID 45, and (b) ID 47. A neural network was optimized for a short segment of several new videos, and compared with the performance of the combined neural network previously described. While the combined neural network does not converge as quickly as the optimized networks, it still reaches its maximum detection rate under two false positives.

## V. DISCUSSION & CONCLUSIONS

Our results support the intuition that it is advantageous to explore different mixed training sets, in order to make an optimal filter that is robust across different environmental conditions. Using the composite Mexican-Hat wavelet filter, we were able to achieve between 60% and 80% detection across several sample videos, including those not used for training. With even a 60% detection rate, the target would be correctly identified in over half of the frames of the video, and many missing frames could be recovered by interpolating the detected target positions in the previous frames, making the task of monitoring sonar images less taxing on an operator. Further optimizing the wavelet filter parameters could result in a better detection rate.

While detection rate was satisfactory for the task of identifying targets in video, neural network responses were not as robust. There are several factors that could explain the limited performance of the "optimized" neural networks. An analysis of the extracted features could be done to ensure that each feature meaningfully contributes to ROI classification. Also, the number of training images could have resulted in a case of overfitting. Because neural

networks rely on an associative learning algorithm, it is possible that too many training examples would decrease the network's flexibility in classifying novel feature values. However, due to the discrepancy in ROI appearance across videos, it does not seem likely that even a not overtrained neural network would be able to make such a generalization.

However, the origins of these videos were not specified, and it was unclear as to why the targets varied so dramatically in appearance across videos. It is possible that these videos were taken by different devices. If so, it would be possible to optimize an ATR system for each device, as was done for each video. While such a system would not be able to perform across different devices, it could be calibrated to perform highly for a specific device. In this case, the optimized approach would be desirable.

Finally, the decisions on filter parameters, features extracted, training sizes and training set were more or less arbitrary. As previously stated, the filter used to detect ROIs is essential to the maximum detection rate of ATR systems using this framework. Future work could be done to further optimize these filters in order to boost the detection rate of this system. The training process and input features for the false positive eliminator could also be revisited and optimized, to balance the robustness of the classifier with the problem of overfitting.

## ACKNOWLEDGMENTS

This research was carried out at the Jet Propulsion Laboratory (JPL), California Institute of Technology under a contract with the National Aeronautics and Space Administration (NASA), and was under the sponsorship of the NASA Undergraduate Student Research Project (USRP) program through JPL.

## REFERENCES

1. Lu, T., Hughlett, C., Zhou, H., Chao, T., and Hanan, J., "Neural network post-processing of grayscale optical correlator," *SPIE Conference* **5908**, (2005).
2. Johnson, O., Edens, W., Lu, T., and Chao, T., "Optimization of OT-MACH Filter Generation for Target Recognition," *SPIE Conference* **7340**, (2009).
3. Zhou, H., Hughlett, C., Hanan, J., and Chao, T., "On the Development of Filter Management Module for Grayscale Optical Correlator," *SPIE Conference* **5437**, 87–94 (2004).
4. Ye, D., Edens, W., Lu, T., and Chao, T., "Neural Network Target Identification System for False Alarm Reduction," *SPIE Conference* **7340**, (2009).
5. Tan, L., Ma, J., Wang, Q., and Ran, Q., "Filtering Theory and Application of Wavelet Optics at the Spatial-Frequency Domain," *Appl. Opt.* **40**, 257-260 (2001)
6. Hinton, G.E. and Salakhutdinov, R. R., "Reducing the Dimensionality of Data with Neural Networks," *Science* **313**, 504 (2006).