Iterative Wavelet Transformation and Signal Discrimination for HRR Radar Target Recognition

Dale E. Nelson, *Member, IEEE*, Janusz A. Starzyk, *Senior Member, IEEE*,
D. David Ensley, *Member, IEEE*

Abstract —This paper explores the use of wavelets to improve the selection of discriminant features in the target recognition problem using High Range Resolution (HRR) radar signals in an air to air scenario. We show that there is statistically no difference among four different wavelet families in extracting discriminatory features. Since similar results can be obtained from any of the four wavelet families and wavelets within the families, the simplest wavelet (Haar) should be used. We further show that a simple box classifier can be constructed from the extracted features and that any feature that classifies four or less training signals can be removed from the classifier without a statistically significant difference in the classifier performance. We use the box classifier to select the 128 most salient pseudo range bins and then apply the wavelet transform to this reduced set of bins. We show that by iteratively applying this approach, classifier performance is improved. We call this the **iterative wavelet transform**. The number of times the feature reduction and transformation can be performed while producing improved classifier performance is small and the transformed features are shown to quickly cause the performance to approach an asymptote.

I. INTRODUCTION

Most of the work in HRR target recognition has been done by or sponsored by the military. The approaches taken by various researchers as summarized by [Mitchell 1997] appear to ignore the benefits that can be gained by proper transformations of the input signal. The wavelet transform [Antonini 1992][Mallat 1992][Mallat 1991] is a new tool which has been used in image compression, edge detection, image classification, and more recently in target recognition. When wavelet transforms are used for image compression the most important goal is to minimize the loss of information. In Automatic Target Recognition (ATR) the most important objective is to separate the various target classes [Szu 1996]. Some researchers have explored the use of wavelets to provide a richer feature space [Stirman 1991][Stirman 1995][Devaney 1997][Szu 1996]. However there is little evidence of widespread use of this technique. Mitchell himself explored the use of a sixth order Daubechies transformation but he limited the analysis to an autoregressive approach to remove low information data from the signature.

Famili [Famili 1997] found that preprocessing the data allows easier subsequent feature extraction and increased resolution. In the past, engineers have used transforms such as the Fourier transform to transform the signal from a time base to a frequency base [Misiti 1996]. Although this is useful for some applications, target recognition of HRR signals improved only a little under this transform. The reason for this lies in the fact that the Fourier Transform tells us that a feature occurs somewhere in the signal, but not where. Wavelets bring a new tool to HRR signal classification. The benefits of using wavelets [Strang 1996] are that the new transforms are local; i.e., the event is connected to the time when it occurs. Researchers who have used wavelets for target recognition (especially for HRR) have found that the original feature space can be augmented by the wavelet coefficients and will yield a smaller set of more robust features in the final classifier [Stirman 1995][Devaney 1994][Etemad 1998]. In addition to computational savings [Devaney 1997], investigators have also found that wavelet methods can improve the probability of correct classification (Pcc) [Stirman 1991], [Stirman 1995].

However, even with improvement in Pcc there can be a bias of the wavelets toward one or two classes to the detriment of others [Stirman 1995].

In considering wavelets for ATR, serious consideration must be given to the selection of a wavelet family and a wavelet in the family. Lu [Lu 1996] explored this issue in the context of image coders. In his paper, Lu compared two wavelets, one from the Biorthogonal family and the other from the Daubechies family. Using two different metrics, Lu observed a slight advantage of the biorthogonal versus the Daubechies. In this paper, using the criterion of improving the probability of correct classification , we show that there is no statistical advantage of one family (out of four) over any other family. Any difference in performance that can be observed in a particular application is due to the statistical nature of normal variations in the data. Stirman, using wavelets for ATR, explored the use of different wavelets from the Daubechies family, and found that the results were similar among the three wavelets [Stirman 1995]. In this paper we show that there is no statistical advantage of one wavelet in a family over another in the same family, thus generalizing Stirman's observation.

Once the input data is transformed, the process of feature selection for the given type of classifier must begin. A very popular approach uses a quadratic classifier [Mitchell 1997]. The quadratic classifier uses statistics of the signal to be classified and compares them to the statistics of a template for the various target classes. This method is fraught with problems. In an actual HRR signal there will be some response in the first few range bins which is not from a target but is noise. The last range bins will have values associated with multi-bounce returns from internal structures and cavities such as engines. The quadratic classifier will try to match these unreliable portions of the return. This increases the error when matched against the templates and can result in either no classification or misclassification. In an effort to solve this problem, Mitchell [Mitchell 1997] uses an autoregressive filter to remove noise and then uses the filter to help in selecting the most valuable range bins for classification.

Other researchers have employed wavelets to assist in HRR target identification [Devaney 1994] [Etemad 1998]. Devaney's approach used a sequential decision process where the log likelihood ratios are computed at each scale in the discrete wavelet transform (DWT) and then hypothesis testing is applied at each scale to yield the target identification. Etemad used the multiscale DWT to reduce the dimensionality of the classification problem. He used the coefficients to build a set of basis functions which yield the largest class separability. These basis functions result in simple and efficient algorithms for classification. The work presented here differs from the efforts of these two researchers. We employ a classifier in this work, but the focus is not on the classifier but on a method to improve upon the DWT itself. Etemad and Devaney applied the multiscale DWT one time we apply it many times. After the first application of the DWT we down-select, using the box classifier, a number of coefficients equal to the orignial number of range bins in the signal. Doing this many times yields a new pseudo wavelet (**iterative wavelet transform**) constructed for the problem presented by the training data.

It was not the purpose of this paper to explore the development of a classifier. However, in order to have a means to test the usefulness of the data transforms, we must have a classifier to test the performance and determine which features to select for further transformation. In order to mitigate the problem of the quadratic classifier, we have chosen to use the simple generalized box classifier [Ichino 1990][Starzyk 1996][Nelson 1997] from which to evaluate the results. Our main objective was to determine which, if any, family of wavelets provided the best feature set for a classifier. A secondary objective was to determine if further wavelet transformations would produce even better classification results. This required the use of a method for down selecting features from which to perform further wavelet analysis. In this paper, using wavelet transformations, we will show:

1) wavelets are useful for generating features that improve classifier performance,
2) what family and which wavelet in the family is best,
3) how to mitigate or eliminate wavelet bias towards some target classes.

II. PROBLEM SPECIFICATION

A. Signal and its Transform

When constructing a classifier, the designer is often able to rely on intuition to select the best features to distinguish the target classes. This works when the sensor used is a "literal" sensor. That is, the sensor provides an output similar to what human senses provide and what the human brain has experience interpreting. When the sensor does not provide this kind of output automated means must be used to select the best features. This paper uses High Range Resolution (HRR) radar signals. A HRR signal is an n-dimensional vector $x = (a_1, a_2, \dots, a_n)$, where $a_i \in \{0, 1, \dots, 255\}$. The HRR radar provides a 1-D picture of what the sensor is looking at. HRR signals are particularly hard to use for target recognition, partly because the 3-D world is projected on to just one dimension. When this is done, there are many ambiguities created which must be resolved. A further complication is that when HRR data is plotted as signal strength vs. range bin, the resulting graph is almost impossible for a human to use for target recognition, mostly because it is a visual 1-D image we have no experience interpreting [Koch 1995]. A better representation would be to present the HRR signal as an audio signal (similar to sonar) because humans have experience interpreting or recognizing this kind of 1-D signal. Szu points out that the human auditory system uses wavelets [Szu 1996] which aid in the recognition process.

A further complication to target identification using HRR is that the signals change considerably with only a small change in azimuth and elevation. This is illustrated in Figure 1.

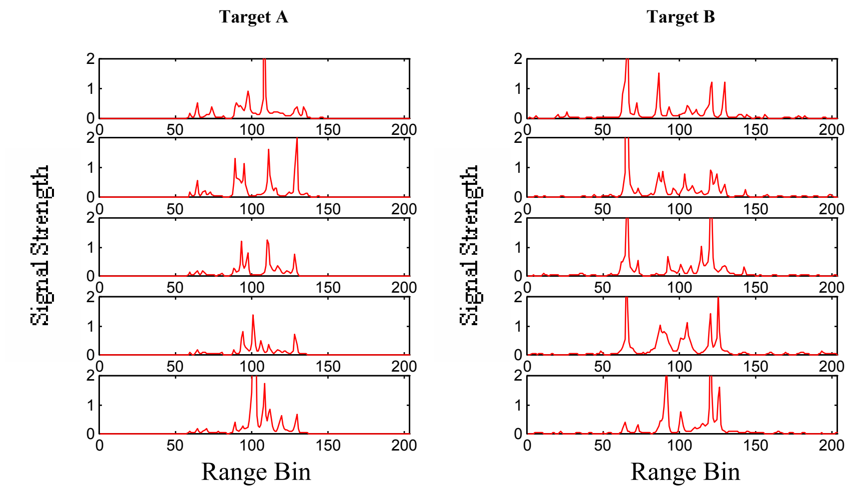

Figure 1. A comparison of two HRR target signals.

The signals shown in Figure 1 are from two different targets. The signals shown for each target were taken at 200 msec intervals. Their significant variations in a short time span illustrates how difficult it would be to construct a target identification system based on these signals.

Wavelet transforms have been found useful in a variety of applications. This is because they provide the analyst with an approximation of the signal and a detail of the signal as well. This helps to identify small anomalies which might be useful. A complete description of Wavelet

Packet Analysis also known as multilevel wavelet analysis as used in this research may be found in [Misiti 1996] and [Strang 1996]. Graphs of the wavelets used are presented in [Misiti 1996].

Prior to selecting features for the target classifier, it is useful to preprocess the original signal. Any operation which increases our ability to separate the classes is desirable. In this paper, we base feature selection on transformations derived from wavelets. Training and test sets were constructed using each of the functions. The utility of each of these wavelets for enhancing the performance of a classifier was then analyzed. An example of the power of a wavelet transformation is illustrated in Figure 2 using the Haar wavelet transform on the original signal.

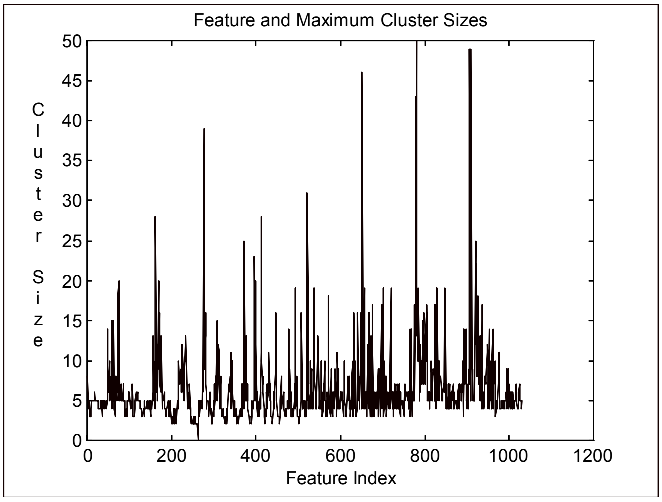

Figure 2. Maximum Cluster Sizes

In Figure 2, the original signal is contained in the first 128 feature index points. The coefficients of the Haar transform are contained in the remaining feature index points. The original signal features show that the largest number of signals in the training set that can be classified by a single feature is 20 out of a maximum of 60. Selecting a single feature from the wavelet coefficients, it is possible to classify 50 out of 60 signals. This is a significant improvement!

B. Training and Test Data Sets

The data used for training and testing is critical to the classifier's performance. Obviously, the best data to be used are measured data. However, these data are in very short supply, very expensive to obtain, and are difficult to ground truth. Ground truth is the recording of the radar parameters and the azimuth and elevation of the target aircraft. It is virtually impossible to gain measured data on enemy aircraft, since the enemy is unlikely to loan us an airplane for the measurements.

The data set used in this research consists of synthetic HRR returns on six targets. For each target there are 1071 range profiles consisting of 128 range bins. The value of each range bin is an

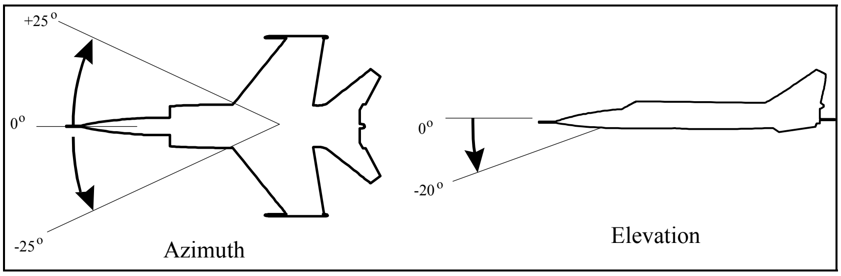

Figure 3. Azimuth and Elevation Ranges

integer between 0 and 255. The pose of the target is head-on with an azimuth range of $\pm 25^{o}$ and elevations of -20^{o} to 0^{o} in one degree increments as illustrated in Figure 3.

This data is divided into two sets, one for training and the other one for testing. The training set consists of 25% of the data and the test set 75% of the data (the remaining data), randomly selected. The small training set permits faster training, facilitating algorithm development and debugging. The training set was constructed by using a random number generator to select 25% of the azimuth and elevation angles and then by selecting signals from each target class with these angles. All remaining signals were placed into the test set.

We have illustrated that wavelets provide a powerful way of looking at the original signal so it makes sense to incorporate wavelet transforms and some statistical properties into the training and test set. The first step towards creating the training and test set is to normalize the original signal x_i using the l_2 norm yielding $\bar{x}_i$. We next calculate six values that characterize the data (2 norm, mean, infinity norm, standard deviation, 1-norm, and the Euclidean norm) $Q(\bar{x}_i) = (q_{1i}(\bar{x}_i), q_{2i}(\bar{x}_i), \ldots, q_{6i}(\bar{x}_i))$ where Q represents these vectors. Using similar notation, the wavelet transforms $W(\bar{x}_i)$ are constructed as described in [Misiti 1996]. The rows in the training set S are defined as the tuple $s_i = (Q(\bar{x}_i), \bar{x}_i, W(\bar{x}_i))$ where each $s_{ij} \in [0,1] \in \Re$. The training and test sets are conveniently represented as a matrix $S = \begin{bmatrix} \vdots \\ s_i \\ \vdots \end{bmatrix}$. We refer to each row of the training and test sets as a signal. The training set S consists of signals having 1030 pseudo range bins.

III. CLASSIFIER DESCRIPTION

The classifier used in this paper is a version of the generalized box classifier [Ichino 1990]. The training set produced as described previously is used to construct the classifier. The first step in constructing the classifier is to sort each column of S from the smallest value to largest value creating a new matrix $\bar{S}$. A matrix $\bar{M}$ is constructed with each element of $\bar{M}$ corresponding to the target type of each element of $\bar{S}$.

The algorithm determining a target classifier is as follows:

Let i denote the target class, and j the feature number. Set $i=j=1$.

Step 1. Search all columns of $\overline{M}$ to find the column with the largest contiguous cluster of the selected target class i. Let $\sigma(j)$ denote the column determined by this procedure (σ is a permutation of the columns of $\overline{S}$). Let $\overline{s}_{n,\sigma(j)}$ denote the minimum value in the contiguous cluster and let $\overline{s}_{k,\sigma(j)}$ denote the maximum value in the contiguous cluster. The indices n and k correspond to the row indices of $\overline{S}$ with the minimum and maximum values. All signals contained in this cluster are removed from further consideration.

Step 2. Define the j^{th} feature of target class i as the set $f_{ij} = (\overline{s}_{n,\sigma(j)}, \overline{s}_{k,\sigma(j)})$. Set $j=j+1$ and repeat this process (go to step 1) until there are no more training signals from target class i.

Step 3. Increment target class i and set $j=1$. Repeat this process (go to step 1) until all target classes are accounted for.

The elements of f_{ij} are called individual features. The feature set F is defined as the set of all f_{ij}. A transformed signal z is said to be classified as target class i when there exists a feature f_{ij} such that $z \in f_{ij}$.

A classifier is tested by classifying each of the transformed test signals, z. An n x n confusion matrix C is constructed to represent the results. To construct the confusion matrix, we first set $C = [0]$. Each test signal is classified and C is modified as follows. If the i^{th} test signal, known to be of class j is classified as the target type j then $C_{jj} = C_{jj} + 1$. If the i^{th} test signal known to be of class j is classified as target type k, then $C_{jk} = C_{jk} + 1$. In other words, the diagonal represents the correctly classified targets. The off-diagonal elements represent misclassification. This process continues until all transformed test signals are classified. In this paper we used equal numbers of signals to represent each target class for both training and test. Therefore, to obtain the final confusion matrix, each element of C is divided by the number of signals for a target class. It should be noted that some test signals might not be classified as any target type. Therefore, it is possible that the rows and columns of the confusion matrix will not sum to one. To evaluate the overall performance of the classifier the probability of correct classification, P_{cc} ,is calculated. P_{cc} is defined for *n* target classes as $P_{cc} = \frac{1}{n}\sum_{i=1}^{n} C_{ii}$.

IV. WAVELET SIGNAL DISCRIMINATION PROPERTIES

As observed in the previous discussion, a wavelet transform improves feature selection for target recognition. The natural question is to identify which wavelet improves target recognition the most. In this section we demonstrate that there is no single wavelet that outperforms all others in this task.

Conjecture 1*: No single wavelet transform has a statistically significant advantage over other wavelets in extracting features for the target classification.*

To verify conjecture 1, classifiers were constructed using training sets from all the wavelet families. Table 1 shows the results obtained upon testing the classifier built from the original signal and the associated wavelet transform. In addition, the mean and standard deviation of P_{CC} for the wavelet family are presented in Table 1. To compare if there is any significant difference among the families we use hypothesis testing of the means [Miller 1965]. The mean, μ , and the standard deviation, σ , of the population are calculated using:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} P_{CC_i} \qquad \sigma = \sqrt{\frac{n\sum_{i=1}^{n} P_{CC_i}^2 - \left(\sum_{i=1}^{n} P_{CC_i}\right)^2}{n(n-1)}}$$

When the mean and standard deviation are computed from samples, μ is replaced by $\bar{x}$ and σ is replaced by s respectively. We are testing the hypothesis; H_0: $\mu_1 = \mu_2$ against the alternative hypothesis; H_1: $\mu_1 \neq \mu_2$. We compute the test statistic as follows:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

We will reject H_0 if $|Z| > 1.96$ (1.96 is for a two-tailed test where the results are significant at a level of .05). The results of this hypothesis testing are presented in Table 2.

Wavelet Name	P_{cc}	Wavelet Name	P_{cc}	Wavelet Name	P_{cc}	Wavelet Name	P_{cc}
Bior1.3	0.72488	Haar	0.77130	Coif1	0.76115	Sym2	0.79153
Bior1.5	0.78045	Db2	0.79576	Coif2	0.78231	Sym3	0.75886
Bior2.2	0.75760	Db3	0.75886	Coif3	0.77133	Sym4	0.77458
Bior2.4	0.78150	Db4	0.79160	Coif4	0.78770	Sym5	0.75800
Bior2.6	0.77400	Db5	0.77567	Coif5	0.76943	Sym6	0.76345
Bior2.8	0.78600	Db6	0.78120			Sym7	0.76591
Bior3.1	0.70550	Db7	0.77460			Sym8	0.78275
Bior3.3	0.77030	Db8	0.76760				
Bior3.5	0.78020	Db9	0.79410				
Bior3.7	0.79410	Db10	0.77630				
Bior3.9	0.79290	Db11	0.75598				
Bior4.4	0.72990	Db12	0.76300				
Bior5.5	0.74150						
Bior6.8	0.73010						
Mean	0.76064		0.77550		0.77438		0.77073
Standard Deviation	0.02890		0.01329		0.01060		0.01272

Table 1. Performance of Wavelets

From the analysis presented we must accept the null hypothesis, that there is no difference in the mean values. This means that there is no statistically significant difference in the performance of the classifiers when different families of wavelets are used to transform the input data. It would be best (from a computational standpoint) to use the simplest form of a wavelet possible. Since there is no difference among the families, the question arises is there any significant difference within each family? By examining the size of the mean and the size of the standard deviation, we see that there is no significant difference among the wavelets within the families. It is safe to conclude that classifier performance would be the same no matter which wavelet we choose. Therefore, it benefits us to use the simplest form of wavelet possible, the Haar (Db1) wavelet.

Wavelet Name	Wavelet Name	$\|Z\|$	Accept or Reject H0
Biorthogonal	Daubechies	1.723060	Accept
Biorthogonal	Coiflet	1.516130	Accept
Biorthogonal	Symlet	1.109050	Accept
Daubechies	Coiflet	0.183654	Accept
Daubechies	Symlet	0.775505	Accept
Coiflet	Symlet	0.540601	Accept

Table 2. Wavelet Family Hypothesis Test

Normally this type of analysis is limited to large samples where the standard deviations of the samples are known. A t-test was also performed which gave the same results. This indicates that the small number of samples did not give us a false acceptance of H_0.

V. FEATURE SIZE DEPENDENCE

When constructing the classifier, there are times when the classifier is selecting features to classify just a few training signals. However, each new feature increases the dimensionality of the statistical feature space in which the signal classification is performed. The increase in space dimensionality reduces the accuracy of the statistical representation of the training data. As a result, it is possible that when the classifier is choosing a feature to classify a few signals, the classifier performance may decrease on the test set. Based on our research we make the following observation:

***Conjecture 2**: Features which classify a small number of training signals do not significantly improve the classifier performance.*

We theorized that this would be true due to the fact that these training signals do not carry much information about the problem to begin with. In statistical learning methods capturing this minute information is very difficult and prone to errors due to the statistically insignificant size of the training data supporting these features. Botha [Botha 1996], constructing a peak based classifier, also found the utility of controlling the number of features used. An analysis was performed to determine what size (number of signals that were classified) feature could be safely ignored. Elimination of features classifying a small number of signals allows for much faster training of the classifier and better generalization. Tables 3-6 show the results of the analysis. Each column in the tables show the probability of correct classification with the minimum feature size indicated in the first row.

Using the Z test, as before, we found that eliminating features that classify only one signal produces no statistically significant difference in the classifiers. However, it was suspected that because of the small sample size it might be necessary to perform a t-test. A t-test is used when either or both of the populations are small and the population variances are unknown. However, it must be assumed that the standard deviations of both populations are the same. The t statistic is:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

Name	Test Pcc	>1 Pcc	>2 Pcc	>3 Pcc	>4 Pcc	>5 Pcc	>6 Pcc	>7 Pcc
Db1	0.77130	0.76903	0.75369	0.71595	0.67470	0.64463	0.61022	0.58327
Db2	0.79576	0.79474	0.77753	0.75514	0.71347	0.67884	0.65625	0.62203
Db3	0.75886	0.75762	0.73772	0.70496	0.67241	0.64131	0.61124	0.58242
Db4	0.79160	0.79060	0.77318	0.73483	0.70124	0.67408	0.65355	0.62640
Db5	0.77567	0.77443	0.75888	0.72965	0.69026	0.65667	0.63760	0.61002
Db6	0.78120	0.78065	0.76759	0.73462	0.69689	0.67035	0.62702	0.60214
Db7	0.77460	0.77381	0.75784	0.72695	0.69378	0.66351	0.63739	0.61624
Db8	0.76760	0.76655	0.74748	0.71327	0.68113	0.63822	0.61313	0.58058
Db9	0.79410	0.79247	0.77712	0.74913	0.70974	0.68175	0.65666	0.62411
Db10	0.77630	0.77526	0.75764	0.72073	0.68549	0.65522	0.62993	0.61210
Db11	0.75598	0.75452	0.73192	0.69771	0.65687	0.62743	0.58783	0.56087
Db12	0.76300	0.76157	0.74456	0.71409	0.66391	0.64318	0.61478	0.56875
Mean	0.77550	0.77427	0.75710	0.72475	0.68666	0.65627	0.62797	0.59908
Std Dev	0.01329	0.01340	0.01496	0.01708	0.01762	0.01762	0.02146	0.02285
Z		0.22515	3.18521	8.12297	13.94261	18.71136	20.24809	23.11965
Alpha=.05	1.96	NOT Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
Alpha=.01	2.58	NOT Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
t s1=s2,normal		0.02601	0.37924	1.00870	1.75033	2.34906	2.74159	3.21468
t .05/22	2.074	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	Sig	Sig	Sig
t .01/22	2.819	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	Sig
F		1.00785	1.12539	1.28459	1.32554	1.32564	1.61402	1.71875
Alpha=.05 2.85 (11,12)		Accept	Accept	Accept	Accept	Accept	Accept	Accept
Alpha=.01 4.54 (11,10)		Accept	Accept	Accept	Accept	Accept	Accept	Accept

Table 3. Significance of Eliminating Features for Daubechies Wavelets

The null hypothesis is rejected if the value of t is greater than 2.074 or 2.85 depending on the level of significance chosen. As seen in Tables 3-6, we can safely eliminate all features which classify less than five training signals.

The t-test depends on the standard deviations being equal so we need to test for this also. To accomplish this we use the statistic

$$F = \frac{s_1^2}{s_2^2}$$

Name	Test Pcc	>1 Pcc	>2 Pcc	>3 Pcc	>4 Pcc	>5 Pcc	>6 Pcc	>7 Pcc
Sym2	0.79153	0.79039	0.77297	0.74601	0.70704	0.68008	0.64691	0.61726
Sym3	0.75886	0.75762	0.73772	0.70496	0.67241	0.64131	0.61124	0.58242
Sym4	0.77458	0.77277	0.75763	0.72985	0.69129	0.65314	0.60380	0.58390
Sym5	0.75800	0.75701	0.73855	0.71056	0.67428	0.64194	0.62079	0.58409
Sym6	0.76345	0.76261	0.74665	0.71949	0.69461	0.65853	0.61811	0.59883
Sym7	0.76591	0.76509	0.74830	0.71927	0.67822	0.63571	0.60585	0.58118
Sym8	0.78275	0.78148	0.76801	0.73919	0.70975	0.67762	0.63906	0.61418
Mean	0.77073	0.76957	0.75283	0.72419	0.68966	0.65548	0.62082	0.59455
Std Dev	0.01272	0.01261	0.01384	0.01492	0.01526	0.01773	0.01646	0.01564
Z		0.22402	3.29655	8.22262	14.13325	18.29834	24.95777	30.26778
Alpha=.05	1.96	NOT Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
Alpha=.01	2.58	NOT Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
t s1=s2,normal		0.02521	0.38028	0.96966	1.67871	2.28800	3.03959	3.62363
t .05/22	2.074	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	Sig.	Sig.	Sig.
t .01/22	2.819	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	Sig.	Sig.
F		1.00860	1.08822	1.17248	1.19974	1.39325	1.29408	1.22955
Alpha=.05 2.85 (11,12)		Accept	Accept	Accept	Accept	Accept	Accept	Accept
Alpha=.01 4.54 (11,10)		Accept	Accept	Accept	Accept	Accept	Accept	Accept

Table 4. Significance of Eliminating Features for Symlets Wavelets

Name	Test Pcc	>1 Pcc	>2 Pcc	>3 Pcc	>4 Pcc	>5 Pcc	>6 Pcc	>7 Pcc
Coif1	0.76115	0.75991	0.74104	0.71139	0.67635	0.64028	0.60566	0.57041
Coif2	0.78231	0.78147	0.76551	0.72819	0.69896	0.67346	0.63241	0.60877
Coif3	0.77133	0.77008	0.75225	0.72675	0.69461	0.65937	0.62371	0.59240
Coif4	0.78770	0.78625	0.77008	0.73835	0.70456	0.68030	0.66371	0.62743
Coif5	0.76943	0.76862	0.74954	0.71699	0.68859	0.65541	0.63386	0.60753
Mean	0.77438	0.77327	0.75568	0.72433	0.69261	0.66176	0.63187	0.60131
Std Dev	0.01060	0.01056	0.01191	0.01047	0.01081	0.01572	0.02105	0.02128
Z		0.25889	4.06272	11.63783	18.70997	20.58158	20.95222	25.22373
Alpha=.05	1.96	NOT Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
Alpha=.01	2.58	NOT Sig.	Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
t s1=s2,normal		0.02663	0.43176	1.19448	1.93590	2.40503	2.77536	3.35826
t .05/22	2.074	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	Sig.	Sig.	Sig.
t .01/22	2.819	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	Sig.
F		1.00351	1.12434	1.01191	1.02042	1.48319	1.98604	2.00790
Alpha=.05 2.85 (11,12)		Accept	Accept	Accept	Accept	Accept	Accept	Accept
Alpha=.01 4.54 (11,10)		Accept	Accept	Accept	Accept	Accept	Accept	Accept

Table 5. Significance of Eliminating Features for Coiflets Wavelets

Name	Test Pcc	>1 Pcc	>2 Pcc	>3 Pcc	>4 Pcc	>5 Pcc	>6 Pcc	>7 Pcc
Bior1.3	0.72488	0.72466	0.71160	0.68548	0.64961	0.61934	0.58160	0.55175
Bior1.5	0.78045	0.77899	0.76178	0.72985	0.68486	0.65355	0.62349	0.60048
Bior2.2	0.75760	0.75639	0.73255	0.70332	0.65377	0.62744	0.59385	0.56337
Bior2.4	0.78150	0.77961	0.76365	0.73047	0.69253	0.66288	0.63697	0.59634
Bior2.6	0.77400	0.77194	0.75577	0.72591	0.69584	0.65874	0.64091	0.60566
Bior2.8	0.78600	0.78397	0.76551	0.73006	0.69482	0.66248	0.63345	0.59551
Bior3.1	0.70550	0.70455	0.68008	0.64484	0.60440	0.56750	0.53391	0.49556
Bior3.3	0.77030	0.76944	0.75182	0.71823	0.68195	0.64381	0.61188	0.58534
Bior3.5	0.78020	0.77899	0.76406	0.73462	0.69461	0.66807	0.63883	0.62515
Bior3.7	0.79410	0.79226	0.77567	0.75224	0.71783	0.69316	0.65936	0.62120
Bior3.9	0.79290	0.79247	0.77837	0.74789	0.70891	0.68361	0.65189	0.62349
Bior4.4	0.72990	0.72944	0.71783	0.69067	0.66848	0.62744	0.60753	0.58473
Bior5.5	0.74150	0.74146	0.73192	0.70869	0.68755	0.65167	0.61622	0.59736
Bior6.8	0.73010	0.73006	0.72032	0.69565	0.66538	0.63553	0.61936	0.58680
Mean	0.76064	0.75959	0.74364	0.71414	0.67861	0.64680	0.61780	0.58805
Std Dev	0.02890	0.02844	0.02857	0.02855	0.02887	0.03113	0.03232	0.03386
Z		0.08970	1.44924	3.96527	6.95551	9.28304	11.41263	13.42897
Alpha=.05	1.96	NOT Sig.	NOT Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
Alpha=.01	2.58	NOT Sig.	NOT Sig.	Sig.	Sig.	Sig.	Sig.	Sig.
t s1=s2,normal		0.01518	0.24566	0.67206	1.18217	1.60943	1.99979	2.38636
t .05/22	2.074	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	Sig.
t .01/22	2.819	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.	NOT Sig.
F		1.01615	1.01172	1.01229	1.00088	1.07728	1.11825	1.171765
Alpha=.05 2.85 (11,12)		Accept	Accept	Accept	Accept	Accept	Accept	Accept
Alpha=.01 4.54 (11,10)		Accept	Accept	Accept	Accept	Accept	Accept	Accept

Table 6. Significance of Eliminating Features for Biorthogonal Wavelets

If this value is less than 2.85 we must accept the null hypothesis that the variances are equal. As seen from Table 2, we accept the null hypothesis for all the tests (the standard deviations are equal), i.e., we can accept the results of the t-test.

The results in Tables 3-6 show no matter what family of wavelet is used there is no statistical performance difference among the classifiers when features which classify four signals or less are removed from the classifier. Removing these features significantly reduces the time required to create the classifier.

VI. ITERATIVE WAVELET TRANSFORM

If the original signals are transformed and then 128 of the most informative pseudo range bins selected for further transformation, a new linear transformation of the input data is created [Starzyk 1998]. This process can be repeated many times and is called the **iterative wavelet transform.** This is similar to the basic assumption of genetic programming where new generations of features related to the most successful features from prior generations may show better qualities than their parents. For any learning process based on a fixed set of data the increase in information represented by learning features is getting smaller as progress towards the optimum is made. At some point there will be no increase, at which point the learning process must stop.

Conjecture 3*: By iteratively selecting the most informative pseudo range bins and transforming them, the informative value of the range bins in general may increase yielding a better classifier.*

An experiment was performed to verify this conjecture. The original 128 range bin signal was transformed (using the Haar wavelet) as previously discussed creating 1024 pseudo range bins. A box classifier was constructed. The range bins used as features were chosen for further transformation. If there are more than 128 pseudo range bin features, then the features which classify the most training signals are selected. If there are fewer than 128 features, then additional pseudo range bins are selected from the middle of the pseudo signal. These 128 range bins were wavelet transformed, a classifier was constructed and tested. This procedure was repeated twelve times and the results are presented in Table 7 and Figure 4.

Iteration	P_{CC}	Target 1	Target 2	Target 3	Target 4	Target 5	Target 6
0	.7713	.9490	.6219	.8219	.6853	.8134	.7363
1	.81361	.9552	.6741	.8555	.6692	.8893	.8383
2	.84762	.9552	.7724	.9128	.7027	.9104	.8321
3	.86421	.9453	.7823	.9452	.7363	.9154	.8607
4	.87976	.9453	.7935	.9465	.7774	.9328	.8831
5	.88618	.9453	.8172	.9552	.7550	.9316	.9129
6	.88453	.9453	.8197	.9601	.7376	.9316	.9129
7	.89095	.9391	.8507	.9664	.7450	.9316	.9129
8	.89261	.9391	.8246	.9664	.7799	.9316	.9142
9	.88867	.9391	.8346	.9651	.7488	.9316	.9129
10	.89717	.9391	.8706	.9639	.7512	.9391	.9192
11	.89365	.9353	.8570	.9639	.7512	.9391	.9154
12	.90049	.9353	.8483	.9689	.7749	.9465	.9291

Table 7. Results of Iterative Application of Haar Transform

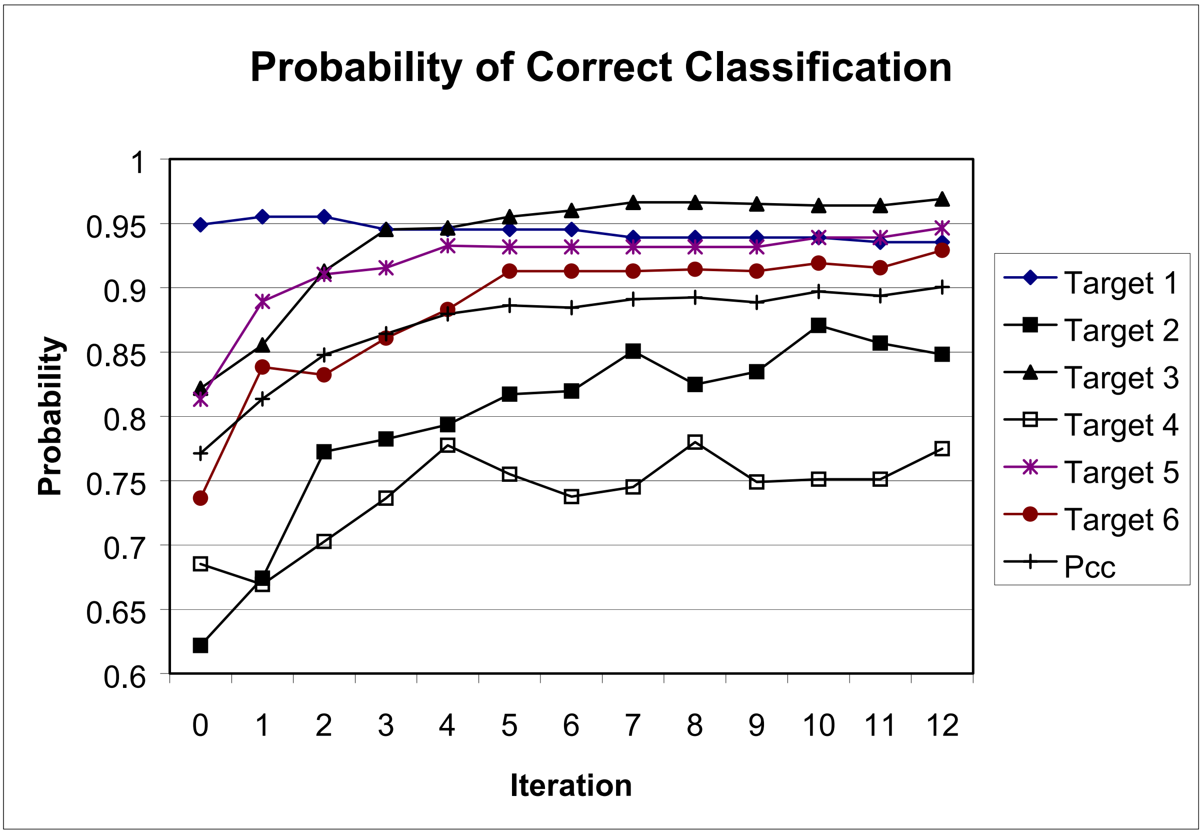

Figure 4. Classification Improvement with Iterated Wavelet Transforms

When using just one wavelet transform on the original signal Stirman showed an increase in Pcc of 6 percentage points [Stirman 1991] and 7.53 percentage points over the baseline classifier [Stirman 1995]. This difference, over the baseline classifier, may have resulted from changing the type of classifier or the use of wavelets. Stirman did not attribute the increase in performance to one or the other, neither did he analyze the significance of using the wavelet transform. In the results presented here, we find that using the same classifier, the improvement in Pcc after one application of the wavelet transform is 4.2 percentage points. This improvement is smaller than the one observed by Stirman, but we can attribute this difference to our use of a different classifier and wavelet (Haar).

The most important curve in Figure 4 is the one for Pcc which represents the performance of the classifier for all target classes at each iteration of the iterative wavelet transform. In this figure iterations 2-12 demonstrate the benefits of the iterative wavelet transform over the use of a single transform. This curve shows an increase in overall classifier performance from .7713 to .89717 by iteration 10. This represents an improvement of 12 percentage points. Furthermore, Target 2 improved by 25 percentage points and Target 6 by 18 percentage points. This is a significant improvement in performance over a single wavelet transform and confirms the benefit of using the iterative wavelet transform.

We questioned why there would be a decrease in performance on some of the targets such as seen on Target 2 between iterations 7 and 8. It is apparent that the iterative wavelet transforms yield an increasing performance in the entire classifier. Individual targets may sacrifice performance while overall performance increases. In general, the momentary decreases are recovered in later iterations. This may be a manifestation of the biasing problem reported by Stirman [Stirman 1995]. If so, by iterating the wavelet transform this problem appears to either be mitigated or eliminated. For our problem, the maximum advantage of iterating the wavelets happens at about ten iterations.

Figure 5. Feature Sizes

We theorized that performance would increase as fewer features are required and the features that are chosen classify more signals. This was confirmed by the experiment and is graphically illustrated in Figure 5. It is easily seen that the targets with the fewest features have the highest performance (Figure 4). The targets with the lowest performance require the most features. This happens when two targets are very similar and the same range bins contain the information required to segment the targets. In such case, an individual feature provides little differentiation between targets, and additional features are required. We also note that performance on a given target improves most significantly when the number of features required for classification decreases as in Target 5, Target 6, and Target 2. Similar observations were made by Fukunaga [Fukunaga 1990] and Etemad [Etemad 1998] which motivated our work to find a minimal number of features for target recognition.

VII. SUMMARY OF RESULTS

The most significant contribution of this paper is the **iterative wavelet transformation** that was used to enrich the feature space and improve classifier performance. Our conjectures were verified using statistical hypothesis testing on synthetic HRR data.. An information entropy approach for the down select of the pseudo range bins has shown similar improvement in classification performance.

We have shown that there is no statistically significant difference in performance of the classifier when different wavelets are chosen. This means that the simplest wavelet to implement will do as good a job as any other wavelet, at least for the HRR target recognition problem. The presented results show that (using a box classifier) features that classify fewer than five training signals can safely be ignored without producing a statistically significant change in the classifier's performance.

The application of the **iterative wavelet transformation** method used here to improve performance could potentially be used in classification of any 1-D signal such as found in echo cardiograms, seismology, sonar, and economic analysis.

REFERENCES

[Antonini 1992] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image Coding Using Wavelet Transform", *IEEE Transactions on Image Processing*, Vol 1, No. 2, 1992, 205-220.

[Botha 1996] E. C. Botha, E. Barnard, and C. J. Barnard, "Feature-based Classification of Aerospace Radar Targets Using Neural Networks", *Neural Networks*. Vol. 9, No. 1, 1996, 129-142.

[Devaney 1994] A. J. Devaney and B. Hisconmez, "Wavelet Signal Processing for Radar Target Identification a Scale Sequential Approach", *SPIE Wavelet Applications*, Vol 2242, 1994, 389-399.

[Devaney 1997] A. J. Devaney, R. Raghavan, H. Lev-Ari, E. Manolakos, and M. Kokar, *Automatic Target Detection and Recognition: A Wavelet Based Approach*, Northeastern University Technical Report AD-A329696, Defense Technical Information Center, 1997.

[Etemad 1998] K. Etemad and R. Chellapa, "Separability-Based Multiscale Basis Selection and Feature Extraction for Signal and Image Classification", *IEEE Transactions on Image Processing*, Vol. 7, No. 10, 1998, 1453-1465.

[Famili 1997] A. Famili, S. Wei-Min, R. Weber, and E. Simoudis, "Data Preprocessing and Intelligent Data Analysis", *Intelligent Data Analysis*, Vol. 1, No. 1, 1997.

[Fukunaga 1990] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, CA, 1990.

[Ichino 1990] M. Ichino, "Optimum Feature Selection for Decision Functions", *Systems and Computers in Japan*. Vol. 21, No. 1, 1990.

[Koch 1995] M. W. Koch., M. M. Moya, L. D. Hostetler, and R. D. Fogler, “Cueing, Feature Discovery, and One-Class Learning for Synthetic Aperture Radar Automatic Target Recognition”, *Neural Networks*. Vol. 8, No. 7/8, 1995, 1081-1102.

[Lu 1996] J. Lu, V. R. Algazi, and R. R. Estes, Jr., “Comaarative Study of Wavelet Image Coders”, *Optical Engineering*, Vol. 35, No. 9, 1996, 2605-2619

[Mallat 1991] S. Mallat, “Zero-Crossings of a Wavelet Transform”, *IEEE Transactions on Information Theory*, Vol. 37, No. 4, 1991, 1019-1033.

[Mallat 1992] S. Mallat and Sifen Zhong, “Characterization of Signals from Multiscale Edges”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 7, 1992, 710-732.

[Miller 1965] I. Miller and J. E. Fruend, *Probability and Statistics for Engineers*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1965

[Misiti 1996] M. Misiti, Y. Nisiti, G. Oppenheim, andJ. Poggi, *Wavelet Toolbox*, The MathWorks, Inc., Natick, MA, 1996.

[Mitchell 1997] R. A. Mitchell, “Hybrid Statistical Recognition Algorithm for Aircraft Identification”, Dissertation Proposal, University of Dayton, 1997.

[Nelson 1997] D. E. Nelson, and J. A. Starzyk, “Advanced Feature Selection Methodology for Automatic Target Recognition”, 29th Southeastern Symposium on System Theory, 1997.

[Starzyk 1996] J. A. Starzyk and D. E. Nelson, “Independent Classifiers in Ontogenic Neural Networks for ATR” Adaptive Distributive Parallel Computing Symposium, 1996.

[Starzyk 1998] J. A. Starzyk, *Evolutionary Feature Extraction for SAR Air to Ground Moving Target Recognition – Statistical Approach*, Sverdrup Technology, Inc. Advanced Systems Group Report, 1998.

[Stirman 1991] C. Stirman and A. Nachman, *Applications of Wavelets to Radar Data Processing*, Technical Report AD-A239297, Defense Technical Information Center, 1991.

[Stirman 1995] C. Stirman and A. Nachman, *Applications of Wavelets to Automatic Target Recognition*, Technical Report AD-A294854, Defense Technical Information Center, 1995.

[Strang 1996] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, 1996.

[Szu 1996] H. H. Szu, "Review of Wavelet Transforms for Pattern Recognition"*, Wavelet Applications III, Proceedings of SPIE*, Vol. 2762, 1996, 2-22.