A Novel Feature Vectors Construction Approach for Face Recognition

Paul Nicholl[1], Afandi Ahmad[2,3], and Abbes Amira[4]

[1] School of Electronics, Electrical Engineering and Computer Science
The Queens University, Belfast, Northern Ireland
P.Nicholl@qub@ac.uk
[2] Department of Electronic and Computer Engineering,
School of Engineering and Design, Brunel University, Uxbridge, United Kingdom
[3] Department of Computer Engineering,
Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia
afandia@uthm.edu.my
[4] Nanotechnology and Integrated Bio-Engineering Centre (NIBEC),
Faculty of Computing and Engineering, University of Ulster,
Jordanstown Campus, Co Antrim, Northern Ireland
A.Amira@ulster.ac.uk

Abstract. This paper discusses a novel feature vectors construction approach for face recognition using discrete wavelet transform (DWT). Four experiments have been carried out focusing on: DWT feature selection, DWT filter choice, features optimization by coefficients selection as well as feature threshold. In order to explore the most suitable method of feature extraction, different wavelet quadrant and scales have been studied. It then followed with an evaluation of different wavelet filter choices and their impact on recognition accuracy. An approach for face recognition based on coefficient selection for DWT is the presented and analyzed. Moreover, a study has been deployed to investigate ways of selecting the DWT coefficient threshold. The results obtained using the AT&T database have shown a significant achievement over existing DWT/PCA coefficient selection techniques and the approach presented increases recognition accuracy from 94% to 97% when the Coiflet 3 wavelet is used.

Keywords: Face recognition, discrete wavelet transform, coefficient selection, feature selection, feature optimization.

1 Introduction

In recent years, the demand for sophisticated security systems has risen significantly. Both commercial and governmental organisations require methods of protecting people and property. These often involve identifying people; to control access to resources or to detect individuals on a watch list. Solutions employing biometric techniques are being used widely to facilitate these needs [1].

M.L. Gavrilova et al. (Eds.): Trans. on Comput. Sci. XI, LNCS 6480, pp. 223–248, 2010.

A variety of biometric approaches have been investigated or adopted. For example, fingerprint recognition [2] has been used in crime solving for many years and is being increasingly installed in consumer devices, such as laptop computers. It is generally accurate and can be deployed with minimal cost. It does, however, suffers from the requirement of many biometric methods that an individual being identified must be compliant in the process – two fingerprints (or sets of fingerprints) must be supplied in order to create a match. Some biometrics are less intrusive – voice scans [3] can be taken without user compliance, although accuracy is currently low.

Face recognition has received a large amount of attention from researchers in recent years [4]. It has the potential to provide a robust biometric which, although unlikely to exceed the accuracy of techniques like iris or fingerprint scanning, could fulfill the needs of many scenarios. Much of the interest in face recognition has been prompted by humans' own remarkable ability to recognize faces [5]. This ability encompasses recognition of faces from thousands of known individuals, even in cases where there is partial occlusion of the face, poor illumination, or there has been a change in appearance. Automatic face recognition also requires less compliance by the individual being identified. A face image for matching can be taken without the individual posing or even knowing that the image is being captured.

A multitude of techniques have been applied to face recognition and they can be separated into two categories: geometric feature matching and template matching. Geometric feature matching involves segmenting the distinctive features of the face, for examples eyes, nose, mouth, etc., and extracting descriptive information about them such as their widths and heights. Ratios between these measures can then be stored for each person and compared with those from known individuals [6].

On the other hand, template matching is a holistic approach to face recognition. Each face is treated as a two-dimensional (2-D) array of intensity values, which is compared with other facial arrays. Techniques of this type include principal component analysis (PCA) [7], where the variance among a set of face images is represented by a number of eigenfaces. The face images, encoded as weight vectors of the eigenfaces, can be compared using a suitable distance measure [8]. In independent component analysis (ICA), faces are assumed to be linear mixtures of some unknown latent variables. The latent variables are assumed non-gaussian and mutually independent, and they are called the independent components of the observed data [9]. In neural network models (NNM), the system is supplied with a set of training images along with correct classification, thus allowing the neural network to ascertain a weighting system to determine which areas of an image are deemed most important [10].

Hybrid multiresolution approaches have received much attention in recent years. The discrete wavelet transform (DWT) [11] has been used along with a number of techniques, including PCA [12], ICA [13] and support vector machines (SVM) [14]. DWT is able to extract features that are localized in both space and frequency by convolving a bank of filters with an image at various locations. However, to date, no systematic examination has been performed which determines how to best employ DWT for face recognition. The effect of employing different filters and scales has not been examined.

This research study attempts to investigate these issues. Initially, experimentation is performed using the Haar [15] and biorthogonal 4.4 [16] wavelets, in order to determine the most appropriate wavelet quadrants and scales. The study is then widened to cover a range of wavelets, with filters examined from the Daubechies [17], symlet [17], Coiflet [17] and biorthogonal [18] families. Results are analysed in order to ascertain whether scales and wavelet filters can be intelligently chosen for face recognition applications. In addition, an approach for face recognition based on DWT coefficient selection is presented and analysed. This operates by attempting to optimize the feature vectors produces by DWT, thereby improving results. An added benefit of the process is that it can automatically segment the face image, eliminating the need to manually crop images and possibly removing useful information.

The remainder of this paper is organized as follows. Section 2 investigates which DWT features should be utilized for face recognition. Section 3 analyses which wavelet filters perform best in this domain. Section 4 addresses the optimization of feature vectors using coefficient selection. Section 5 investigates how to choose a threshold for coefficient selection. Section 6 provides concluding remarks.

2 DWT Feature Selection

2.1 Concepts

In order to assess whether DWT can enhance face recognition system performance, a study is performed which attempts to determine how to employ it for this purpose. A number of variables are assessed, including: quadrant – which DWT quadrant(s) should be used for feature extraction?; scale – which scale(s) should be used for feature extraction?; and filter – which wavelet filters produce the best results?. This section attempts to address the first two points and experiments are conducted on the AT&T database. Each experiment is performed on coefficients taken from a specific wavelet scale and quadrant. A high-level overview of the recognition approach adopted is given in Figure 1.

2.2 Experiments

The experiments start with system training. For this stage, each training image undergoes wavelet transformation to the x^{th} scale. DWT coefficients from the specified quadrant at the x^{th} scale undergo PCA, producing a set of principal components. The training images are then projected onto the set of principal components, producing a weight vector for each image, which represents the features for the image. Probe images are processed in a similar manner, with each image decomposed to the same scale and coefficients from the same quadrant

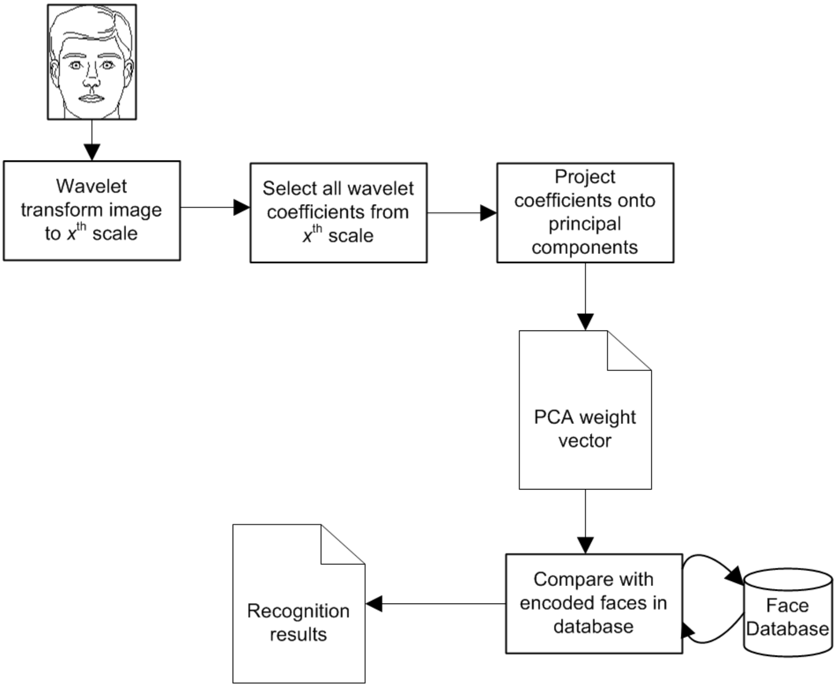

Fig. 1. Overview of recognition approach

extracted. This image is projected onto the same principal components and a weight vector is produced. This vector can be compared with those of training images. For the experiments described here, the Euclidean distance measure is employed.

For this study, five randomly-selected training images are used for each individual, with the remaining five being used as probe images. Other than minor re-scaling, the images undergo no preprocessing. As assessing wavelet filters is not the object of the experiments in this section, only two filters are adopted: Haar and biorthogonal 4.4. PCA is then employed, reducing the feature set further. The images used for system training also from the gallery set. As there are 200 training images, up to 200 principal components can be used to encode each face. However, when there are fewer than 200 features (pixels or wavelet coefficients) per training image (as is the case for higher wavelet scales), using more than 200 principal components is redundant. Encoded gallery images are compared with probe images using the Euclidean distance measure.

2.3 Results

This section presents recognition results for the Haar and biorthogonal 4.4 wavelets. The Haar wavelet has been chosen for its simplicity. The biorthogonal 4.4 wavelet has been chosen to represent a more sophisticated filter. Results are presented for the first five scales. It is worth mentioning that recognition accuracies for the sixth scale are significantly lower, due to the reduced number of coefficients at this scale.

The results for the Haar wavelet are shown in Figures 2 to 6. As can be seen from the graphs, the choice of scale does have a significant effect on recognition rate. For example, in Figure 2, the first and second scales perform better initially, however, the recognition rates fall sharply as more eigenvectors are used. This would suggest that, for HH quadrants, most of the useful information in these scales is encoded within the first 20 eigenvectors. The results for the third scale are more consistent, although they do not match the second scale peak recognition rate of 66.5%. The results for the fourth scale and fifth scale are lower, with the recognition rates leveling off at 64 and 16 eigenvectors respectively, due to the number of coefficients per quadrant at these scales being 16 and 64, respectively.

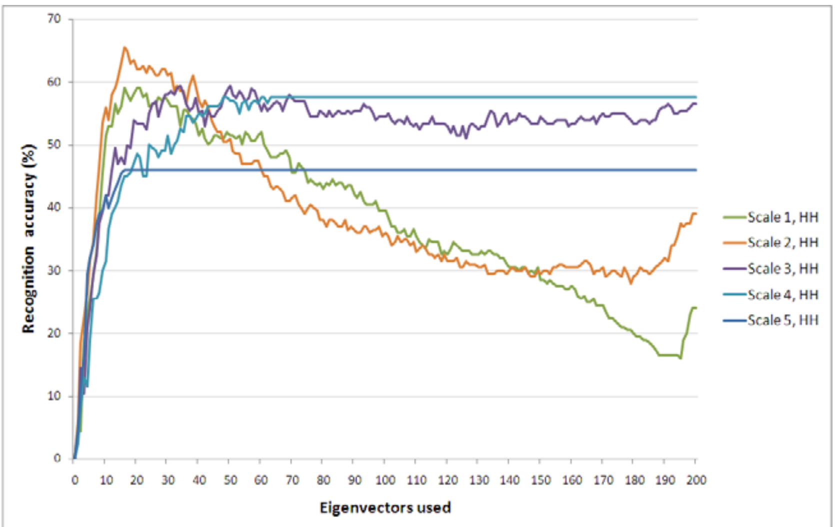

Fig. 2. Recognition results for Haar wavelet and HH quadrant

Figures 3 and 4 show results for LH and HL quadrants. In both cases, the fourth scale produces the best results, followed by the third. In addition, recognition rates for the first two scales significantly deteriorate as eigenvectors increase, whereas this does not occur for the third, fourth and fifth scales. Peak recognition rates are higher than for HH, with 74% of faces recognized correctly using the LH quadrant and 78% with the HL quadrant. Figure 5 provides results for the LL quadrant. By a significant margin, the best results with the Haar wavelet are achieved using this quadrant. The best scale for LL is the third, producing recognition rates up to 95%. Scales 2, 1 and 4 produce similar performance, with maximum recognition rates of 94%, 93% and 93%,

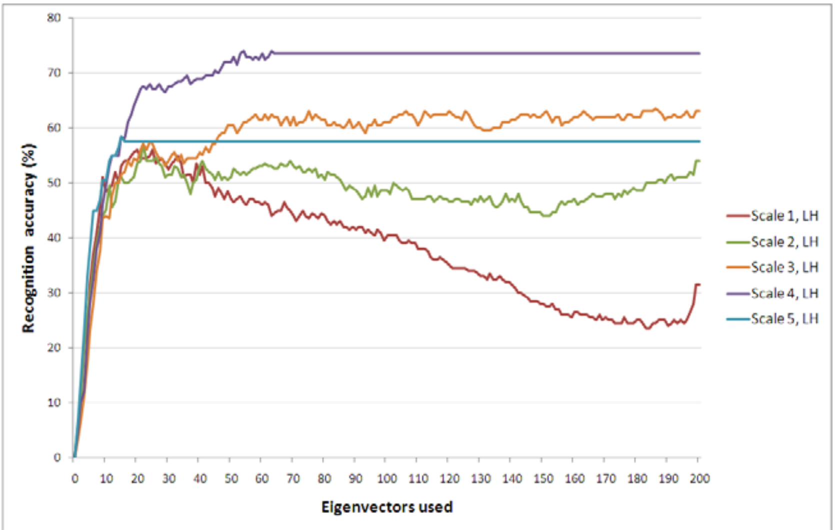

Fig. 3. Recognition results for Haar wavelet and LH quadrant

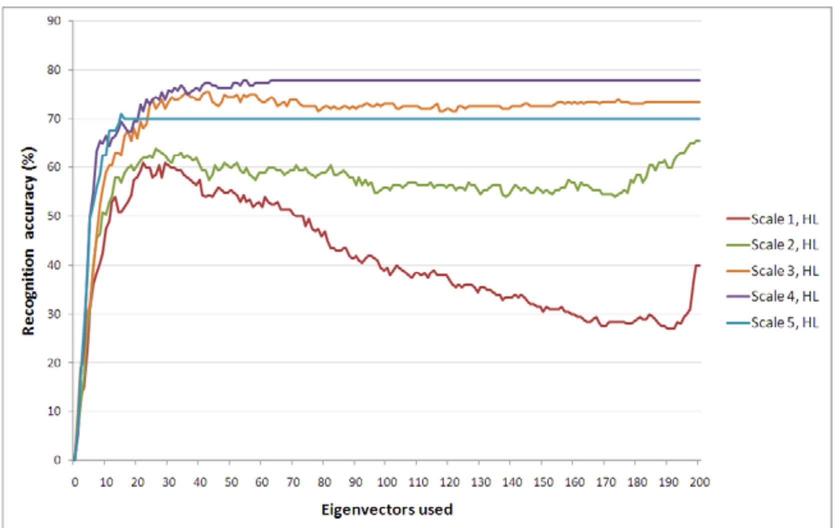

Fig. 4. Recognition results for Haar wavelet and HL quadrant

respectively. Scale 5 matches up to 87.5% of faces correctly. Figure 6 illustrates the extent to which the LL quadrant outperforms other quadrants.

Figures 7 and 8 show performance for the biorthogonal 4.4 wavelet. For coefficients from the LL quadrant, recognition rates are similar to those for Haar: scales 2 and 3 correctly recognize up to 94.5% of faces, with scales 1 and 4 recognizing 94% and 93%, respectively. From Figure 8, it can be seen that

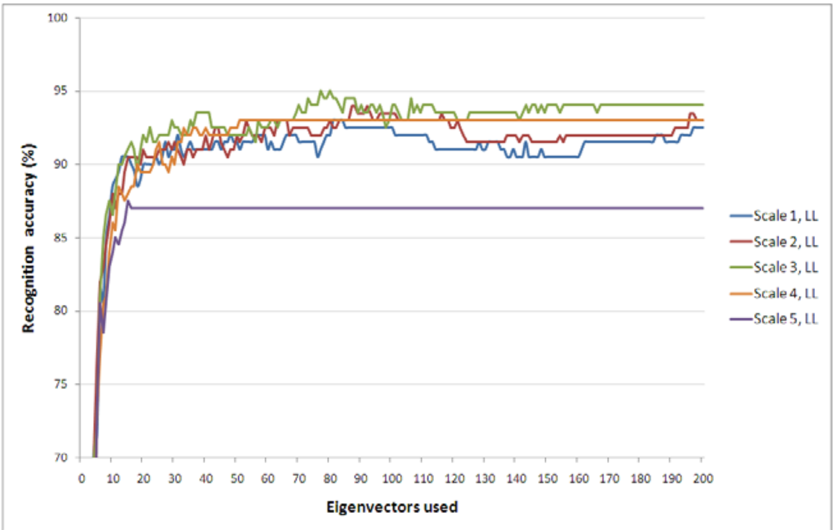

Fig. 5. Recognition results for Haar wavelet and LL quadrant

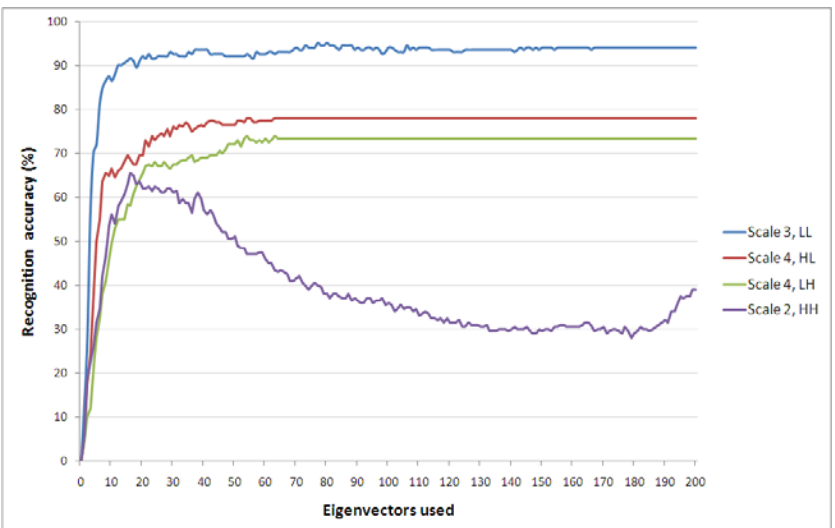

Fig. 6. Best recognition results for each quadrant using the Haar wavelet

the LL quadrant significantly outperforms the other quadrants. When compared with the Haar results, the most significant difference is that the best-performing scales for the LH and HL quadrants are the second and fifth respectively, as opposed to the fourth. However, the significance of this is minimal, due to the wide margin between these quadrants' results and those for LL.

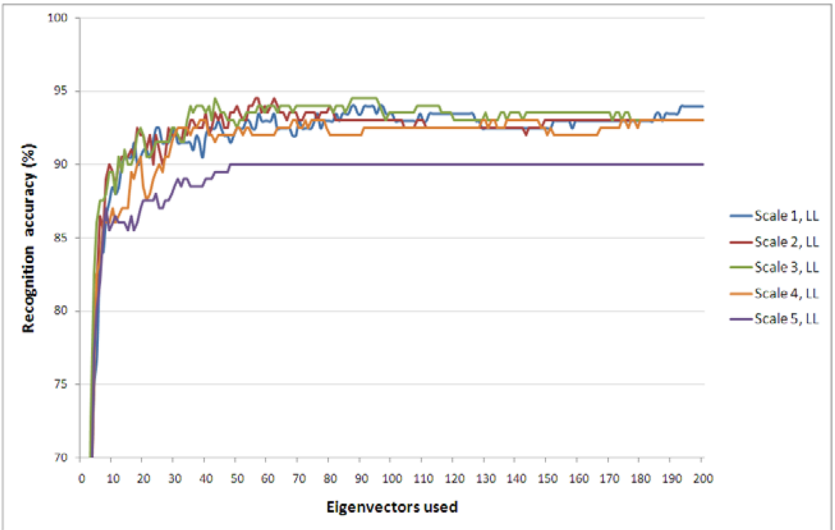

Fig. 7. Recognition results for biorthogonal 4.4 wavelet and LL quadrant

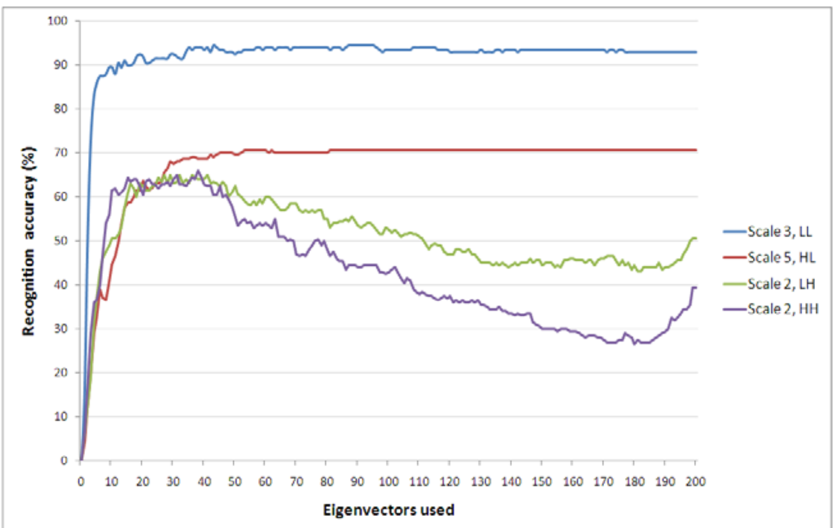

Fig. 8. Best recognition results for each quadrant using the biorthogonal 4.4 wavelet

The results achieved for these experiments help to guide decisions regarding the experiments that are still to be performed. When DWT coefficients are used for training a PCA-based recognition system, those from the LL quadrant appears to be much more discriminative in the process of face classification. As other quadrants isolate high-frequency features such as edges, small errors in alignment or facial expression between the images will significantly detract from accuracy. Conversely, the LL quadrant benefits from the removal (or reduction in impact) of high-frequency features. The consequence from these conclusions is that quadrants other than LL need not be investigated further. Remaining experiments will focus on observations from the LL quadrant. The effect of scale in the LL quadrant is less clear. Although the third scale produced best results for both wavelet filters tested, there was less variation between results for different scales than there was for different quadrants. It would therefore be appropriate to investigate the effect of scale further in remaining experiments.

3 DWT Filter Choice

3.1 Concepts

In this section, a study is performed to determine whether the choice of wavelet filter has a significant effect on recognition accuracy. Various wavelet families exist, each providing a different compromise between compactness and smoothness. Within a family, individual wavelets vary in the number of vanishing moments they contain. A vanishing moment limits a wavelet's ability to represent polynomial behavior or information in a signal. For example, a wavelet with one moment easily encodes polynomials of one coefficient, or constant signal components. A two moment wavelet encodes polynomials with two coefficients, i.e. constant and linear signal components; and three moment wavelets encode 3-polynomials, i.e. constant, linear and quadratic signal components.

Most wavelets can be described as orthonormal, meaning that they have a unit magnitude and are orthogonal. The consequence of having a unit magnitude is that convolution of a signal with a wavelet does not change the total energy of the signal. Orthogonality indicates that the inner product of the wavelet basis functions at different scales is zero. A signal can therefore be completely represented using a finite number of wavelet basis functions. The same wavelet filters are generally used for decomposition and reconstruction.

3.2 Experiments

Four wavelets are tested from each of the following wavelet families shown in Table 1. Matlab is used for experimentation and the filters are provided by the Matlab wavelet toolbox. As before, the AT&T database is used for experimentation, with five training images and five testing images used for each individual. Only the LL quadrant is used for feature extraction, at scales 1 to 5.

3.3 Results

This section presents recognition results for the examined wavelet filters. Figures 9 to 12 provide the results. Choice of wavelet family seems to have little effect on the maximum possible recognition rate – filters from the Daubechies and biorthogonal wavelet families matched up to 96.5% of faces correctly, whereas

Table 1. Wavelets filters descriptions

Wavelets	Descriptions
Daubechies	Designed to have the maximum possible number of vanishing moments for a given support size. Daubechies wavelets are widely used for solving a broad range of problems, such as identifying self-similarity properties of a signal, signal discontinuities, etc.
Symlet	Designed to be more symmetrical than Daubechies wavelets, yet compactly supported. Their increased symmetry makes them more adept at analysis of a signal than Daubechies wavelets.
Coiflet	Designed to have the maximum number of vanishing moments for both the wavelet and scaling filters, they are efficient at extracting the low frequency information in a signal.
Biorthogonal	Use separate filters for decomposing (or analyzing) and reconstructing signals. This allows each filter to be optimized for its specific purpose. Although the decomposition filters are not orthogonal with respect to the reconstruction filters, the two sets of filters are orthogonal with respect to themselves. Biorthogonal wavelets have been found to be effective in image processing applications such as fingerprint image compression [19].

filters from the symlet and coiflet families recognized 97%. The choice of filter within a wavelet family seems to be more significant. For example, although the biorthogonal 5.5 wavelet matches up to 96.5% of faces correctly, the biorthogonal 3.3 wavelet only reaches 93%. The exact nature of the relationship between wavelet and recognition performance however is unclear.

The number of non-zero coefficients in a wavelet filter (known as support size) has a number of effects on the performance of the wavelet. Filters with a larger support size are more adept at analyzing and representing complex features contained within the signal/image, however, they are more likely to be affected by artifacts at the edge of the image. Computational complexity of the wavelet transform is also increased when filters with larger support sizes are used.

Figure 13 illustrates the relationship between the support size of the low-pass filter for each wavelet with the maximum recognition rate for the filter. As all coefficients are taken from LL quadrants, only the low-pass filter is employed to calculate them. The graph shows all the maximum recognition rates achieved, for all wavelets and scales tested. The graph reveals little correlation between the two parameters. Although accuracy is highest for wavelets with a support size of 30 (Daubechies and symlet 15) and 40 (Daubechies and symlet 20), it drops again for wavelets with a support size of 50 (Daubechies and symlet 25). However, as larger filters are known to be more affected by boundary conditions, and the images used for experimentation are relatively small, this is not unexpected. Figure 14 presents the maximum recognition rates for each scale. Accuracy for

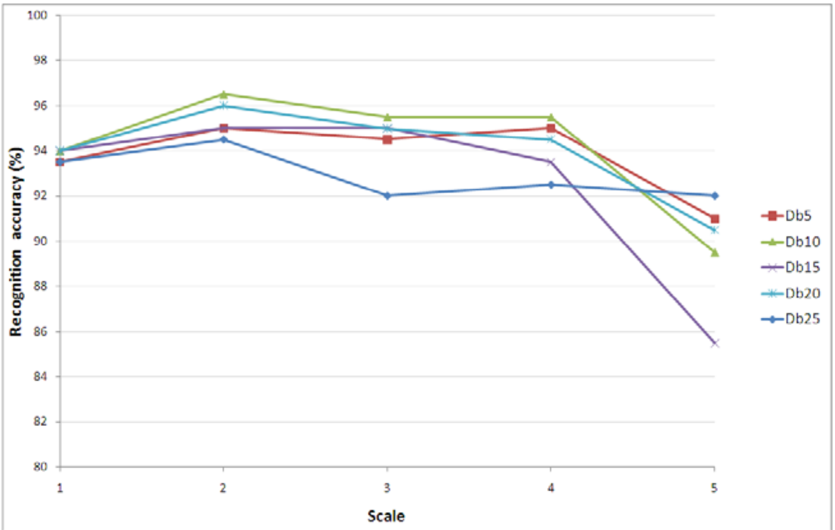

Fig. 9. Maximum recognition rates for selected Daubechies wavelets

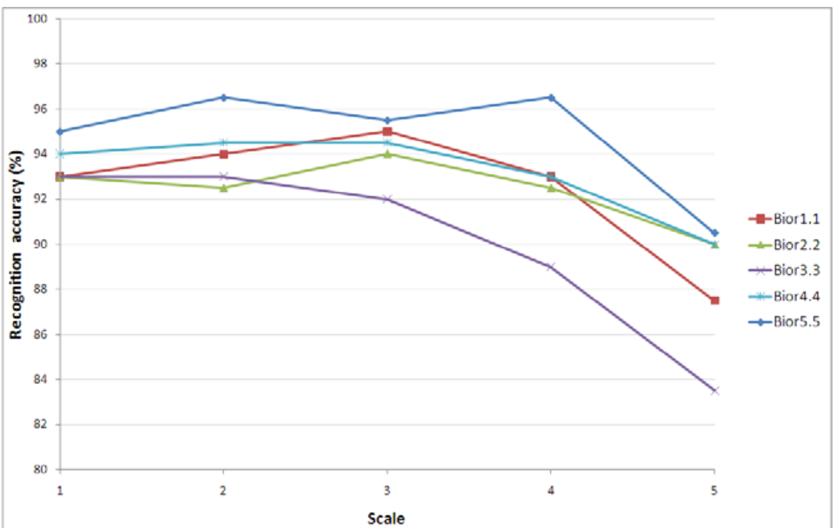

Fig. 10. Maximum recognition rates for selected biorthogonal wavelets

scales 2, 3 and 4 are similar to each other, with scale 1 providing slightly lower accuracy and scale 5 significantly lower than the best three. Scale 2 appears to provide somewhat more consistent accuracy than scales 3 or 4.

Figure 15 compares one of the best-performing wavelet filters (biorthogonal 5.5, 4^{th} scale) against recognition in the spatial domain. As can be seen form the graph, recognition accuracy is increased significantly, with maximum recognition

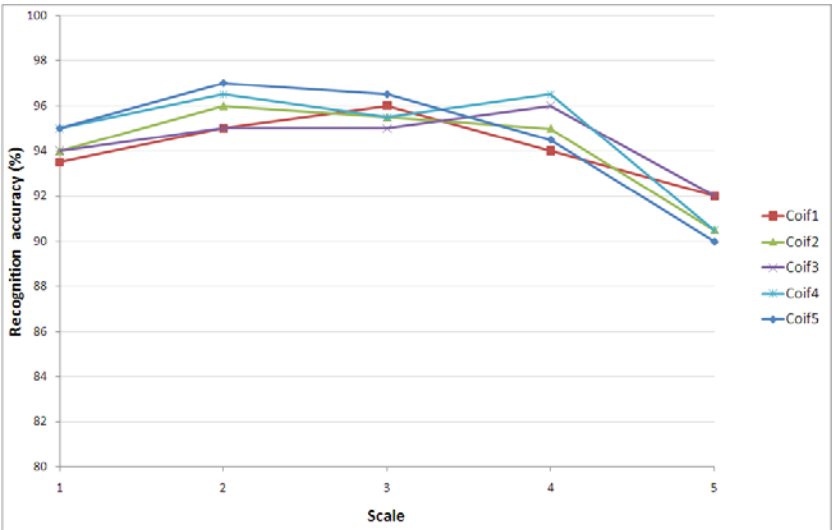

Fig. 11. Maximum recognition rates for selected coiflet wavelets

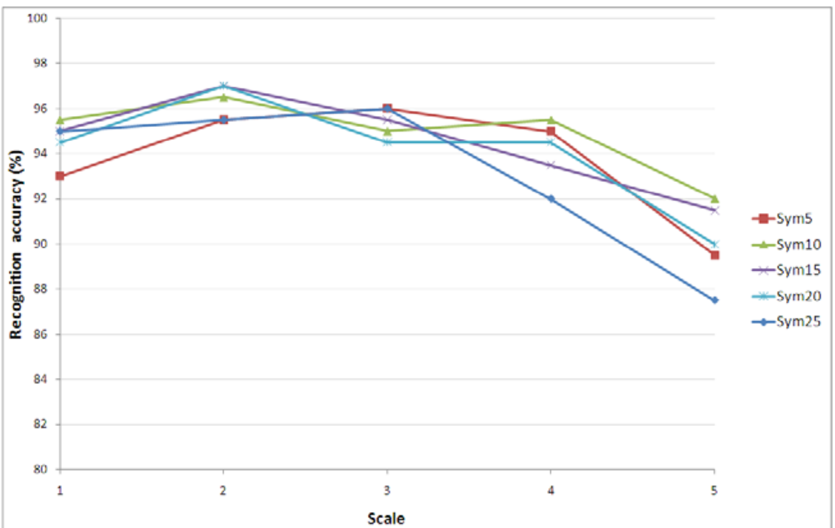

Fig. 12. Maximum recognition rates for selected symlet wavelets

rates increasing from 93% to 96.5%. This corresponds to 50% decrease in the number of incorrectly classified images.

Sample execution times are provided in Table 2. Training and classification times are given for images in the spatial and wavelet domains (biorthogonal 5.5 wavelet). Training time decreases from 0.122 seconds per image in the spatial domain to between 0.0480 and 0.0610 seconds in the wavelet domain. Although

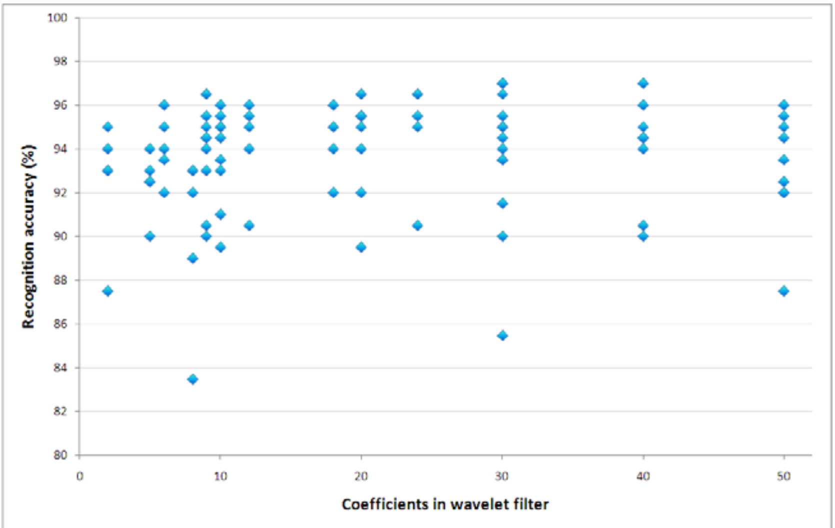

Fig. 13. Maximum recognition rate and support size of low-pass filter for all tested wavelets and scales

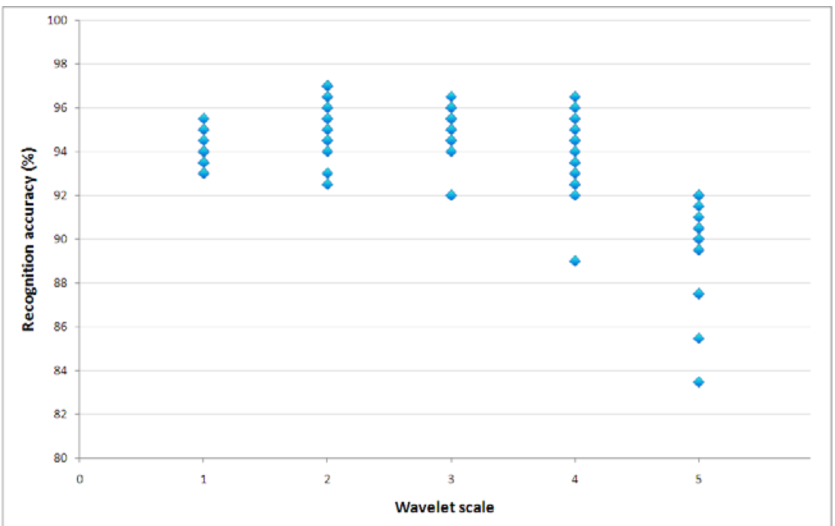

Fig. 14. Maximum recognition rate and scale for all tested wavelets and scales

there is a time penalty involved in performing DWT, this is offset by the resulting reduction in coefficients during PCA training. Classification time per spatial domain image is 0.0375 seconds and ranges from 0.0293 to 0.0479 seconds for images in the wavelet domain. The execution times were obtained using a single processor 2.4 GHz Pentium 4, with 512 MB of RAM. (Although the hardware

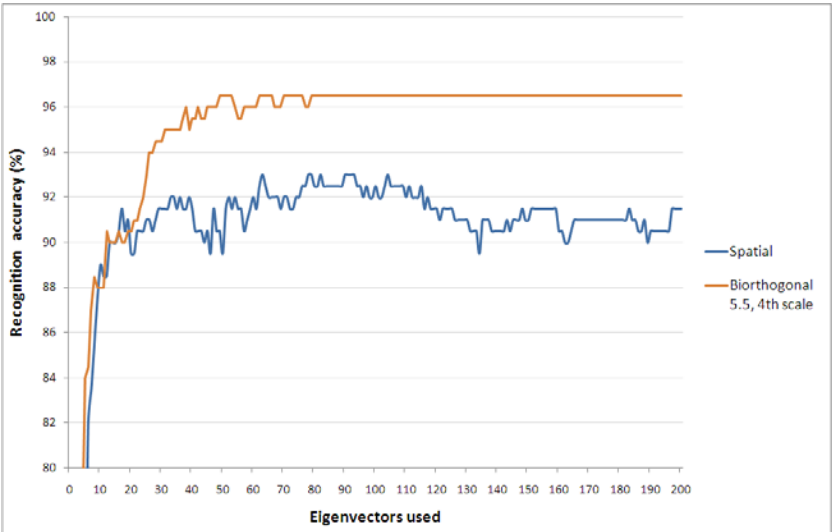

Fig. 15. Comparison of recognition results for biorhogonal 5.5 wavelet, 4^{th} scale with results for recognition in the spatial domain

Table 2. Comparison of training and classification times for AT&T database images in the spatial and wavelet domains (biorthogonal 5.5 wavelet)

Scales	Training time per image (s)	Classification time per image (s)
Spatial	0.163	0.0375
DWT, 1^{st} scale	0.0610	0.0293
DWT, 2^{nd} scale	0.0509	0.0373
DWT, 3^{rd} scale	0.0482	0.0402
DWT, 4^{th} scale	0.0480	0.0420
DWT, 5^{th} scale	0.0517	0.0479

used is not of a high specification, the execution times would differ only in magnitude if more powerful equipment had been used.)

To summarize, it is clear that DWT has the potential to significantly enhance recognition rates for PCA-based face recognition. For the AT&T database, maximum recognition rates increase from 93% for recognition in the spatial domain to 97% in the wavelet domain. There is not a substantial difference between recognition rates for the wavelet families tested, although coiflet filters produced slightly more consistent results. Across all the tested wavelet filters, there was no strong correlation between the support size of the low-pass filter and the results. Scale did appear to have an effect on results, with the 2^{nd} scale slightly outperforming the 3^{rd} and 4^{th} scales. The 1^{st} scale produced slightly lower results, with the 5^{th} scale performing significantly worse.

4 Optimizing Features by Coefficient Selection

4.1 Concepts

In this section, an approach for face recognition based on coefficient selection for DWT is presented and analyzed. One problem with many face recognition techniques is that the areas of the face images to be used for recognition have to be chosen. Images are often cropped by creating an arbitrary bounding box around the face and discarding the information outside the box [20]. There is often a trade-off between ensuring that the most relevant parts of a face image are selected for recognition and removing information that is not useful or may detract from the process.

Similarly, with PCA, eigenvectors are ordered by their corresponding eigenvalues, with the vectors with the highest eigenvalues being used to encode the face images [7]. A number of variations of this approach include excluding the initial eigenvector, or choosing eigenvectors based on their energy values. However, the number of eigenvectors chosen and the number discarded are often arbitrary choices.

The recognition approach is based on standard DWT/PCA face recognition. Figure 16 provides a general overview of the system. As can be seen from the diagram, face images firstly undergo DWT coefficient selection, followed by PCA coefficient selection. The output from this stage is a coefficient vector, which is compared with those of the gallery face images. Recognition results are returned as the identities of the most likely matches in the database.

The purpose of DWT coefficient selection is to select the most discriminative DWT coefficients. Each training image undergoes wavelet decomposition to

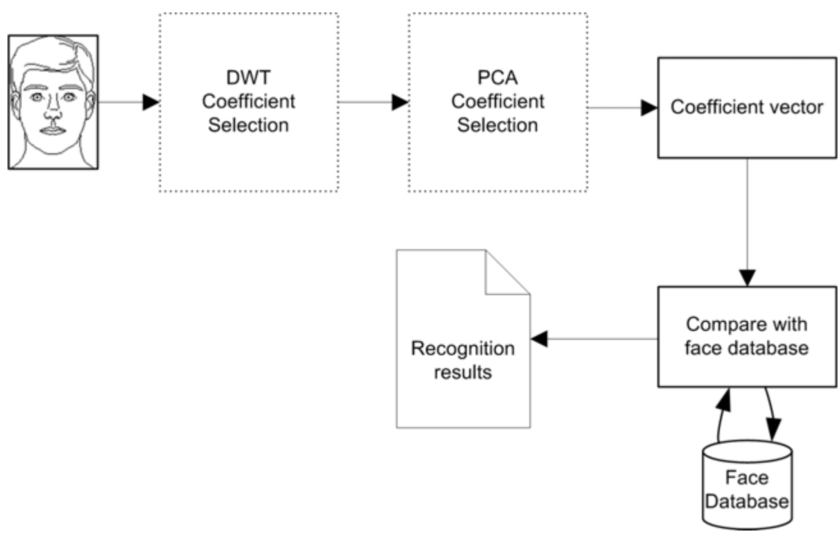

Fig. 16. System overview

a specified scale, with the low-pass coefficients being selected to form the image's observation vector. The distribution of these coefficient values is then examined to determine each coefficient's discriminative power. The inter-class and intra-class standard deviations for each coefficient are calculated and the ratio of these two values is determined. This ratio indicates how tightly the coefficient's values are clustered within each class, compared to the spread within the complete training dataset. The selection of DWT coefficients is therefore based on the maximisation of the following criterion:

$$J = \frac{\sigma_{inter}\left(A_m\right)}{\sigma_{intra}\left(A_m\right)} \tag{1}$$

where $\sigma_{inter}\left(A_m\right)$ and $\sigma_{intra}\left(A_m\right)$ represent inter-subject and intra-subject standard deviation spanned by DWT coefficients in the feature space A_m respectively. The DWT coefficients with the highest ratios are the most discriminative and chosen for recognition.

Figure 17 shows the steps involved in DWT coefficient selection. Figure 18 provides an illustration of the ratios calculated for a set of faces. Brighter areas in the image represent DWT coefficients with higher inter-class to intra-class ratios. As can be seen, brighter areas include the eyes, nose, mouth and the outline of the face. As would be expected, the image background and to a lesser extent, areas such as the forehead have lower values and have therefore been deemed to be less discriminative.

The second component of the approach is PCA coefficient selection. This initialises by performing PCA on the selected DWT coefficients, creating a set of eigenvectors and associated eigenvalues. Each training face's DWT coefficients are then projected onto the eigenvectors, producing a projection vector for each image. Eigenvectors are generally ordered by descending corresponding eigenvalues and selected using one of a number of approaches:

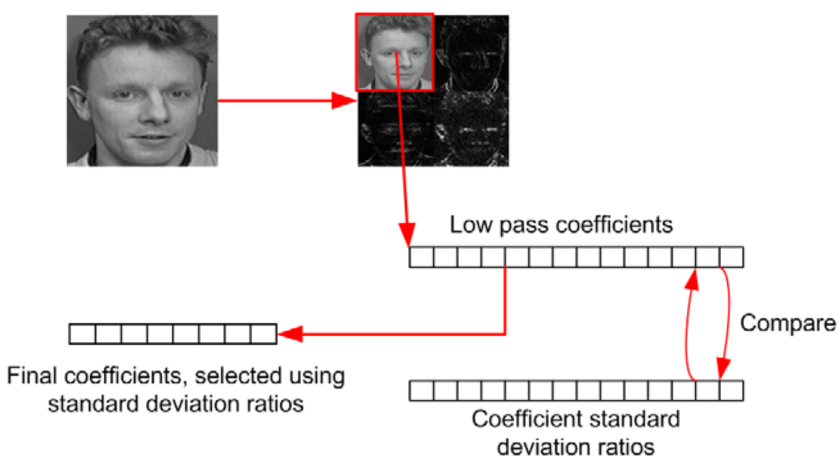

Fig. 17. DWT coefficient selection

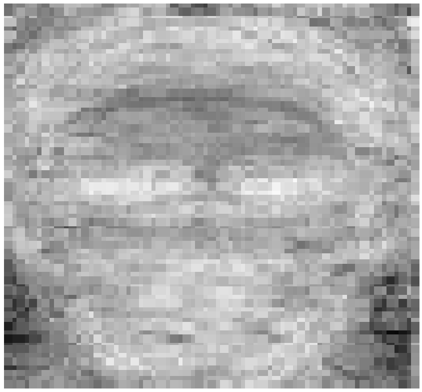

Fig. 18. Inter to intra-class ratios of DWT coefficients

- All eigenvectors corresponding to non-zero eigenvalues are used to create the eigenspace;
- The first x eigenvectors are chosen, where x often corresponds to 60% of the total eigenvector set [8];
- All eigenvectors are used apart from the first, which usually represents mostly variation in illumination [8];
- Eigenvectors are chosen based on energy values, with the first y being selected so that their cumulative energy exceeds a predetermined percentage of the total energy of all eigenvectors [21];
- Eigenvectors can be chosen based on their stretch values, where the stretch of an eigenvector is the ratio of its eigenvalue over the maximum eigenvalue [21]; and
- Eigenvectors can be chosen based on the ratios of inter-class to intra-class variance values, where those with the highest values are deemed most discriminative and selected [22].

The approach adopted for this study is based on the inter-class to intra-class standard deviation ratios. As with DWT coefficient selection, the ratios of inter-class to intra-class standard deviations are calculated. Projection coefficients with the highest ratios indicate that the associated eigenvector is highly discriminative and may contribute to better recognition accuracy. This method eliminates the need to guess which eigenvectors represent mostly variation in image illumination. Once training is complete and the most discriminative eigenvectors have been selected, classification can be performed using a simple distance measure, such as Euclidean. The adoption of this approach brings together similar coefficient selection strategies for both stages of the feature vector selection – DWT coefficient selection and PCA eigenvector selection.

4.2 Experiments

Experiments are performed which determine the benefits of DWT coefficient selection and PCA eigenvector selection separately, as well as in a combined

Table 3. Comparison of DWT coefficient selection recognition rates with those of standard DWT/PCA approach, along with percentages of DWT coefficients required to achieve maximum rate

		Recognition Rate (%)			
Wavelet	**Scale**	**Standard Approach**	**Coefficient Selection**	**Increase (%)**	**Coefficients Required (%)**
Haar	**1**	93	95	2	66
	2	94	95	1	50
	3	95	96	1	58
	4	93	95.5	2.5	68
Biorthogonal 4.4	**1**	94	96	2	69
	2	94.5	96	1.5	78
	3	94.5	96.5	2	83
	4	93	94	1	95
Coiflet 3	**1**	94	97	3	73
	2	95	97	2	85
	3	95	97	2	98
	4	96	96	0	95
Daubechies 10	**1**	94	95.5	1.5	66
	2	96.5	97.5	1	99
	3	94	96.5	2.5	75
	4	95.5	97	1.5	98
Symlet 10	**1**	95.5	96.5	1	99
	2	96.5	96.5	0	90
	3	95	95	0	90
	4	95.5	95.5	0	92
			Average Increase (%):		1.37

framework. As the technique is more suited to face data sets with little variation in pose/location, the AT&T database of faces is used for experimentation. The images contain variation in lighting, expression and facial details (for example, glasses/no glasses). For the experiments described in this study, five images for each individual are used for system training, with the other five used for testing.

4.3 Results

A number of wavelet filters are investigated, and decomposition is performed to between one and four levels. Selection percentages from 1% to 100% are tested and PCA is used for classification. Where the selection percentage is 100%, this is equivalent to no coefficient selection being applied. Results are shown in Table 3.

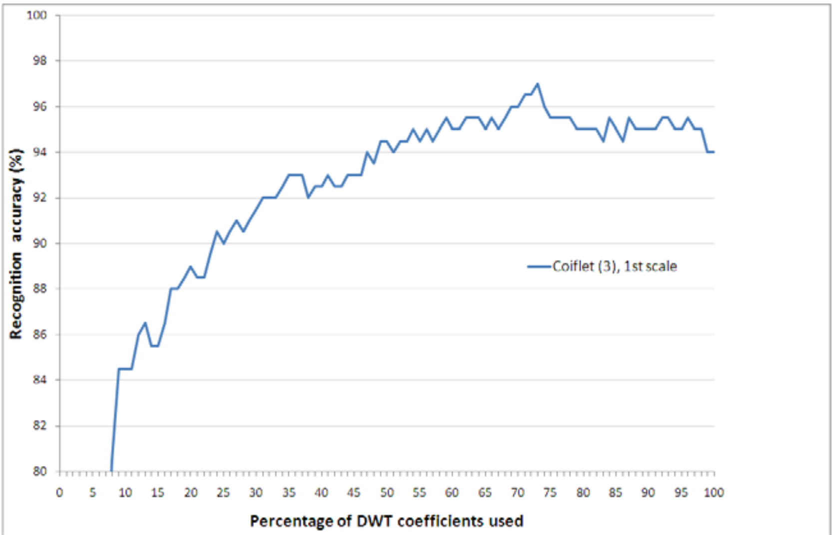

Fig. 19. Recognition rates for various DWT coefficient selection percentages, using Coiflet 3 wavelet, 1st scale

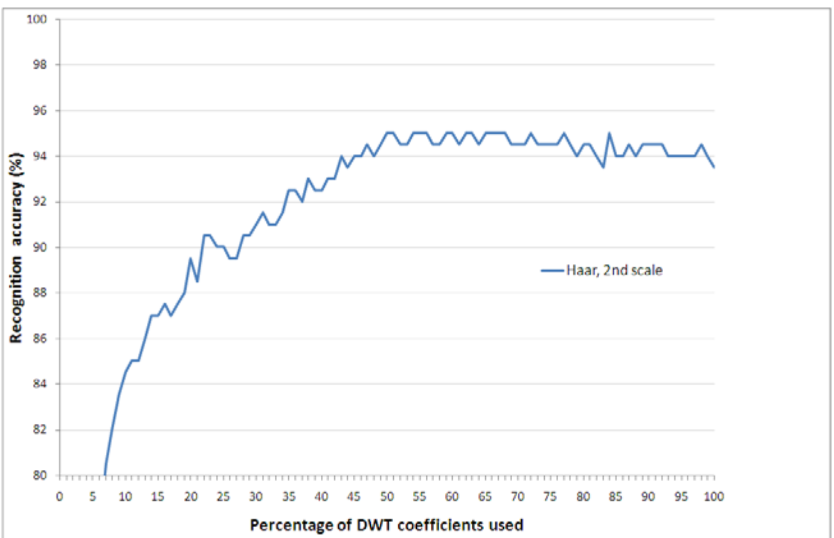

Fig. 20. Recognition rates for various DWT coefficient selection percentages, using Haar wavelet, 2nd scale

The results show that DWT coefficient selection has increased maximum recognition rate in 16 out of the 20 cases tested. The percentages of coefficients required to achieve the new maximum are also shown. In one case: Coiflet 3, 1st

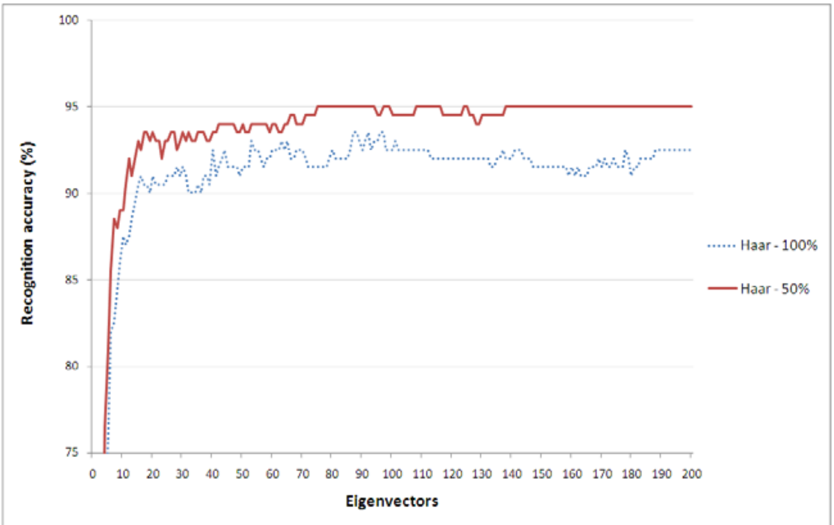

Fig. 21. Haar 2nd scale, all coefficients vs. top 50% coefficients

scale, the recognition rate has risen from 94% to 97%, which corresponds to 43% reduction in incorrectly-classified faces. The graph in Figure 19 provides more detail for this case. As the percentage of DWT coefficients increases, recognition accuracy also increases until 73% of coefficients are used. The trend then changes, with the recognition rate generally decreasing as the remaining coefficients are added. Another example: Haar, 2nd scale as can be seen in Figure 20, where the maximum recognition rate of 95% is reached with 50% of coefficients. This case is shown in more detail in Figure 21, which compares recognition accuracies for each of the two cases (50% vs. 100% of coefficients) for varying numbers of eigenvectors. The graph illustrates the full benefit of the approach, as the recognition rate for 50% of coefficients is consistently better than that for 100%.

Table 4. Comparative results on AT&T database

Method	Accuracy (%)	References
DCT/HMM	84	[23]
ICA	85	[24]
Weighted PCA	88	[25]
Gabor filters & rank correlation	91.5	[26]
2D-PHMM	94.5	[27]
NMF	96	[28]
TNPDP	96.5	[29]
LFA	97	[30]
DWT/PCA with coefficient selection	97.5	Proposed

The average increase in accuracy for this case is 2.7%. Similar improvements were seen across other wavelets and scales.

As Table 4 shows, the approach described compares well with other techniques from the literature that have used this training set. It should be noted that although the AT&T database is relatively small, the technique could be extended to other face databases. However, the coefficient selection approach is particularly suited to data sets with little variation in pose and alignment, therefore, images would have to undergo a normalization step prior to recognition. If this was performed, it is expected that results for other databases would be similar to those for the AT&T database.

5 Feature Threshold

In this section, a study is performed to investigate ways of choosing the DWT coefficient selection threshold. Although the recognition increases offered by DWT coefficient selection are significant, they are only achievable through a judicious choice of threshold. The maximum possible increases in accuracy offered by DWT coefficient selection can be seen in Table 3. Increases in recognition accuracy range from 0% to 3%, with the average increase being 1.37%. However, the results presented are the best for each wavelet and scale, found after tests employing varying numbers of DWT coefficients. For coefficient selection to be viable, the number of DWT coefficients to use as features must be chosen automatically. Two approaches are investigated for choosing this threshold.

5.1 Percentage Midpoint Average (PMA)

The first approach is referred to as percentage midpoint average (PMA). PMA assumes that a number of tests runs have been carried out with appropriate wavelets and scales, and full accuracy data obtained. For each test set, the minimum percentage of DWT coefficients required to produce the maximum recognition accuracy is recorded. The highest percentage of DWT coefficients producing the same maximum accuracy is also noted. The average of these two figures is then calculated, as the percentage midpoint for the current test set. The average of the percentage midpoints for all the test runs is calculated, with this percentage being chosen as the selection threshold.

Tests are performed on the AT&T database to determine the effectiveness of this approach. The PMA value is calculated from recognition results obtained previously, and determined to be 81.36%. DWT coefficient selection results, using 81.36% of coefficients, are shown in Table 5. The results indicate that this approach is not effective, with recognition accuracy decreasing by an average of 0.025% from the results obtained without DWT coefficient selection. This is not unexpected, as the approach is not sophisticated. It assumes that the same percentage of coefficients should be chosen in each case, regardless of the choice of wavelet filter and scale or the individual characteristics of the data set, such as the amount of background (non-face) in the image.

Table 5. Maximum recognition rates using DWT coefficient selection with PMA threshold

		Recognition Rate (%)		
Wavelet	**Scale**	**All coefficients**	**PMA**	**Increase (%)**
Haar	**1**	93	94	1
	2	94	94	0
	3	95	94.5	-0.5
	4	93	94.5	1.5
Biorthogonal 4.4	**1**	94	95	1
	2	94.5	95	0.5
	3	94.5	96.5	2
	4	93	92	-1
Coiflet 3	**1**	94	94.5	0.5
	2	95	97	2
	3	95	94	-1
	4	96	94	-2
Daubechies 10	**1**	94	94.5	0.5
	2	96.5	96	-0.5
	3	94	96.5	2.5
	4	95.5	93	-2.5
Symlet 10	**1**	95.5	95.5	0
	2	96.5	95	-1.5
	3	95	93.5	-1.5
	4	95.5	94	-1.5
		Average Increase (%):		-0.025

5.2 Optimal Ratio Average (ORA)

The second approach is referred to as optimal ratio average (ORA). As with PMA, ORA assumes that a number of tests runs have been carried out with appropriate wavelets and scales, and full accuracy data obtained. As explained in previously, DWT coefficient selection operates by calculating the ratios of inter-class to intra-class standard deviations for each coefficient: this value is used to select the most discriminative coefficients. In ORA, the cut-off ratio that produces the highest recognition rate for each test run is recorded. The average of the cut-off ratios for all test runs is chosen as the selection threshold.

Tests are performed on the AT&T database to determine the effectiveness of this approach. The ratio threshold value is calculated from the DWT coefficient selection results obtained previously. Unlike with PMA, a different percentage of DWT coefficients may be chosen for each wavelet and scale, depending on how discriminative its coefficients are. Results are provided in Table 6 and indicate

Table 6. Maximum recognition rates using DWT coefficient selection with ORA threshold

		Recognition Rate (%)		
Wavelet	**Scale**	**All coefficients**	**ORA**	**Increase (%)**
Haar	**1**	93	94.5	1.5
	2	94	94.5	0.5
	3	95	94.5	-0.5
	4	93	94	1
Biorthogonal 4.4	**1**	94	95.5	1.5
	2	94.5	95	0.5
	3	94.5	95.5	1
	4	93	93.5	0.5
Coiflet 3	**1**	94	95.5	1.5
	2	95	96.5	1.5
	3	95	96	1
	4	96	96	0
Daubechies 10	**1**	94	96	2
	2	96.5	96	-0.5
	3	94	96.5	2.5
	4	95.5	96	0.5
Symlet 10	**1**	95.5	94.5	-1
	2	96.5	95	-1.5
	3	95	95	0
	4	95.5	95.5	0
		Average Increase (%):		0.6

that the approach is effective, increasing recognition accuracy by an average of 0.6% over recognition without DWT coefficient selection. However, this is less than 50% of the maximum possible increase of 1.37% that DWT coefficient selection could provide. Although ORA is more flexible than PMA in handling varying datasets, it is likely that an optimized system would utilize one specific wavelet and scale for both system training and identification. This would allow a more relevant threshold ratio to be chosen, which would increase recognition accuracy.

6 Conclusions

In this paper, a novel feature vectors construction approach for face recognition using DWT has been discussed. The first set of experiments performed focused on the choice of DWT features. It is reveals that, where direct coefficient values were used for recognition, the LL quadrant provided the best results. For the wavelet filters tested, the highest recognition rate achieved for this quadrant was 95%. The highest accuracies for the HL, LH and HH quadrants were 78%, 74% and 66%, respectively. However, these tests did not provide enough information to indicate whether particular scales perform consistently better than others.

The second set of tests has been designed to identify which wavelet filters were the most effective at extracting features for face recognition with the specified database. The maximum recognition rates were compared for five wavelet filters each from the Daubechies, symlet, Coiflet and biorthogonal wavelet families. LL coefficients were used as features, with the first five scales investigated. The results indicated that there was no strong link between choice of wavelet family and recognition rate, although Coiflet wavelets produced the most consistent performance, across various filters and scales. When the results from all wavelet families and filters were examined together, there was no obvious correlation between the support size of the scaling filter and the maximum recognition rates.

The choice of scale did appear to have some effect, with the second, third and fourth scales outperforming the first scale by a small margin and the fifth scale by a significant margin. In case of feature optimisation by coefficient selections, the results show that DWT coefficient selection has increased maximum recognition rate in 16 out of the 20 cases tested. For instance, recognition accuracy increased from 94% to 97% for the Coiflet 3 wavelet, 1st scale.

Finally, for the feature threshold, two approaches have been investigated which are PMA and ORA. Results obtained shown that the PMA is ineffective approach, with recognition accuracy decreasing by an average of 0.025% from the results obtained without DWT coefficient selection. Unlikely, results for ORA approaches indicate better recognition accuracy by an average of 0.6%.

References

1. Jain, A., Bolle, R., Pankanti, S.: Biometrics: Personal Identification in Networked Society. Kluwer Academic Publishers, Dordrecht (1999)
2. Maltoni, D., Maio, D., Jain, A., Prabhakar, S.: Handbook of Fingerprint Recognition. Springer, New York (2003)
3. Campbell, W., Sturim, D., Reynolds, D.: Support Vector Machines Using GMM Supervectors for Speaker Verification. IEEE Signal Processing Letters 13(5), 308 (2006)
4. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face Recognition: A Literature Survey. ACM Comput. Surv. 35(4), 399–458 (2003)
5. Kanwisher, N., Moscovitch, M.: The Cognitive Neuroscience of Face Processing: an Introduction. Cognitive Neuropsychology 17(1), 1–11 (2000)
6. Amira, A., Farrell, P.: An Automatic Face Recognition System Based on Wavelet Transforms. In: IEEE International Symposium on Circuits and Systems, ISCAS 2005, vol. 6, pp. 6252–6255 (May 2005)
7. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience 3, 71–86 (1991)

8. Moon, H., Phillips, P.: Computational and Performance Aspects of PCA-based Face Recognition Algorithms. Perception 30, 303–321 (2001)
9. Yuen, P.C., Lai, J.H.: Face Representation Using Independent Component Analysis. Pattern Recognition 35(6), 1247–1257 (2002)
10. Kussul, E., Baidyk, T., Kussul, M.: Neural Network System For Face Recognition. In: ISCAS, vol. (5), pp. 768–771 (2004)
11. Mallat, S.: A Wavelet Tour of Signal Processing, 3rd edn. The Sparse Way. Academic Press, London (2008)
12. Feng, G., Yuen, P., Dai, D.: Human Face Recognition Using PCA on Wavelet Subband. Journal of Electronic Imaging 9, 226–233 (2000)
13. Harandi, M.T., Ahmadabadi, M.N., Araabi, B.N.: Face Recognition Using Reinforcement Learning. In: ICIP, pp. 2709–2712 (2004)
14. Kemal Ekenel, H., Sankur, B.: Multiresolution face recognition. Image and Vision Computing 23(5), 469–477 (2005)
15. Haar, A.: Zur Theorie der orthogonalen Funktionensysteme. Mathematische Annalen 69(3), 331–371 (1910)
16. Daubechies, I., et al.: Orthonormal Bases Of Compactly Supported Wavelets. Comm. Pure Appl. Math. 41(7), 909–996 (1988)
17. Daubechies, I.: Ten Lectures on Wavelets. Society for Industrial Mathematics (1992)
18. Dahmen, W., Aachen, R., Micchelli, C.: Biorthogonal Wavelet Expansions. IGPM Preprint (114) (May 1995)
19. Hopper, T., Brislawn, C., Bradley, J.: WSQ Gray-Scale Fingerprint Image Compression Specification. Federal Bureau of Investigation Tech. Rep. IAFIS-IC-0110-V2 (Criminal Justice Information Services, Washington, DC) (1993)
20. Bartlett, M., Movellan, J., Sejnowski, T.: Face recognition by Independent Component Analysis. IEEE Transactions on Neural Networks 13(6), 1450–1464 (2002)
21. Kirby, M.: Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns. John Wiley & Sons, Inc., New York (2000)
22. Wang, J., Plataniotis, K., Venetsanopoulos, A.: Selecting Discriminant Eigenfaces For Face Recognition. Pattern Recognition Letters 26(10), 1470–1482 (2005)
23. Nefian, A., Hayes, M.: Hidden Markov Models For Face Recognition. In: ICASSP 1998, pp. 2721–2724 (1998)
24. Kim, J., Choi, J., Yi, J., Turk, M.: Effective Representation Using ICA for Face Recognition Robust to Local Distortion and Partial Occlusion. IEEE Trans. Pattern Anal. Mach. Intell. 27(12), 1977–1981 (2005)
25. Wang, H.Y., Wu, X.J.: Weighted PCA Space And Its Application In Face Recognition. In: Proceedings of 2005 International Conference on Machine Learning and Cybernetics, Washington, DC, USA, pp. 4522–4527. IEEE Computer Society, Los Alamitos (2005)
26. Ayinde, O., Yang, Y.H.: Face Recognition Approach Based On Rank Correlation Of Gabor-Filtered Images. Pattern Recognition 35(6), 1275–1289 (2002)
27. Samaria, F.: Face Recognition using Hidden Markov Models. PhD thesis, Cambridge University Engineering Department (1994)
28. Xue, Y., Tong, C.S., Chen, W.S., Zhang, W.: A Modified Non-negative Matrix Factorization Algorithm for Face Recognition. In: ICPR 2006: Proceedings of the 18th International Conference on Pattern Recognition, Washington, DC, USA, pp. 495–498. IEEE Computer Society, Los Alamitos (2006)

29. Lu, J., Tan, Y.P.: Enhanced Face Recognition Using Tensor Neighborhood Preserving Discriminant Projections. In: 15th IEEE International Conference on, Image Processing, ICIP 2008. pp. 1916–1919 (October 2008)
30. Ersi, E.F., Zelek, J.S.: Local Feature Matching For Face Recognition. In: CRV 2006: Proceedings of the The 3rd Canadian Conference on Computer and Robot Vision (CRV 2006), Washington, DC, USA, p. 4. IEEE Computer Society, Los Alamitos (2006)