DWT/PCA Face Recognition using Automatic Coefficient Selection

Paul Nicholl[1] and Abbes Amira[2]

[1]School of Electronics, Electrical Engineering and Computer Science,
Queen's University, Belfast, United Kingdom
[2]School of Engineering and Design,
Brunel University, Uxbridge, United Kingdom
p.nicholl@qub.ac.uk, abbes.amira@brunel.ac.uk

Abstract

In PCA-based face recognition, there is often a trade-off between selecting the most relevant parts of a face image for recognition and not discarding information which may be useful. The work presented in this paper proposes a method to automatically determine the most discriminative coefficients in a DWT/PCA-based face recognition system, based on their inter-class and intra-class standard deviations. In addition, the eigenfaces used for recognition are generally chosen based on the value of their associated eigenvalues. However, the variance indicated by the eigenvalues may be due to factors such as variation in illumination levels between training set faces, rather than differences that are useful for identification. The work presented proposes a method to automatically determine the most discriminative eigenfaces, based on the inter-class and intra-class standard deviations of the training set eigenface weight vectors. The results obtained using the AT&T database show an improvement over existing DWT/PCA coefficient selection techniques.

Keywords: face recognition, principal component analysis, discrete wavelet transform

1 Introduction

Face recognition has been the subject of intense interest in recent years. Biometric techniques such as fingerprint or iris recognition have a high degree of accuracy; however, they require the subject to comply with the recognition process. Face recognition can be performed with a lesser degree of compliance, meaning that people can be identified without their consent or even their knowledge[1].

A multitude of techniques have been applied to face recognition. Many of these techniques fall into the template matching category, where faces are treated as two-dimensional arrays of intensity values. These include principal component analysis (PCA), where the variance among a set of face images is represented by a number of eigenfaces [2], and independent component analysis (ICA) [3], where faces are assumed to be linear mixtures of some unknown latent variables. In neural network models (NNMs), 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 most important [4]. In hidden Markov models (HMMs) [5], a statistical model characterizes each identity in the training database. Hybrid multiresolution approaches have received much attention in recent years. The Discrete Wavelet Transform (DWT) has been used along with a number of techniques, including PCA [2], ICA [3] and support vector machines (SVMs)[6]. DWT provides the advantage of extracting data that is useful for recognition and discarding that which is not.

In face recognition, there is often a trade-off between including all useful pixels / DWT coefficients and excluding those that detract from the process (for example, image background). This is often achieved by manual cropping of the images. It is the purpose of the work described in this paper to investigate techniques for automatically selecting DWT coefficients for DWT/PCA face recognition. In addition, selection of eigenfaces may be performed using techniques such as removing the last 40% of eigenvectors, removing the first eigenvector and assessing the energy dimension or stretching dimension [7]. This paper presents an investigation of the combination of DWT coefficient selection with eigenface selection based on inter and intra-class standard deviation ratios.

The remainder of this paper is organised as follows: in section 2, the mathematical background is explained. Section 3 describes the recognition approach used. Experimental results and analysis are presented in section 4. Section 5 contains concluding remarks.

2 Mathematical Background

2.1 Discrete Wavelet Transform

DWT allows a signal to be localized in both time and frequency. It operates by convolving a target function with wavelet kernels to obtain wavelet coefficients representing the contributions of wavelets in the function at different scales and orientations.

Figure 1a. Face image from AT&T Database of Faces

Figure 1b. 1-level Haar DWT applied to figure 1a

Figure 1c. Complete Haar decomposition applied to figure 1a

DWT can be mathematically expressed by Equation 1:

$$DWT_{x(n)} = \begin{cases} d_{j,k} = \sum x(n)h_j^*(n-2^jk) \\ a_{j,k} = \sum x(n)g_j^*(n-2^jk) \end{cases}$$ (1)

The coefficients $d_{j,k}$ refer to the detail components in signal $x(n)$ and correspond to the wavelet function, whereas $a_{j,k}$ refer to the approximation components in the signal. The functions $h(n)$ and $g(n)$ in the equation represent the coefficients of the high-pass and low-pass filters respectively, whilst parameters j and k refer to wavelet scale and translation factors. For the case of images, the 1-D DWT can be extended to two dimensions. Figure 1 illustrates the effect of applying the non-standard wavelet transform to an image from the AT&T Database of Faces [8].

2.2   Principal Component Analysis

PCA is a dimensional-reduction technique. It works by finding a new coordinate system for a set of data, where the axes (or *principal components*) are ordered by the variance contained within the training data. To achieve this, the set of face images $\{x_i\}$ is represented as a matrix X, where:

$$X = [x_1 x_2 x_3 ... x_M]$$ (2)

and X is of dimension N x M, with N being the number of pixels in an image. The 'average' face is calculated and subtracted from each face in X, giving X':

$$X' = [(x_1 - \bar{x})(x_2 - \bar{x})(x_3 - \bar{x})...(x_M - \bar{x})]$$ (3)

The principal components of this set are found by calculating the eigenvectors of the covariance matrix C, where:

$$C = \sum_{i=1}^{M} X' X'^T$$ (4)

The calculated eigenvectors are used as an orthogonal basis to represent the training set faces. The training set faces are projected onto the k eigenvectors with the highest associated eigenvalues (i.e. those representing the greatest variance).

3   Recognition Approach

The recognition approach is based on standard DWT / PCA face recognition. Figure 2 provides a general overview of the system. Face images initially undergo DWT coefficient selection, followed by PCA coefficient selection. The output is a coefficient vector which can be compared with those in the face database.
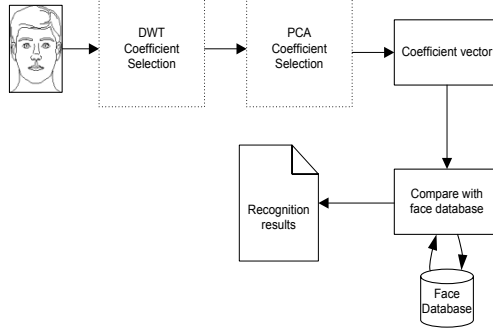

Figure 2 System Overview

3.1   DWT Coefficient Selection

The purpose of DWT coefficient selection is to select the most discriminative DWT coefficients. This proceeds by wavelet transforming each training image and extracting the low-pass coefficients to form the image's observation vector. The inter-class and intra-class standard deviations for each coefficient are calculated and the ratio of these values is determined. This ratio indicates how tightly the coefficient's values are clustered within a class compared to the spread within the complete training dataset. The coefficients with the highest ratios are deemed to be the most discriminative and are used for recognition. Figure 3 illustrates the ratios visually. The brighter areas in the image represent DWT coefficients with higher inter-class to intra-class ratios.

Figure 3 Inter to intra class ratios of DWT coefficients

3.2   PCA Coefficient Selection

The second component of the technique is PCA coefficient selection. This begins by performing PCA on all the DWT coefficients. Each training face's DWT coefficients are then projected onto the eigenvectors. Those eigenvectors with the highest eigenvalues represent the greatest variation within the training set, however, they indicate the variance across the entire dataset without taking into account

the class of the training set data. In the proposed technique, the inter-class and intra-class standard deviation are calculated for each eigenface. The ratios of these values are calculated and the eigenfaces with the highest ratios are used for recognition.

4   Results and Analysis

4.1   Datasets

Experiments were carried out using the AT&T Database of Faces [8], which contains ten greyscale images each of forty individuals. Five images for each individual were used for system training, with the other 5 used for testing.

4.2   DWT Coefficient Selection

The first set of experiments was designed to test whether DWT coefficient selection improved results. The training set images were wavelet-transformed using either the Haar, Biorthogonal 9/7 or Coiflet (3) wavelet, to the 1st, 2nd or 3rd scale. The top 90%, 70% or 50% of DWT coefficients were selected and passed to standard PCA for recognition. In addition, the results for 100% of coefficients (i.e. no coefficient selection) were also obtained. Results are shown in Table 1, with the best recognition rates for each wavelet and scale shown in bold. As can be seen from the results, DWT coefficient selection improved results in all cases, apart from when the Haar wavelet / 3rd scale were used. The best results were obtained with the Coiflet (3) wavelet, where results improved from a maximum of 94% to a maximum of 96%.

Table 1 Maximum recognition rates for DWT coefficient selection, for various selection percentages

| Wavelet | Scale | Coefficient selection | | | |
		None	90%	70%	50%
Haar	1	93	93.5	**95**	93
	2	94	94.5	94.5	**95**
	3	**95**	94	94	93
Bior9/7	1	94	**95**	95	94.5
	2	94.5	**95.5**	95	94.5
	3	94.5	**95.5**	**95.5**	93
Coif(3)	1	94	95	**96**	94.5
	2	95	**96**	**96**	94.5
	3	95	**96**	94.5	90.5

However, the increase in maximum recognition rate does not convey the full improvement in recognition performance. Figure 4 shows the recognition rates for Haar, scale 2, using no coefficient selection versus 50% coefficient selection. As can be seen from the graph, the recognition rate using coefficient selection

is *consistently* better than that without selection. In addition, the greatest difference between the two recognition rates is 5%, which is a highly significant improvement. Improvements of this magnitude were seen across other wavelets and scales.
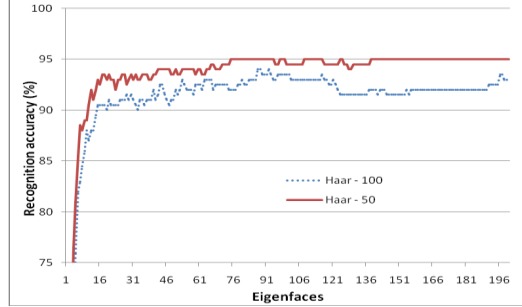

Figure 4 Haar scale 2 no selection vs. top 50% coefficients

4.3   PCA Coefficient Selection

The next set of experiments involved testing PCA coefficient selection, using all (rather than selected) DWT coefficients. Results are shown in table 2, with the best recognition rate for each wavelet and scale shown in bold. As can be seen from the results, there was an improvement in the results for 7 out of the 9 cases tested, with no change in the other 2 cases. The improvements were smaller than those brought about by DWT coefficient selection, with the biggest improvement being shown for the Coiflet(3) wavelet at the 1st scale, where maximum recognition accuracy increased from 94% to 95.5%.

Table 2 Maximum recognition rates for PCA coefficient selection

Wavelet	Scale	Standard DWT/PCA	DWT/PCA with PCA coefficient selection
Haar	1	93	**93.5**
	2	94	**94.5**
	3	95	95
Bior9/7	1	94	**95**
	2	94.5	**95**
	3	94.5	94.5
Coif(3)	1	94	**95.5**
	2	95	**96**
	3	95	**95.5**

4.4   Combined Coefficient Selection

The aim of the final set of experiments was to combine the two coefficient selection techniques to determine whether they had a synergistic effect. One experiment was performed for each wavelet and scale, using the selected DWT coefficient percentage that

was found to produce the best results (see table 1). Results are shown in table 3, and show the combined coefficient selection result beside the best DWT coefficient result for each wavelet and scale. The best recognition rate for each wavelet and scale is shown in bold. In 4 out of the 9 cases tested, recognition rates increased over those for DWT coefficient selection, whereas in 1 case, there was a slight decrease. In the other 4 cases, there was no change. The best result was for the Biorthogonal 9/7 wavelet and the 2nd scale, where accuracy increased from 95.5% to 96.5%.

Table 3 Maximum recognition rates for combined coefficient selection

Wavelet	Scale	DWT coeffs (%)	DWT coefficient selection	Combined coefficient selection
Haar	1	70	95	**95.5**
	2	50	95	95
	3	70	94	94
Bior9/7	1	70	95	**95.5**
	2	90	95.5	**96.5**
	3	70	95.5	95.5
Coif(3)	1	70	**96**	95.5
	2	70	96	**96.5**
	3	90	96	95.5

From the results, it can be seen that there is only a small benefit to using both DWT coefficient selection and PCA coefficient selection. However, combined coefficient selection tended to 'flatten out' the recognition rate graphs, eliminating the dips as the number of eigenfaces increased towards the maximum. As the number of eigenfaces used in a real-world situation would be predetermined, without having the luxury of looking at accuracy graphs, this is a useful benefit. As can be seen from table 4, DWT/PCA with coefficient selection compares well with existing techniques. Computation time for training is reduced from 0.760s per image for standard DWT/PCA recognition to 0.426s for DWT coefficient selection and 0.446s for combined coefficient selection (50% selected coefficients). Recognition time is virtually unaffected with computation times of 0.11 to 0.12s per image. The reason that computation times do not increase as a result of the extra step is due to the decreased number of coefficients used for PCA training and recognition.

5   Conclusions

In this paper, a novel technique for selecting DWT coefficients and PCA coefficients in the area of face recognition has been presented. Tests have been performed on the AT&T Database, showing that recognition rates are increased when the technique is deployed. In one case recognition accuracy increased from 93% to 95.5%, which represents a significant drop in the error rate.

Table 4 Comparative Results

Method	Accuracy(%)	Ref
DCT/HMM	84	[5]
ICA	85	[3]
Weighted PCA	88	[2]
Gabor Filters & Rank Correlation	91.5	[9]
DWT/PCA with Coefficient Selection	96.5	(Proposed)

6   References

[1] W. Zhao, R. Chellappa, P.J. Phillips and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp 399–458 (2003).

[2] H.Y. Wang and X.J. Wu, "Weighted PCA space and its application in face recognition," in *Proceedings of 2005 International Conference on Machine Learning and Cybernetics,* Washington, DC, USA: IEEE Computer Society, pp 4522–4527 (2005).

[3] J. Kim, J. Choi, J. Yi and M. Turk, "Effective representation using ICA for face recognition robust to local distortion and partial occlusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp 1977–1981 (2005).

[4] E. Kussul, T. Baidyk, and M. Kussul, "Neural network system for face recognition." in *ISCAS* (5), pp 768–771 (2004).

[5] A. Nefian and M. Hayes, "Hidden markov models for face recognition," in *ICASSP98*, pp 2721-2724 (1998).

[6] H.K. Ekenel and B. Sankur, "Multiresolution face recognition." *Image Vision Comput.*, vol. 23, no. 5, pp 469–477 (2005).

[7] M. Kirby, *Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Pattern*, Wiley, New York, 2000.

[8] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in *IEEE Workshop on Applications of Computer Vision*, Sarasota, Florida (1994).

[9] O. Ayinde and Y.H. Yang, "Face recognition approach based on rank correlation of gabor-filtered images." *Pattern Recognition*, vol. 35, no. 6, pp 1275–1289 (2002).