Combination of PCA and Wavelet Transforms for Face Recognition on 2.5D Images

Chi-Fa Chen[1], Yu-Shan Tseng[1] and Chia-Yen Chen[2]
[1]Department of Electrical Engineering,
I-Shou Univeristy, Kaohsiung, Taiwan
cfchen@isu.edu.tw
[2]CITR, Tamaki Campus,
The University of Auckland, Auckland, New Zealand
yen@cs.auckland.ac.nz

Abstract

This work presents a method by which increased accuracy in face recognition using the *Principal Components Analysis* (PCA) on wavelet transforms is achieved by using 2.5D depth maps as the source of facial features. Comparable or better results are yielded in less processing time under tested conditions. A variety of classifiers are employed, such as the *nearest centre* (NC), *nearest feature line* (NFL) and *linear discriminant analysis* (LDA) classifiers. The photometric stereo method was used to acquire 2.5D depth maps.

Keywords: Principal Components Analysis, wavelet transform, photometric stereo

1 Introduction

Face recognition has traditionally been performed with facial feature extracted from two dimensional images. The features are placed in sets, and decision boundaries are created by which input images may be classified and identified.

By acquiring a 2.5D depth map through such methods as the photometric stereo method, more information related to facial features may be extracted. The increase in information can be used to aid the face recognition process.

In this system, the facial depth information is extracted and used to replace traditional two dimensional gray level images in face recognition. The extra depth information allows for a more accurate representation and identification of an individual's face.

2 Procedures

Using the photometric stereo method, three images were used to create a depth map. Six viewing directions were used for each of the 101 subjects. The depth maps were then used to create a face database.

Three out of the six viewing directions for each individual were used as training sets for face recognition. Features were extracted and then compared, using the nearest centre, nearest feature line, linear discriminant analysis and a combination of the linear discriminant and the nearest feature line classifiers.

2.1 Image Acquisition

The photometric stereo method is used to capture 2.5D image representations of faces. Facial features are later extracted from the images captured. Using the photometric stereo method, three light sources are required to construct an acceptable environment. The method involves a calibrated camera which captures two dimensional gray level images from the same viewing angle under different light source directions. The photometric stereo method is then used to reconstruct a 2.5D representation of the desired subject from the images.

In the acquisition of the images used herein, different viewing directions were used. This was performed to approximate real world conditions where ideal viewing angles are not necessarily obtainable.

As 2.5D depth maps contain more surface information than a traditional 2D gray level images, since it also contains information on surface geometry, which may be useful in matching faces with different orientations. It is expected that the matching accuracy would show an improvement from using just 2D images. In this system, the extracted depth information is represented by gray levels to generate a depth image. The depth image replaces the traditional gray level images used during the recognition process.

2.2 Face Database Creation

A face database was created from the results of each image acquisition. The experimental database of 2.5D face depth information include the following:

- 101 subjects; 94 males and 7 females.
- Six different viewing directions for each subject. The viewing directions used resulted in the subject's face looking straight ahead, tilted slightly upwards and being tilted 15° and 30° to both the left and the right.
- The dimensions of each captured image is 256×256 pixels, the reconstructed depth map has the same dimensions.

In total, the database contained 606 depth map images (101 subjects × six images per subject). Examples of depth map images used in the database may be seen in Fig. 1.

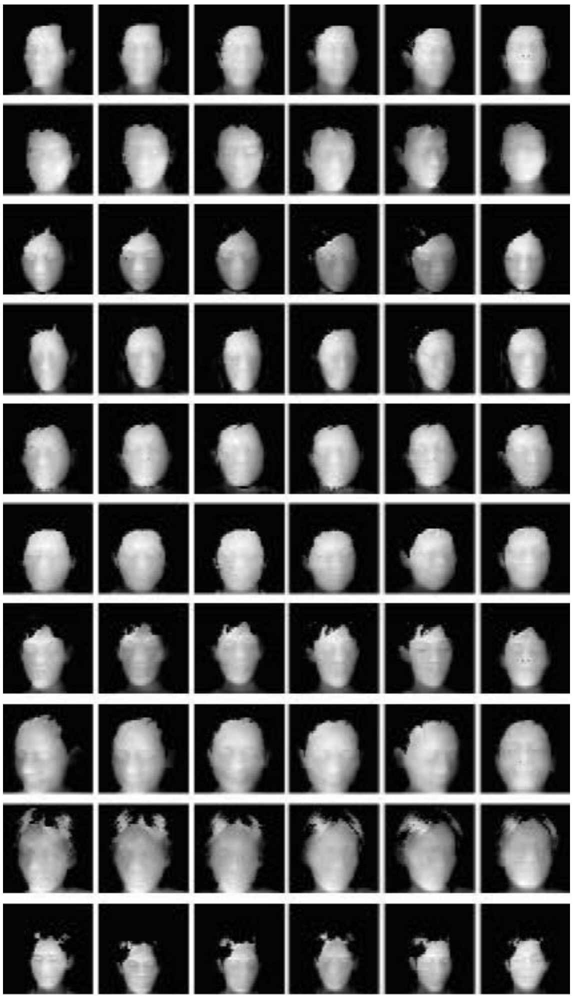

Figure 1: Example images from the face database, depth maps for each subject are acquired from six different viewing directions.

Using the photometric stereo method, each depth map was reconstructed from three input images with different illumination directions, there, there is a total of 1818 two dimensional gray level images (606 depth maps × three images per depth map).

2.3 Training Set Selection

Three out of the set of six depth maps per subject were used as part of the training set for the face recognition system. The remaining three depth maps per subject were used as the input images for testing the accuracy of the face recognition system. Three sets of experiments were performed in an attempt to refine the results. The experiments were:

- The first experiment has training set comprising of the left 30°, right 30° and upwards facing faces. This results in a training set of 303 depth maps, with the remaining 303 depth maps being used as input data for testing the recognition system.
- The second experiment has training set comprising of the left 30°, right 30° and forwards facing faces. This also results in a training set of 303 depth maps with the remaining 303 depth maps used as input data. The results from this training set were not as accurate as the first experiment. The faces selected are denoted by a cross under their images as shown in Fig. 2.
- The third experiment has training set comprising of the left 30°, right 30° and forwards facing faces. However, the upward facing faces has excluded from the testing set. The reason is because the upwards facing faces would often be mismatched. The recognition rate has improved when forwards facing faces are eliminated from the testing set.

The results for each of the above three experiments are shown below.

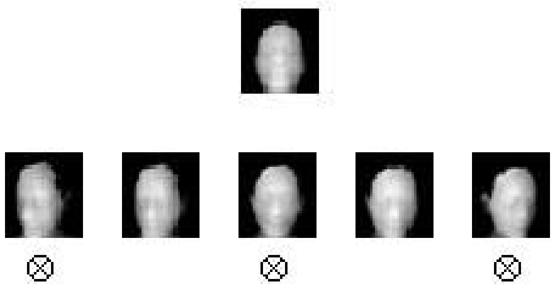

Figure 2: The cross marks the selection of face depth maps used in the second experiment.

2.4 Feature Extraction

Feature extraction on the training set is performed by combining the method of principal component analysis (PCA) and the method of wavelet transforms. Principal component analysis has been used in face recognition, as described by Turk and Pentland [1]. It involves arranging eigenvectors of an image in such a way that *eigenfaces* are formed. The eigenfaces are so called due to their resemblance to a face.

The wavelet transform concentrates the energy of the image signals into a small number of wavelet coefficients. Using two dimensional wavelet transforms, an image $f(x,y)$ can be represented as:

$$\begin{aligned} f(x,y) &\approx s_J \phi_J(x,y) + \sum_{j=1}^{J} d_j^v \psi_j^v(x,y) \\ &\quad + \sum_{j=1}^{J} d_j^h \psi_j^h(x,y) + \sum_{j=1}^{J} d_j^d \psi_j^d(x,y) \\ &= S_J + \sum_{j=1}^{J} D_j^v + \sum_{j=1}^{J} D_j^h + \sum_{j=1}^{J} D_j^d \ , \end{aligned} \tag{1}$$

where the two dimensional wavelets are the tensor product of the one dimensional wavelets as below:

$$\begin{aligned} \phi(x,y) &= \phi(x) \times \phi(y) \\ \psi^v(x,y) &= \phi(x) \times \psi(y) \\ \psi^h(x,y) &= \psi(x) \times \phi(y) \\ \psi^d(x,y) &= \psi(x) \times \psi(y) \ , \end{aligned} \tag{2}$$

where J are the stages of the wavelets. The first stage is called the approximation image, the other three are called the vertical, horizontal and diagonal images. The energy of the original image concentrates within the approximation image. Images showing the most significant components (significant component map) in using only three stages of wavelets are shown in Fig. 3 .

The feature extraction method used the approximation component of the wavelet coefficients in the principal component analysis. Assuming that the approximation is $a_i, i = 1, \ldots, M$, and that there are M images in the training set, then we have an image feature such that:

$$y = w^T (a - A) \ , \tag{3}$$

where the mean image is:

$$A = \frac{1}{M} \sum_{i=1}^{M} a_i \ , \tag{4}$$

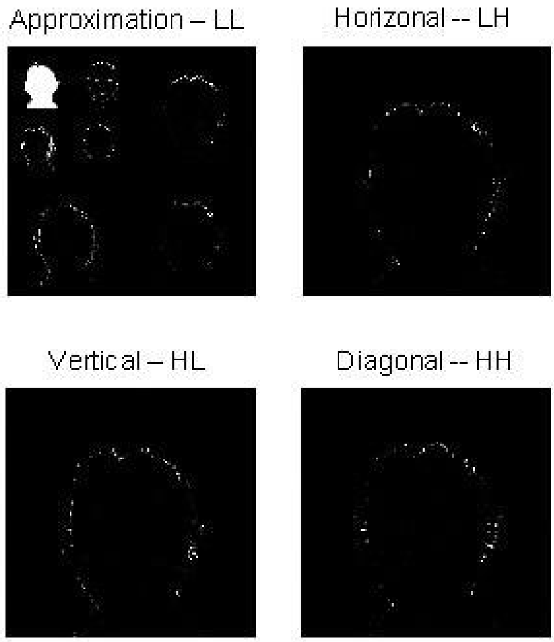

Figure 3: Significant component maps using only three stages of wavelets.

and w represents the eigenvectors corresponding to large eigenvalues of the covariance matrix $(a - A)(a - A)^T$.

2.5 Wavelet Stage Selection

Experiments were performed to determine the best wavelet stage to use. The experiments were performed using Daubechies wavelet and the nearest feature line classifier. The results are shown in Table 1. The columns represent:

- Stage: The wavelet stage used to determine the results.
- Success: The recognition success rate expressed as a percentage of correct matches out of all inputs.
- $Time_{proc}$: The amount of processing time in seconds required for the distinct features of an input image to be extracted.
- $Time_{recog}$: The amount of time in seconds required to compare the extracted features of an input image to those from the training set to make a decision.

Performance was measured with respect to a Celeron 533Mhz with 256Mb of RAM.

Table 1: Experimental results for estimation of the wavelet stage.

Stage	Success (%)	$Time_{proc}$(s)	$Time_{recog}$(s)
2	70.3	195	0.8713
3	74.3	192	0.8416
4	80.9	199	0.8647
5	**86.5**	**209**	**0.9043**
6	79.9	223	0.9505

From the results in Table 1, it can be seen that approximation components of the fifth stage yield the highest success rates. As a result, the fifth stage was selected.

2.6 Comparison Methods

Depth maps were converted to traditional two dimensional gray level depth images by replacing depth values with corresponding gray levels. The depth images were then used in the face recognition process.

For comparison, two dimensional images that were captured for the photometric stereo method image acquisition were also used in the face recognition process. The training set used for the two dimensional images are the viewing equivalents for each of the face angles.

3 Results

Results from conventional recognition methods that use gray level images are compared with recognition results using depth images.

Different classifiers were employed and compared. Each classifier used the resulting feature set from the combination of the wavelet and principal component analysis. They are listed in the tables below:

- WP: Using the nearest center classifier.
- WPN: Using the nearest feature line classifier.
- WPL: Using the linear discriminant analysis classifier.
- WPLN: Using the nearest feature line plus the linear discriminant analysis classifier.

Training sets were selected from the face database. The training sets were used to extract key features from each individual. The key features were then later used for face recognition comparisons.

Three experiments were performed. The results are placed below in the same order as the training set selection.

3.1 Experiment One

The 2.5D training set is as described in Section 2.3. This experiment uses the initial training set. The two dimensional training set consists of images taken from equivalent viewing directions. The test set has 1515 images, resulting from the 1818 images of the original less the two dimensional training set. The results are as shown in Table 2.

Table 2: Results for experiment one. The two dimensional results are in brackets.

Classifier	Success (%)	$Time_{proc}$(s)	$Time_{recog}$(s)
WP	75.3 (67.8)	220 (266)	0.76 (0.88)
WPN	86.5 (72.4)	221 (266)	0.89 (1.09)
WPL	83.2 (78.7)	206 (273)	0.68 (0.89)
WPLN	84.2 (85.0)	231 (266)	0.97 (1.09)

3.2 Experiment Two

The 2.5D training set is as described in Section 2.3. This experiment uses the second training set. The two dimensional training set consists of images taken from equivalent viewing directions. The test set has 1515 images. The results as shown in Table 3.

Table 3: Results for experiment two. The two dimensional results are in brackets.

Classifier	Success (%)	$Time_{proc}$ (s)	$Time_{recog}$ (s)
WP	77.2 (68.9)	221 (267)	0.70 (0.88)
WPN	86.5 (74.2)	223 (268)	1.02 (1.09)
WPL	86.2 (80.8)	248 (269)	0.77 (0.89)
WPLN	88.8 (86.1)	248 (268)	1.02 (1.10)

3.3 Experiment Three

The 2.5D training set is as described in Section 2.3. This experiment uses the last training set. The two dimensional training set consists of images taken from equivalent viewing directions. Again, the test set has 1515 images. The results as shown in Table 4.

Table 4: Results for experiment three. The two dimensional results are in brackets.

Classifier	Success (%)	$Time_{proc}$(s)	$Time_{recog}$(s)
WP	82.7 (74.7)	208 (268)	0.70 (0.88)
WPN	93.1 (79.1)	210 (268)	0.93 (1.09)
WPL	92.1 (84.6)	265 (273)	0.80 (0.90)
WPLN	95.1 (89.4)	270 (269)	1.01 (1.10)

4 Conclusion

The experiments performed in this work have shown that face recognition with training sets created from 2.5D depth maps are more accurate and faster when compared to traditional two dimensional images. The most accurate recognition rate for 2.5D data was 95.1%, as compared to 89.4% from two dimensional data.

5 Future Possibilities

Several areas of improvement and investigation are possible.

- Under tested conditions, every input face was guaranteed to match a face within the given database. Future possibilities include testing faces not in the database to identify the rate of false positive and negative matches.

- Two dimensional comparisons were performed. Further investigation could include extending face recognition methods to compare 2.5D data.

References

[1] M. A. Turk and A. P. Pentland, ”Eigenfaces For Recognition”, in *Journal of Cognitive Neuroscience*, 3(1), pp. 71–96, (1991).

[2] A. F. Eriksson and D. M. Weber, ”Towards 3-Dimensional Face Recognition”, in *Proc. 5th IEEE AFRICON*, 1, pp. 401–406, (1999).

[3] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Massachusetts, (1993).

[4] S. Z. Li and J. Lu, ”Face Recognition Using the Nearest Feature Line Method”, in *IEEE Transactions on Neural Networks*, 10(2), pp. 439-443, (1999).

[5] G. D. Guo, S. Z. Li and K. L. Chan, ”Support Vector Machines For Face Recognition”, in *Image and Vision Computing*, 19, pp. 631-638, (2001).

[6] G. Bellaire, and K. Schlüns, ”3-D Object Recognition by Matching the Total View Information”, in *Proc. Int. Conf. Patter Recognition*, pp. 534-539, (1996).

[7] R. Brunelli and T. Poggio, ”Face Recognition: Features Versus Templates”, in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(10), pp. 1042-1052, (1993).

[8] Q. Wu, Z. Liu, Z. Xiong, Y. Wang, T. Chen and K. R. Castleman, ”On Optimal Subspace for Appearance-Base Object Recognition”, *Proceedings of Int. Conf. Image Processing*, (2002).

[9] C. S. Burrus, R. A. Gopinath and H. Guo, ”Introduction to Wavelets and Wavelet Transforms”, Prentice-Hall International, (1998).

[10] A. Graps, ”An Introduction to Wavelets”, *IEEE Computational Science and Engineering*, 2(2), pp. 50-61, (1995).