Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 2, No. 2, pp. 277–288, May 2006

AUTOMATIC PARAMETERS SELECTION FOR EIGENFACES

Ronny Tjahyadi, Wanquan Liu and Svetha Venkatesh

Abstract: In this paper, we investigate the parameters selection for Eigenfaces. Our focus is on the eigenvectors and threshold selection issues. We will propose a systematic approach in selecting the eigenvectors based on relative errors of the eigenvalues for the covariance matrix. In addition, we have proposed a method for selecting the classification threshold that utilizes the information obtained from the training data set. Experimentation was conducted on two benchmark face databases, ORL and AMP, with results indicating that the proposed automatic eigenvectors and threshold selection methods produce better recognition performance in terms of precision and recall rates. Furthermore, we show that the eigenvector selection method outperforms energy and stretching dimension methods in terms of selected number of eigenvectors and computation cost.

Key words: *face recognition, eigenfaces, principal component analysis, matrix approximation*

Mathematics Subject Classification: *68T10, 65F15*

1 Introduction

Research into face recognition has flourished in recent years due to the increased need for surveillance and more secure systems and has attracted multidisciplinary research efforts. Many approaches have been proposed to represent and recognize faces under various viewing conditions and numerous algorithms have been developed such as statistical-based, neural networks and feature-based algorithms [5].

Eigenfaces is a popular face recognition method based on Principal Components Analysis (PCA) [19]. This approach is one of the preferred techniques in face recognition due to its simplicity, ability to perform in real-time situations and robustness under varying illuminations. Many algorithms have been proposed to the extension of Eigenfaces such as in Modular Eigen-spaces [14], Global Eigen approach [11], and Bayesian modeling [13].

There are several important issues in the process of implementing Eigenfaces. The first is the issue of eigenvector selection [18, 19, 20]. Selecting appropriate eigenvectors to create the eigenspace is crucial to the computational cost, threshold selection and performance of Eigenfaces. Although several approaches have been introduced to overcome this issue [20], there are still no methods for selecting the appropriate number of eigenvectors systematically and still not adequate to represent a high dimensional space with lowest dimensional subspace with good approximation.

The next issue is the threshold selection [4, 16, 19]. Manual threshold selections based on ad hoc heuristics and Receiver Operating Characteristics (ROC) curves are the common approaches. However, such methods are not suitable in automatic face recognition systems as they require a manual step and are tedious in practice.

In this paper, we will address these two basic issues in Eigenfaces, namely the eigenvector and threshold selection. First, we develop a systematic approach for selecting the eigenvectors based on the relative errors of the eigenvalues. This is based on observations from algebraic knowledge that relative error is one way of characterization for the capacity of subspace spanned by the eigenvectors corresponding to large eigenvalues. Second, we propose a technique to automatically select the threshold that utilizes the information obtained from the training database. The performance of the proposed techniques will be investigated using five datasets derived from ORL and AMP databases. Finally, the performance of the eigenvectors selection method is compared with other existing methods.

The rest of paper is organized as follows. Section 2 explains a brief description of Eigenfaces and summarizes the related work on eigenvectors and threshold selection. Section 3 describes the proposed methods in selecting the eigenvectors and the classification threshold. Experimental results are presented in Section 4. Finally, Section 5 concludes this paper.

2 Preliminaries

2.1 Eigenfaces

The motivation of Eigenfaces was taken from the work by Sirovich and Kirby in 1987 [9]. Turk and Pentland [19] developed a PCA face recognition system known as Eigenfaces. The PCA extracts the eigenvectors and eigenvalues from a covariance matrix constructed from the original training images. The first orthogonal dimension of this eigenspace captures the greatest amount of variance in the database whilst its last one captures the least amount of variance in the database. The training images are then projected into the eigenspace, thus creating a lower dimensional space. To test whether an image corresponds to images in the training database, the test image is projected into the eigenspace and the Euclidean distance between the test image and training image used as a basis for matching [19].

2.2 Eigenvectors Selection

Eigenfaces encounters an issue in the selection of eigenvectors to represent the optimal subspace [18, 19, 20]. Several approaches have been proposed for selecting the eigenvectors. However, none of these methods are scalable in the sense of those related to the variance with the database. Five of the representative approaches among them are described below [20]:

1. Standard eigenspace projection:
   All eigenvectors corresponding to non-zero eigenvalues are used to create the eigenspace.

2. Removing the last 40% of the eigenvectors:
   Since the eigenvectors are sorted by the corresponding descending eigenvalues, this method removes the last 40% of the eigenvectors that contains the least amount of variance among the images [15].

3. Removing the first eigenvector:
   This method removes the first eigenvector with the assumption that information in this eigenvector is affected by lighting conditions and degrades classification [15].

4. Energy Dimension:
   This method uses the minimum number of eigenvectors to guarantee that the energy (e) is greater than a threshold. A typical threshold is 0.9. The accumulated energy of

the i^{th} eigenvector is the ratio of the sum of the first i eigenvalues over the sum of all the eigenvalues [8].

$$e_i = \frac{\sum_{j=1}^{i} \lambda_j}{\sum_{j=1}^{k} \lambda_j} \tag{1}$$

where k is the number of non-zero eigenvalues.

5. Stretching Dimension:
   This method calculates the stretch (s) of an eigenvector. The stretch of the i^{th} eigenvector is the ratio of the i^{th} eigenvalues λ_i over the maximum eigenvalue λ_1 [8]. A common threshold for the stretching dimension is 0.01.

$$s_i = \frac{\lambda_i}{\lambda_1} \tag{2}$$

Removing unnecessary eigenvectors also reduces the computation cost in Eigenfaces. The first and third eigenvector selection methods do not remove the unnecessary eigenvectors that have the least amount of variance amongst the images, and thus have high computation cost. None of these methods are adequate to the variance with a database.

Specifically, the second approach is based on a fact that the dimension for the covariance matrix organized from the training data is very high and even 60% of it is still very large. In order to reduce the dimension of the covariance matrix, a new approach called two dimensional PCA is proposed recently based on different presentation of the training images [21].

2.3 Threshold Selection

Selecting the optimum threshold for facial classification is crucial to the performance of the recognition system [4, 16, 19]. The optimum here indicates that the testing performance can be balanced in terms of precision and recall rates [3]. Each face database requires a unique threshold value. Currently, ad hoc methods are used for the threshold selection and most of them are manual and required more extra testing to validate the selected threshold. Another method to select the classification threshold is by using the Receiver Operating Characteristic (ROC) curves [6, 10, 12]. An ROC graph illustrates the trade off between True Positive (TP) versus False Positive (FP), and it is usually obtained using a subset of the training database. The point on the ROC curve where TP and FP values are high and low respectively is selected as classification threshold. This threshold has shown to perform well [12] but its main shortcoming is found manually and not suitable for changing databases, in particular where new images are added or old ones are taken out.

3 Methodology

In this section, we present a systematic approach for eigenvector and threshold selection by using the information obtained from the training database set.

3.1 Eigenvectors Selection

The motivation of Eigenfaces is to optimally represent high dimensional space with lowest dimensional subspace with good approximation under Euclidean measure criteria. Due to the fact that the number of eigenvalues is a reflection of subspace capacity, we expect to obtain a better approximation subspace via selecting suitable number of eigenvalues. Here, we use the relative errors of eigenvalues to achieve this aim. The relative error usually has a significant change when the eigenvalues vary and this change should be an indicator for the subspace capacity.

The eigenvectors selection algorithm is based on the relative errors obtained from 60% of the eigenvalues. These 60% of eigenvalues correspond to the eigenvectors that contain the greater amount of variance among the images [15]. The relative error of R_k is defined as

$$R_k = \frac{\lambda_k - \lambda_{k+1}}{\lambda_k} \times 100\% \tag{3}$$

where $k = 1, \ldots, N$ and N represents 60% of the eigenvalues for the covariance matrix. With these data, K-means clustering algorithm is then used to cluster $\{R_1, \ldots, R_N\}$ into 2 classes [7]. The class that contains the relative errors corresponding to the larger eigenvalues is then chosen and the index of the chosen class is selected as the number of eigenvectors required to create the eigenspace (M'). All these operations can be done automatically. Here it should be noted that the K-means will produce the required two classes automatically. The performance of the proposed algorithm will be discussed in Section 4.

3.2 Threshold Selection

In order to compute the threshold for Eigenfaces, the intra and inter class information gathered from the training database is used. The intra class (D_i) is a set including the distances between the images of the same individual. This class gives an indication of how similar the images of the same individual are. The other is the inter class (P_i), including the distances between the images of an individual against the images of other individuals. This class indicates how different each image of an individual is when compared to images of other individuals in the database. The individual threshold values (θ_i) are then calculated from each individuals intra and inter class information which will be described below. The minimum of θ_i over all individuals is then defined as the classification threshold (δ) for the database. Usually the training database requires two or more images per individual as greater number of images contributes more distance information, and hence may result in a better estimation for the threshold. The algorithm for classification threshold selection is as follows:

Denote that

I	$=$	number of individuals
K	$=$	number of images per individual

For each image F_{ik}, where $i = 1, \ldots, I$ and $k = 1, \ldots, K$, we compute w for image ik^{th} based on the selected eigenvectors in section 3.1. w is image projection into the eigenspace which is defined in section 3.3.

Then we compute the intra class distances

$$d_{ik}^{ik'} = \|w_{ik} - w_{ik'}\|^2 \tag{4}$$

where $i \in I, k \neq k'$ and $k = 1, \ldots, K$, and the inter class distance

$$p_{ik}^{jl} = \|w_{ik} - w_{jl}\|^2 \tag{5}$$

where $j = 1, \dots, I$, $j \neq i$ and $l = 1, \dots, K$

Now, we sort

$$D_i = \{d_{ik}^{ik'}\}, \quad \text{and} \quad P_i = \{p_{ik}^{jl}\}$$

in ascending order and denote that

$$D_{imax} = \max\{D_i\}, \qquad P_{imin} = \min\{P_i\}$$

Through D_{imax} and P_{imin}, we can define the individual threshold (θ_i) as follows. If $D_{imax} > P_{imin}$, then (θ_i) is defined as:

$$\theta_i = \frac{D_{imax} + P_{imin}}{2} \tag{6}$$

If $D_{imax} < P_{imin}$, then FP will occur. Therefore, we need to find a value between D_{imax} and P_{imin} to balance the TP and FP. As shown in Figure 1, the point on the ROC curve whose derivative is less equal than 1 indicates that small changes in TP lead to significant changes in FP. The point in the ROC curve where its derivative is greater than 1 indicates that small changes in FP values will cause large changes in TP. So, a point to balance TP and FP is one where its derivative should be equal to one.

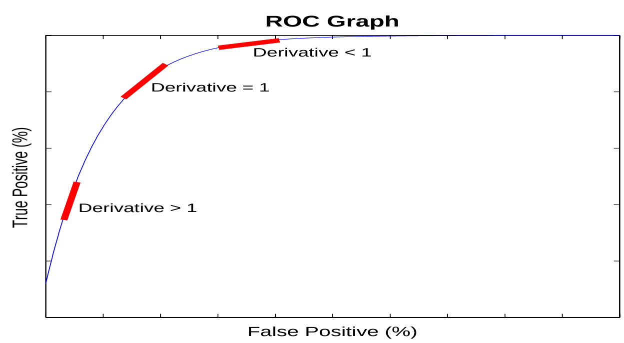

Figure 1: ROC graph

As an illustrative example, consider $D_i = \{a, b, c\}$ and $P_i = \{e, f, g\}$. Now if a is chosen as the threshold, then all elements in P_i and less than a, will be classified within the class and this will be incorrect (FP). All elements in D_i and less than a will be classified correctly and correspond to TP. The number of elements in the intra class depends on the number of images per individual. The higher number of images per individual gives the better approximation of the threshold as more intra class distances can be used to calculate the threshold. This, however, contributes to higher computational cost. To overcome this costive computation issue, only a subset of D_i will be used as possible threshold candidates to measure the TP and FP values in this paper.

In order to find the optimal point on ROC curve to balance TP and FP, its derivative is calculated as:

$$T_x = \left\{ \frac{\Delta TP}{\Delta FP} \times \frac{max(FP)}{100} \right\} \tag{7}$$

where $\Delta TP = TP(x+1) - TP(x)$ and $\Delta FP = FP(x+1) - FP(x)$; If $FP(x) = 0$ or $\Delta FP = 0$ then we define $T_x = 0$

Here the term $\frac{max(FP)}{100}$ is used to stretch the ROC curve evenly. In practice, it is impossible to chose the point on the ROC curve with its derivative being exactly equal to one. Therefore, we need to chose a nearby point as described in the following. Selecting the θ_i as the value in D_i corresponding to the first occurrence of T_x lesser than 1, may result a higher FP. To obtain an optimal performance, the θ_i is set as the value in D_i corresponding to the previous T_x value. If none of the T_x value is less equal than 1, then the θ_i is set as the last value in D_i.

Once θ_i for a training database has been calculated, the testing threshold δ is set as the $\min_i(\theta_i)$.

3.3 Computation Cost Analysis

In this section, we investigate the computational advantage in using the proposed eigenvector selection approach. Let M denote the total number of images in the training data set where each Γ_i is $N \times N$. First, the average image is computed as below

$$\Psi = \frac{1}{M} \sum_{i=1}^{M} \Gamma_i \tag{8}$$

with dimension $N^2 \times 1$. Hence, we construct the normalized image, $\Phi_i = \Gamma_i - \Psi$ and $A = [\Phi_1, \Phi_2, \ldots, \Phi_M]$ with dimension $N^2 \times M$. Let the covariance matrix $B = A^T A$. The eigenspace (u_l) is computed as

$$u_l = \sum_{k=1}^{M} v_{lk} \Phi_k, l = 1, 2, \ldots, M \tag{9}$$

with dimension $N^2 \times M$, where v_{lk} are eigenvectors of B. The projection of an image w_k is

$$w_k = u_k^T (\Gamma - \Psi), k = 1, 2, \ldots, M' \tag{10}$$

The computation cost reduction for an image will be

$$w_k = u_k^T (\Gamma - \Psi), k = M', M'+1, \ldots, M_2 \tag{11}$$

if the number of eigenvectors obtained in Eigenfaces is $M_2 > M'$, and M' is the number of eigenvectors obtained by the proposed approach. This computation cost reduction for an image will be roughly in order $(O(M_2 - M') \times N^2)$. Due to the fact that N is usually very large, if $M_2 - M'$ is large, the computation cost reduction will be significant when the computation reduction is carried over into large database, threshold selection and comparison processes.

4 Experimental Results

4.1 Face Databases

Experiments were carried out on five datasets created from ORL [2] and AMP face database [1] as shown in Table 1. We divided the two databases into five datasets here in order to testify that the proposed approaches for selecting eigenvectors and threshold are working properly under different situations. This is required from the statistical validation point of view if we can not prove the effectiveness theoretically. The ORL face database consists of 400 images compiled of 40 individuals with 10 images each. All the images used were taken against a dark homogeneous background with various lighting conditions, facial expression (open/closed eyes, smiling/non smiling) and face details (glasses/no glasses). The faces are in a frontal upright position with some head tilting, rotation and scaling. The AMP database contains 975 images comprising 13 individuals with 75 images each. The images were captured over a range of different facial expressions, including smiling, frowning, stern, laughing, shocked and neutral. The query effectiveness is evaluated using precision and recall statistics.

Face Data-base	**Data-set**	**# of person**	**# of images /person**	**# of training images**	**# of testing images**
ORL	1	36	5	180	220
	2	36	5	180	220
	3	30	5	150	250
AMP	1	11	15	164	810
	2	11	10	110	865

Table 1: Five datasets created from ORL and AMP face database

4.2 Eigenvetors Selection

The eigenvector selection is based on the relative errors of the eigenvalues. Figure 2 shows the relative errors of the eigenvalues for the first ORL dataset. As illustrated in Figure 2, the class 1 consists of the relative errors corresponding to larger eigenvalues. The index of this class is 17, which is then set as the number of eigenvectors (M'). This value is determined by the K-means clustering algorithm. The performance of the selected (M') value with the proposed threshold method is shown in Figure 3. The precision and recall for the (M') value are 85.24% and 86.40% respectively.

For comparison, we implemented the Energy and Stretching Dimension [20] to select the (M'). The thresholds for these methods were selected based on the selection method in Section 3.2. The comparative results of these methods and the proposed method are shown in Table 2. The performance for the Energy Dimension is 79.64% for precision and 89.87% for recall while the performance for Stretching Dimension is 86.49% in precision and 85.91% in recall. These performance rates are comparable to the proposed algorithm. However, the proposed algorithm uses much fewer eigenvectors and much lower computation cost for comparable classification results as can be seen in Table 2. In implementing this simulation, the first ORL dataset requires 29.22 seconds with the proposed algorithm whist it requires

65.99 seconds for Energy Dimension and 72.03 seconds for Stretching Dimension. The simulation was conducted on AMD Athlon 1.7GHz, 512 Mb DDRRAM running Windows XP with Matlab version 6.5. The time was measured as the total time required for training and testing each dataset.

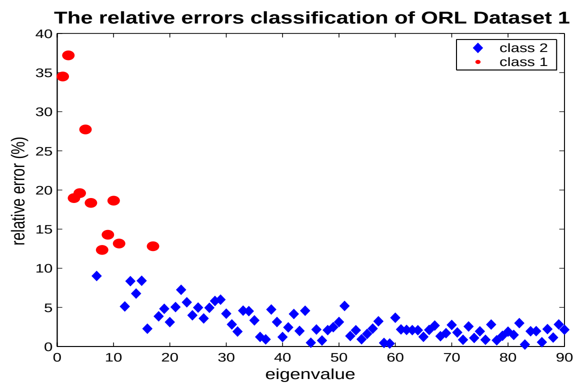

Figure 2: The relative errors classification for the first ORL dataset

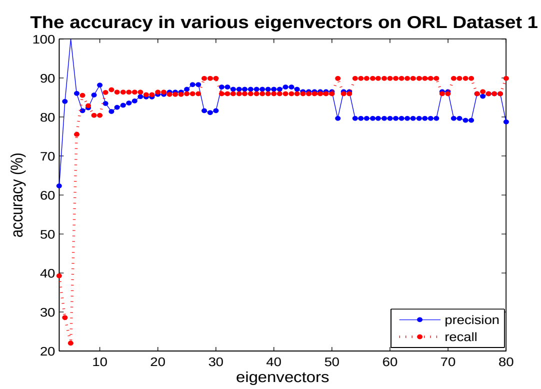

Figure 3: The accuracy in various eigenvectors for the first ORL dataset

4.3 Threshold Selection

The threshold value varies depending on the number of eigenvectors chosen to create the eigenspace. Prior to the threshold selection, the eigenvectors selection is performed to find the suitable number of eigenvectors (M').

The appropriate (M') for the first ORL dataset was selected as 17. Figure 4 shows the individual threshold (θ_i) values for the first ORL dataset derived using our threshold selection criteria. The classification threshold (δ) was chosen as the minimum of the (θ_i) values, which was 1.68.

As shown in Figure 5, the precision and recall rates of the first ORL dataset vary under different threshold values and the (θ_i) value gives recognition with the precision and recall

Method	M′	Precision (%)	Recall (%)	Time (secs)
Proposed Algorithm	17	85.24	86.40	29.22
Energy Dimension	61	79.64	89.87	65.99
Stretching Dimension	70	86.49	85.91	72.03

Table 2: Comparison with other eigenvectors selection methods on the first ORL dataset

rates of 85.24% and 86.4% respectively. The threshold selection has performed significantly well as the precision and recall rates are balanced.

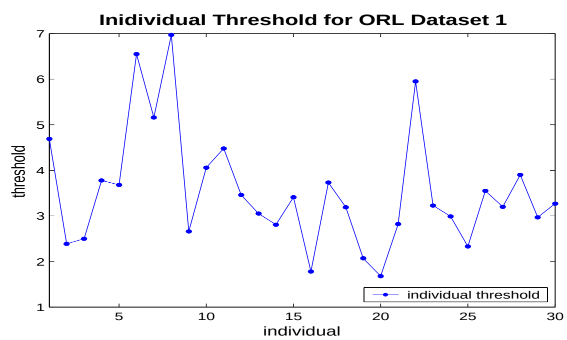

Figure 4: The individual threshold values for the first ORL dataset

Data-set	M′	Threshold	Precision (%)	Recall (%)	Time (secs)
ORL 2	12	1.77	92.41	83.91	27.18
ORL 3	10	2.17	76.47	79.76	34.66
AMP 1	15	0.59	99.85	100	35.19
AMP 2	13	0.44	100	100	29.29

Table 3: The performance with four different datasets

4.4 Evaluation of the Threshold and Eigenvectors Selection on Other Datasets

The performance of the eigenvectors and threshold selection has shown a balanced recognition with the first ORL dataset. In this subsection, we evaluate the performance of the eigenvectors and threshold selection with other four different datasets. The performance results of these datasets are shown in Table 3. The precision and recall of the second ORL dataset are 92.41% and 83.91% respectively. The performance of the third ORL dataset shows 76.47% in precision and 79.76% in recall. The precision and recall for the first AMP

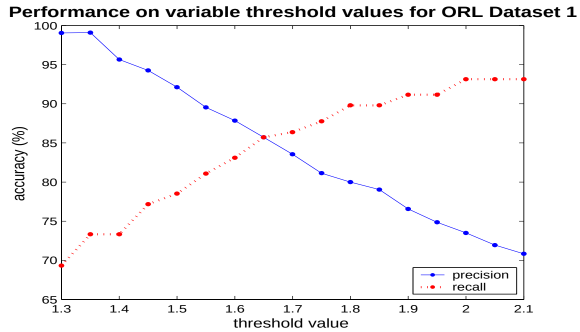

Figure 5: The performance on various threshold values for the first ORL dataset

dataset shows a near optimal with the rate of 99.85% and 100% respectively while for the second AMP dataset, it reaches 100% in both precision and recall. The results in these datasets have shown that our eigenvectors and threshold selection have performed well with much fewer eigenvectors.

The results for the energy and stretching dimension methods in selecting the eigenvectors are shown in Table 4. Both methods achieve comparable precision and recall rates with the proposed eigenvectors selection method. In contrast, the proposed method is able to remove unnecessary eigenvectors in creating the eigenspace and thus in less computational cost when compared to the other eigenvectors selection methods.

Dataset	**Method**	**M′**	**Precision (%)**	**Recall (%)**	**Time (secs)**
ORL 2	Energy Dimension	71	95.48	83.15	80.79
	Stretching Dimension	74	95.48	83.15	83.29
ORL 3	Energy Dimension	66	75.84	81.82	80.17
	Stretching Dimension	66	75.84	81.82	80.40
AMP 1	Energy Dimension	22	99.85	100	41.71
	Stretching Dimension	33	99.85	100	60.15
AMP 2	Energy Dimension	20	100	100	30.28
	Stretching Dimension	33	98.62	100	41.17

Table 4: The results of other eigenvectors selection methods on other datasets

5 Conclusions

This paper has investigated Eigenfaces for some implementing issues. Specifically, we studied the eigenvector selection issue and introduced an alternative approach for selecting appropriate number of eigenvectors. In addition, we proposed a method in selecting the classification threshold using the distance information obtained from the training data set. Experimentation has been conducted with the ORL and AMP face databases. Results have shown that the automatic selection methods in eigenvectors and threshold can produce balanced recognition performance with less computational costs. In addition, the eigenvectors selection method has shown to outperform the energy and stretching dimension methods in terms of computational cost, precision and recall rates. Further, the idea of threshold selection algorithm has been used successfully in our recent paper [17].

References

[1] Advanced Multimedia Processing (AMP Database), [Online] `http://amp.ece.cmu.edu`.

[2] AT&T Laboratories Cambridge (ORL Database), [Online] `http://www.cam-rol.co.uk/facedatabase.html`.

[3] M. Buckland and F. Gey, The relationship between Recall and Precision, *Journal of the American Society for Information Science*, 45(1) (1999) 12–19.

[4] T.E. Campos and R.S. Feris and R.M. Jr. Cesar, Eigenfaces versus eigeneyes: first steps toward performance assessment of representarions for face recognition, *Lecture Notes in Artificial Intelligence*, 1793 (2000) 197–206.

[5] R. Chellappa and C.L. Wilson and S. Sirohey, Human and machine recognition of faces: a survey, *Proceeding of the IEEE*, 83(5) (1995) 705–740.

[6] Z.M. Hafed and M.D. Levine, Face Recognition Using the Discrete Cosine Transform, *International Journal of Computer Vision*, 43(3) (2001) 167–188.

[7] J.A. Hartigan and M.A. Wong, A k-means clustering algorithm, *Applied Statistics*, 28 (1979) 100–108.

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

[9] M. Kirby and L. Sivorich, Application of the Karhunen-Loéve procedure for the characterization of human faces, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1) (1990) 103–108.

[10] J. Kittler and Y. P. Li and J. Matas, On matching scores for LDA-based face verification, in *Proc British Machine Vision Conference BMVC2000*, 2000.

[11] L. Lorente and L. Torres, A global eigen approach for face recognition, in *International Workshop on Very Low Bit-rate Video Coding*, Urbana, Illinois, 1998.

[12] G.L. Marcialis and F. Roli, Fusion of LDA and PCA for face verification, in *Biometric Authentication*, LNCS 2359, 2002, pp. 30–37.

[13] B. Moghaddam, T. Jebara and A. Pentland, Bayesian modeling of facial similarity, in *Advances in Neural Information Processing Systems (NIPS'98)*, 1998, pp. 910-916.

[14] B. Moghaddam and A. Pentland, Face recognition using view-based and modular eigenspaces, in *Automatic Systems for the Identification and Inspection of Humans*, Vol. 2277, SPIE, 1994.

[15] H.J. Moon and P.J. Phillips, Analysis of PCA-based face recognition algorithms, in *EEMCV98*, 1998.

[16] G. Shakhnarovich and J.W. Fisher and T. Darrell, Face recognition from long-term observations, in *Proceedings of the European Conference on Computer Vision*, Vol. 3, 2002, pp. 851–868.

[17] R. Tjahyadi, W.Q. Liu, S. An and V. Svetha, Face recognition based on ordinal approach, in *Proc. of ISSNIP2005*, IEEE, 2005, pp. 349–354.

[18] M. Turk, A random walk through eigenspace, *IEICE Transaction Information and Systems*, E84-D(12) (2001) 1586–1595.

[19] M. Turk and A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, 13(1) (1991) 71–86.

[20] W. S. Yambor, *Analysis of PCA-Based and Fisher Discriminant-Based Image Recognition Algorithms*, 2000, Technical report CS=00=013, Computer Science Department, Colorado State University, Fort Collins.

[21] J. Yang, D. Zhang, A.F. Frangi and J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1) (2004) 131–137.

Manuscript received 4 March 2005
revised 24 October 2005, 24 February 2006
accepted for publication 31 October 2005

Ronny Tjahyadi
Department of Computing, Curtin University of Technology, GPO Box U1987 Perth WA 6845
E-mail address: tjahyadi@cs.curtin.edu.au

Wanquan Liu
Department of Computing, Curtin University of Technology, GPO Box U1987 Perth WA 6845
E-mail address: wanquan@cs.curtin.edu.au

Svetha Venkatesh
Department of Computing, Curtin University of Technology, GPO Box U1987 Perth WA 6845
E-mail address: svetha@cs.curtin.edu.au