New Modelling of Modified Two Dimensional Fisherface Based Feature Extraction

Arif Muntasa

Informatics Engineering Department, Enginering Faculty, University of Trunojoyo, Ry Telang Po. Box 2 Kamal, Bangkalan Corresponding author, email: arifmuntasa@if.trunojoyo.ac.id

Abstract

Biometric researches have been interesting field for many researches included facial recognition. Crucial process of facial recognition is feature extraction. One Dimensional Linear Discriminant Analysis is one of feature extraction method is development of Principal Component Analysis mostly used by researches. But, it has limitation, it can efficiently work when number of training sets greater or equal than number of dimensions of image training set. This limitation has been overcome by using Two Dimensional Linear Discriminant Analysis. However, search value of matrix identity R and L by using Two Dimensional Linear Discriminant Analysis takes high cost, which is $O(n^3)$. In this research, the seeking of "Scatter between Class" and "Scatter within Class" by using Discriminant Analysis without having to find the value of R and L advance are proposed. Time complexity of proposed method is $O(n^2)$. Proposed method has been tested by using AT&T face image database. The experimental results show that maximum recognition rate of proposed method is 100%.

Kata kunci: Biometric, Face Recognition, Principal Component Analysis, Two Dimensional Fisherface

1. Introduction

The biometrics survey results show that facial recognition has big prospect to be developed in advanced. Facial recognition technology has also applied in many sectors, such as banking, government, military and even industrial [1]. Low level feature extraction such as appearance method is the most popular method to extract feature in facial recognition, which is Principal Component Analysis (PCA) [2], [3]. It is simple method to extract facial features by dimensional reduction. This method has been cited by many researches. However, it has limitations. The first, dimensional reduction could not be conducted, when the number of training sets more or equal than image dimension. The last, separation classes could not be efficiently conducted.

Some appearance methods have been proposed to overcome it, which are Linear Discriminant Analysis (LDA) [4], [5], [6], Linear Preserving Projection (LPP) well known as Laplacianfaces [7], and Orthogonal Neighborhood Preserving Projection (ONPP) [8]. These methods have been overcome the last problem of PCA. However, the first problem could not be overcome when number of training set samples more than image dimensionality.

In order to overcome the last problem of PCA, two dimensionality appearance methods has been proposed, such as Two Dimensional - Principal Component Analysis (2D-PCA) [9],[10] and Two Dimensional – Linear Discriminant Analysis (2D-LDA) [11], [12], [13], [14]. Feature extraction without transformation employ on image samples was first proposed by J. Yang [10]. Besides 2D-LDA, further development of 2D-PCA was also proposed by J. Ye, this is Generalized Low Rank Approximations of Matrices (GLRAM) [15]. 2D-LDA was iteratively implemented to find the right (R) and the left (L) sides of the matrix to achieve the optimal matrix projection. Unfortunately, it needs high cost to obtain the L and R values as feature extraction. In this research, modification of 2D-LDA is proposed to obtain the optimal matrix projection, without finding the value of L and R.

2. One Dimensioanl Appearance Method

The most popular of appearance feature extraction method is Principal Component Analysis, well known as PCA [2], [3]. PCA implemented on face is also well known as eigenfaces. Suppose an image is represented by using f(x, y), for each data input f(x,y) is

transformed from two dimensional into one dimensional. An image f(x,y) can be expressed in the matrix form as seen

$$f = \begin{pmatrix} f(1,1) & f(1,2) & \cdots & f(1,w) \\ f(2,1) & f(2,2) & \cdots & f(2,w) \\ \vdots & \vdots & \ddots & \vdots \\ f(h,1) & f(h,2) & \cdots & f(h,w) \end{pmatrix}$$
(1)

In this case, image weight and height are represented by using *h* and *w*. The result of transformation is one dimensional matrix form, which has size (1,n) or (n,1). The symbol of *n* represents the result of multiplication between *h* and *w* ($n=h^*w$). If number of face images used as training sets is *k*, then training sets matrix can be written by using

$$f_{k,n} = \begin{pmatrix} f(1,1) & f(1,2) & \cdots & f(1,n) \\ f(2,1) & f(2,2) & \cdots & f(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ f(k,1) & f(k,2) & \cdots & f(k,n) \end{pmatrix}$$
(2)

PCA method has been developed by many researchers such as Linear Discriminant Analysis (LDA) is well known as Fisherface. The feature of Fisherface can be obtained by using maximized of scatter between class (S_b) and minimized scatter within class (S_w) as seen on equation

$$S_{b} = \frac{1}{N} \sum_{i=1}^{c} (\mu c_{i} - \mu)^{T} * (\mu c_{i} - \mu)$$
(3)

$$S_{w} = \frac{1}{N} \sum_{i=1}^{c} \sum_{X_{j} \in c_{i}} (X_{j} - \mu c_{i})^{T} * (X_{j} - \mu c_{i})$$
(4)

The results of Equation (3) and (4) can be utilized to compute the Eigen value and vector as seen follows

$$S_{B}S_{W}^{-1}w = J(w)w$$

$$S_{B}S_{W}^{-1}w = \lambda w$$
(5)

Eigen value achieved by Equation (5) must be ordered decreasingly $(\lambda_1 \ge \lambda_2 \ge \lambda_2 \ge ... \ge \lambda_k)$ and followed by Eigen vector (Λ). The dimension of Eigen vector (Λ) is *fea*, where *fea* is defined as *k-c*.

However, LDA has crucial problem, it will fail to maximized scatter between class and minimized scatter within class when the discriminatory information is not in the mean of the training sets but rather in the variance of the training sets. If this condition has occurred, than LDA could not efficiently work.

3. Two Dimensional Appearance Method

Basically, problem occurred on LDA can be overcome by using Two Dimensional Discriminant Analysis (2D-LDA) (H. Kong et al., 2005; M. Li, 2005; H. Xiong et al., 2005; Liang Z, 2008). 2D-LDA method is feature extraction based on appearance without transformation from two dimensional into one dimensional matrix vector. However, 2D-LDA has also limitation, which is the running time for feature extraction needs high cost, which is $O(n^3)$. The total

number of training sets is *M*, and for each class *i*-th has M_i training sets. The i-th image training is represented by using A_i and mean i-th image of training sets are $\overline{A}^{(i)}$, whereas \overline{A} represents mean image of all training sets. 2D-LDA can be calculated:

$$S_{W}^{R} = \sum_{i=1}^{c} \left(A_{i} - \overline{A}^{(i)} \right) * R * R^{T} * \left(A_{i} - \overline{A}^{(i)} \right)^{T}$$
(6)

$$S_B^R = M^{(i)} * \sum_{i=1}^c \sum_{A_i \in c_i} (\overline{A}^{(i)} - \overline{A}) * R * R^T * (\overline{A}^{(i)} - \overline{A})^T$$
(7)

The initialization value of R can be defined as identity matrix that has k dimension, where k represents number of eigenfaces used. The results of Equations (20) and (21) can be used to compute the covariance as seen

$$C^{R} = \frac{S^{R}_{B}}{S^{R}_{W}}$$
(8)

The result of Equation (8) can be utilized to calculate the eigenvector. These eigenvectors are used as initialization value of L $\,$

 $L = \Lambda_1^R, \Lambda_2^R, \Lambda_3^R, \dots, \Lambda_m^R$ To compute the value of C^L , it is necessary to calculated first the value of S_W^L and S_R^L

$$S_{W}^{L} = \sum_{i=1}^{c} \left(A_{i} - \overline{A}^{(i)} \right) * L * L^{T} * \left(A_{i} - \overline{A}^{(i)} \right)^{T}$$
(9)

$$S_B^L = M^{(i)} * \sum_{i=1}^c \sum_{A_i \in c_i} (\overline{A}^{(i)} - \overline{A}) * L * L^T * (\overline{A}^{(i)} - \overline{A})^T$$
(10)

And the value of C^{L} can be computed by using

$$C^{L} = \frac{S^{L}_{B}}{S^{L}_{W}} \tag{11}$$

The value of R can be updated by using the eigenvector of Equation (11)

$$R = \Lambda_1^L, \Lambda_2^L, \Lambda_3^L, \cdots, \Lambda_m^L$$
(12)

The values of L and R are used to achieve the projection matrix. However this method has limitation, which is the value of L and R depend on the number of iterations. Time complexity to achieve feature extraction (*L* and *R*) is $O(n^3)$.

4. Proposed Method

Main idea of proposed method is modify of 2D-LDA. The training set is not converting into one dimensional vector. For each class is computed the average of the training set and for all classes are also computed the average of the training sets. The covariance of training set can be computed by using the following equation

$$C = \left(\frac{S_{W}^{2D} * (S_{W}^{2D})^{T}}{S_{B}^{2D} * (S_{B}^{2D})^{T}}\right) * \left(\frac{S_{W}^{2D} * (S_{W}^{2D})^{T}}{S_{B}^{2D} * (S_{B}^{2D})^{T}}\right)^{T}$$
(13)

In this proposed method, the zero mean of training sets has been modified by multiplication S_W^{2D} with its transpose and followed by multiplication S_B^{2D} with its transpose as seen on equation

$$z = \left(\frac{S_W^{2D} * (S_W^{2D})^T}{S_B^{2D} * (S_B^{2D})^T}\right)$$
(14)

In this case, the value of S_W^{2D} and S_B^{2D} can be calculated by using the following equation

$$S_{W}^{2D} = \sum_{i=1}^{c} \left(A_{i} - \overline{A}^{(i)} \right) * \left(A_{i} - \overline{A}^{(i)} \right)^{T}$$
(15)

$$S_B^{2D} = \sum_{i=1}^c M^{(i)} \sum_{A_i \in c_i} (\overline{A}^{(i)} - \overline{A}) * (\overline{A}^{(i)} - \overline{A})^T$$
(16)

 S_W^{2D} and S_B^{2D} represent scatter within class and scatter between class of two dimensional matrix, They can be obtained from the training sets without transformation. Time complexity to obtain the values of S_W^{2D} and S_B^{2D} is $O(n^2)$. The results of Equation (13) are used to compute the eigenvector. The eigenvector is used as the optimal projection matrix, whereas the weight of training set can be achieved by using the following equation

$$W = A_i * \Lambda^{2D} \tag{17}$$

The eigenvector of Equation (13) is used to achieve the weight of the testing sets. The results of Equation (17) is measured the similarity of the weight between the training and testing sets. Feature extractions resulted of proposed method is two dimensional matrixes, the results of feature extraction of training sets has same size with original image. To achieve high recognition rate, it is necessary to chose dominant feature. The most dominant feature has correlation to the largest eigenvalue. If number of feature chosen is d, then number of vector element is d^*h , where h represents image height.

To find out of the testing sets class, it is necessary to measure the weight between the training and the testing set. In this research, four methods were used to measure the similarity, which are Euclidian Distance (D_1) and Manhattan (D_2) as seen in the following equation

$$D_{1} = \sqrt{\sum_{j=1}^{H} \left(W_{j}^{2D-Train} - W^{2D-Test} \right)^{2}}$$
(18)

$$D_{2} = \sum_{j=1}^{h} \left| \left(W_{j}^{2D-Train} - W^{2D-Test} \right) \right|$$
(19)

The final decision of the similarity measurements is the smallest value of the result for each equation.

5. Experimental Results and Analysis

In this research, 400 images have been used as data experiment, for both as the training and the testing sets. These images are made by AT&T - Cambridge University Computer Laboratory, it founded in 1986 as the Olivetti Research Laboratory, better known as ORL. Images are made from 40 people. For each people has been taken 10 times with different poses.



Figure 1. Samples of ORL Face Database

The proposed method has been proved with the facial image reconstruction from ORL face image database as seen in Figure 1.



Figure 2. Face Image Reconstruction By Using Proposed Method

In this paper, 30 features have been used to reconstruct facial image. In Figure 2 on the first row image, from the left to the right describes facial image reconstruction using 1 until 10 features. Facial image reconstruction using 11 until 20 features is described in Figure 2 on the second row image from the left to the right. Similarly on the third row image, facial image reconstruction used 21 until 30 features.

Based on facial image reconstruction, the greater features used, the better facial image reconstruction yielded as seen in Fgure 2. The other hand, it can be said that the greater features used, the smaller error yielded as seen in Figure 3.



Figure 3. Facial Image Reconstruction Error Rate of Proposed Method

In order to evaluate robustness of proposed method, experiments were measured with 2 different similarity measurements, which are Euclidian Distance and Manhattan. For each similarity measurements utilized 2, 3, 4, 5, 6 and 7 training sets as seen in Table 1.

		5		
No	Number of Training	Number of Testing		
	Sets Used	Sets Used		
1	2	8		
2	3	7		
3	4	6		
4	5	5		
5	6	4		
6	7	3		

Table 1. Experimental Model for each Similarity Measurements Method

For each training sets, number of features used are 2, 3, 4, 5, 6, 7, 8, 9, and 10 features as dominant characteristics respectively. The first scenario, Euclidian distance method is used to measure the similarity of the feature extraction. The last scenario the results of feature extraction were measured the similarity by using Manhattan method. The experimental results show that the number of training sets influence recognition rate. Based on Figure 4, it can be shown that the minimum recognition rate for 2 features obtained 84.69%, whereas the maximum recognition rate for 2 features achieved 100%.



Figure 4. The First Scenario: Experimental Results of Proposed Method Using Euclidian Distance

The number of training sets has influence to recognition rate achieved. The more training sets used, the more recognition rate achieved. Similarly, it can be also seen for all features, except experimental results using 4 features. The recognition rate for experimental using 6 training sets is higher than experimental using 7 training sets. The number of features used to measure similarity does not have influence significant to recognition rate. It can be seen in Figure 4. The same phenomena is also occurred, when the similarity measurements using Manhattan as seen in Figure 5. Deviation of the experimental results show that recognition rate of 6 training sets is higher than recognition rate of 7 training sets as seen in Figure 5. The mistake of recognition occurred, because of accessories usage on the testing set. The highest recognition rate occurred, when 2 features is used as parameter for similarity measurement.



Figure 5. The Second Scenario: Experimental Results of Proposed Method Using Manhattan

The experimental results of proposed method have been also compared to other methods, which are PCA, LDA, LPP, Orthogonal Laplacianfaces, Feature Fusion [17], 2D-PCA [16] and 2D-LDA [16]. The Comparison results show that, recognition rate for 5 training sets on the first scenario is 96%, it is superior to other methods expect to 2D-PCA and 2D-LDA. But recognition rate for the second scenario is superior to other method except to 2D-LDA. In this case, maximum recognition rate of our proposed method is 97%, whereas recognition rate of 2D-LDA is 97.33%. For 6 and 7 training set, proposed method is superior to other methods for both the first and the second scenario as seen in Table 2. Proposed method cannot select the best features for each column yet, so increasing features used has given bad effect to recognition results.

Table 2. Experimental Comparison Results				
Method	Number of Training Sets Used			
	5	6	7	
PCA	76.5	81.25	87.5	
LDA	94.4	95.62	98.83	
LPP	83	90.63	92.5	
Orthogonal Laplacianfaces	91	97.5	99.17	
2D-PCA [16]	96.15	97.71	98.87	
2D-LDA [16]	97.33	98.75	99.16	
Feature Fusion [17]	95.5	98.13	99.17	
First Scenario of Proposed method	96	99.38	100	
Second Scenario of Proposed method	97	100	100	

6. Conclusions and Future Research

Based on experimental results and analysis, the highest recognition rate occurred when two vector features used as parameter for similarity measurements, for both the first and the second scenarios. It shows that the more features used, the lower recognition rate achieved. Decreasing of recognition rate is caused of the usage non dominant features.

The more number of training sets used, the higher recognition rate achieved. The highest recognition rate is 100%. It occurred when experiment uses 7 training sets on the first scenario, and 6 and 7 training sets on the second scenario.

Proposed method will be developed to improve their limitations, when number of features used more than 2 features. It is necessary to select the dominant feature vectors for each column. It is conducted to increase the recognition rate and to decrease computation time for similarity measurement.

References

- [1] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips. Face recognition: A literature survey. ACM Comput. Surv. 2003; 35(4): 399-458.
- [2] M. Turk and A. Pentland. Eigenfaces for recognition. J. Cognitive Neuroscience. 1991; 3(1): 71-86.
- [3] A.M. Martinez and A.C. Kak. PCA versus LDA. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001; 23(2): 228-233.
- [4] P.N. Belhumeur, J. Hespanha and D.J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class pecific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997; 19(7): 711-720.
- [5] H.C. Kim, D. Kim and S. Y. Bang. Face recognition using LDA mixture model. Pattern Recognition Letters. 2003; 24: 2815-2821.
- [6] M. Loog and R.P.W. Duin. Linear Dimensionality Reduction via a Heteroscedastic Extension of LDA: The Chernoff Criterion. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004; 26(6): 732-739.
- [7] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang. Face Recognition Using Laplacianfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005; 27(3): 328-340.
- [8] E. Kokiopoulou and Y. Saad. Orthogonal Neighborhood Preserving Projections: A projectionbased dimensionality reduction technique. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007; 29(12): 2134-2156.
- [9] J. Yang and J. Y. Yang. From image vector to matrix: A straightforward image projection technique-IMPCA vs. PCA. Pattern Recognition. 2002; 35(9): 1997-1999.
- [10] J. Yang, Zhang D, Frangi AF, and Yang JY. Two-dimensional PCA: A New Approach to Appearancebased Face Representation and Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004; 26: 131-137.
- [11] H. Kong, L. Wang and E. K. Teoh. A Framework of 2D Fisher Discriminant Analysis: Application to Face Recognition with Small Number of Training Samples. CVPR. 2005.
- [12] M. Li and B. Yuan. 2D-LDA: A novel statistical linear discriminant analysis for image matrix. Pattern Recognition Letter. 2005; 26(5): 527-532.
- [13] H. Xiong, M.N.S. Swamy and M.O. Ahmad. Two-dimensional FLD for face recognition. Pattern Recognition. 2005; 38: 1121 –1124. [14] Liang Z, Li Y, and Shi P. A Note on Two-Dimensional Linear Discriminant Analysis. Pattern
- Recognition. 2008; 9: 2122-2128.
- [15] J. Ye. Generalized Low Rank Approximations of Matrices. Machine Learning. 2005; 61: 167-191.
- [16] Hythem Ahmed, Mohamed Jedra and Nouredine Zahid. Kernel Weighted Scatter-Difference-Based Two Dimensional Discriminant Analyses for Face Recognition. International Journal of Signal Processing, Image Processing and Pattern Recognition. 2012; 5(3).
- [17] Arif Muntasa, Indah Agustien Sirajuddin, and Mauridhi Hery Purnomo. Appearance Global and Local Structural Fusion for Face Image Recognition. Telkomnika Journal. 2011; 9(1): 125-132.