PROCEEDINGS

of the

2018 Symposium on Information Theory and Signal Processing in the Benelux

May 31-1 June, 2018, University of Twente, Enschede, The Netherlands

https://www.utwente.nl/en/eemcs/sitb2018/

Luuk Spreeuwers & Jasper Goseling (Editors)

ISBN 978-90-365-4570-9

The symposium is organized under the auspices of

Werkgemeenschap Informatie- en Communicatietheorie (WIC)

& IEEE Benelux Signal Processing Chapter

and supported by

Gauss Foundation (sponsoring best student paper award)

IEEE Benelux Information Theory Chapter

IEEE Benelux Signal Processing Chapter

Werkgemeenschap Informatie- en Communicatietheorie (WIC)

UNIVERSITY OF TWENTE.

IEEE

The Behavior of Principal Component Analysis and Linear Discriminant Analysis (PCA-LDA) for Face Recognition

Nova Hadi Lestriandoko[1,2] Luuk Spreeuwers[1] Raymond Veldhuis[1]
[1]University of Twente
Faculty of EEMCS, DMB group
Netherlands
[2]Indonesian Institute of Sciences
Research Center for Informatics
Indonesia
[1](n.h.lestriandoko; l.j.spreeuwers; r.n.j.veldhuis)@utwente.nl
[2]nova002@lipi.go.id

Abstract

This paper presents the analysis of PCA-LDA behavior for face recognition using Singular Value Decomposition (SVD). The experimental results is shown to analyze face recognition performance, i.e. the impact of number of subjects, images per subject, training set size, and trade-off between the number of subjects and the number of images per subject on recognition performance, in relation with the number of PCA-LDA coefficients. The comparison of three classifiers, i.e. Euclidean Distance, Cosine Similarity, and Likelihood Ratio, are presented to obtain knowledge about their characteristics. All experimental evaluations are in the verification context. Based on the experimental results, the larger number of subjects and images per subject produced the better recognition performance. Regarding the number of subjects and images per subject trade-off, its indicated both of them influence the recognition performance. Otherwise, the image size also affect to recognition performance. PCA-LDA can perform low resolution image well up to 15x15 pixels and breaks down afterward. Regarding the p and ℓ coefficients, PCA-LDA has different behavior for each classifier.

1 Introduction

Face recognition based on eigenfaces or Principal Component Analysis (PCA) was introduced by M.Turk and A.Pentland [1]. This method reduced the dimensionality by transforming the features from a higher dimensionality space to a lower dimensionality space. The PCA projects face images onto a feature space spanned by the eigenfaces that are the eigenvectors of the covariance matrix of the vector space of face images. The recognition is performed by measuring similarity using classification techniques. Still in the linearly projection area, Fisherfaces or Linear Discriminant Analysis (LDA), introduced by Belhumeur et al[2], maximizes the ratio of the within class and the between class to obtain the best separation of the classes. The dimensionality reduction is done by choosing the most significant coefficients: p largest PCA eigenvalues and l smallest LDA eigenvalues from remaining p.

Commonly, there are many LDA improvements implemented in Biometrics applications. Veldhuis et al.[3] demonstrated the feasibility of hand-geometry recognition based on contour parameters. The integrating LDA with likelihood ratio as classifier was presented in Veldhuis[4] and Spreeuwers et al[5]. Moreover, Sharma et al.[6] proposed a two-stage linear discriminant analysis technique that regularize the between-class scatter and within-class scatter matrices in parallel to produce two orientation matrices, which is concatenated afterward. Still in the context of LDA improvement, Ioeffe [7] presented the probabilistic LDA by modelling both within class and between class variations to solve recognition problems on classes that unseen before. On the other hand, in the context of LDA performance analysis, Zanetti et al.[8] presented the reports on the impact of the number of individuals, the number of images per individual, and trade-off between them to face recognition. However, there are no explanation about PCA-LDA behavior in relation with the number of coefficients.

This paper presents the behavior of PCA-LDA on three classifiers and analyzes the effect of the number of subjects, the number of images per subject, images size, and trade-off between the number of subjects and the number of images per subject on recognition performance. This paper is organized as follows: Section 2 presents the PCA for face recognition. Section 3 continues to LDA and its implementation. Section 4 deals with Similarity Score. The experiments and results are showed in the section 5 and 6, followed by the conclusion at the end of paper.

2 Principle Component Analysis for Face Recognition

PCA is a statistical approach, introduced by M.Turk *et al* [1], used to extract the most relevant features to describe faces. In PCA, every image in the training set is represented as a linear combination of weighted eigenvectors called eigenfaces. PCA can be written as:

$$\mathbf{M_t} \approx \mu_\mathbf{t} + \sum_{i=1}^{k} \mathbf{u_i}\omega_i \tag{1}$$

where k is the number of eigenfaces (eigenvectors) and $k \leq d$. Then, the weight ω_i can be computed easily because of orthonormality as:

$$\omega_j = \mathbf{u_j}^T(\mathbf{M_t} - \mu_\mathbf{t}) \tag{2}$$

where ω_j are weighted features, $\mathbf{u_j}$ are the eigenvectors, $\mathbf{M_t}$ are facial images with dimensionality d, and $\mu_\mathbf{t}$ is the average of the training set.

These eigenvectors are obtained from the covariance matrix of a training set. The weights are obtained after selecting a set of most relevant Eigenfaces. For the verification recognition, the score is obtained by projecting a test image onto the subspace spanned by the eigenfaces and then classification is done by calculating the similarity score.

2.1 Training

Below are the steps to train PCA as feature extraction for face recognition:

1. Prepare the training faces.
   Obtain face images, preprocess them to get centered face in the same size with dimension d.

2. Prepare dataset.
   For every face image in the database, transform into a vector $\mathbf{m_i}$ with size $d \times 1$ and place into a training set $\mathbf{M_t}$.

$$\mathbf{M_t} = \{\mathbf{m_1}, \mathbf{m_2}, \mathbf{m_3}, ..., \mathbf{m_n}\} \tag{3}$$

where n is the number of training samples. So, $\mathbf{M_t}$ is an $d \times n$ matrix.

3. Compute the average of face vector.
   The average of face vector $\mu_\mathbf{t}$ can be calculated by using the following formula:

$$\mu_\mathbf{t} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{m_i} \tag{4}$$

4. Centering the data (subtract the average face vector).
   To get centered data, the face vector is subtracted by the average of face vector. Then, the result is stored in $\mathbf{Z}$.

$$\mathbf{Z_i} = \mathbf{m_i} - \mu_\mathbf{t} \tag{5}$$

   and for matrix $\mathbf{A}$:

$$\mathbf{A} = \{\mathbf{Z_1}, \mathbf{Z_2}, \mathbf{Z_3}, ..., \mathbf{Z_n}\} \tag{6}$$

   Now, the size of $\mathbf{A}$ is $d \times n$.

5. Calculate the covariance matrix.
   The covariance matrix can be obtained by following formula:

$$\mathbf{C} = \frac{1}{n-1} \sum_{i=1}^{n} \mathbf{Z_i}\mathbf{Z_i}^T = \frac{1}{n-1}\mathbf{A}\mathbf{A}^T \tag{7}$$

   Now the size of matrix $\mathbf{C}$ is $d \times d$.

6. Calculate the Eigenvectors and Eigenvalues.
   The matrix $\mathbf{C}$ has size $d \times d$, so it will have d eigenvalues. For this case, the computationally intensive is very hard because of a large dimensional matrix.

   PCA reduces the dimensionality by calculating eigenvectors of matrix $\mathbf{A}^T\mathbf{A}$. Both of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ have the same eigenvalues λ. If $\mathbf{u_i}$ is eigenvector of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{v_i}$ is eigenvector of $\mathbf{A}^T\mathbf{A}$, then the relation of $\mathbf{u_i}$ and $\mathbf{v_i}$ is described in the following equations:

$$\mathbf{A}^T\mathbf{A}\mathbf{v_i} = \lambda_i \mathbf{v_i} \tag{8}$$

   Multiplying both sides by $\mathbf{A}$,

$$\mathbf{A}\mathbf{A}^T\mathbf{A}\mathbf{v_i} = \lambda_i \mathbf{A}\mathbf{v_i} \tag{9}$$

   From this equation, $\mathbf{A}\mathbf{v_i}$ are the eigenvectors of $\mathbf{C} = \mathbf{A}\mathbf{A}^T$ and both of $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ have same eigenvalues λ_i, thus:

$$\mathbf{u_i} = \mathbf{A}\mathbf{v_i} \tag{10}$$

   We can use Singular Value Decomposition (SVD), that is a robust approach, to calculate eigenvalues and eigenvectors. SVD is a decomposition of a real or complex matrix that factorize a matrix into three matrices $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$. The columns of $\mathbf{U}$ and $\mathbf{V}$ are orthonormal and the matrix $\mathbf{S}$ is a diagonal matrix with positive real entries. Both columns of $\mathbf{U}$ and $\mathbf{V}$ form an orthogonal set. The matrices $\mathbf{U}$ and $\mathbf{V}$ are also called left singular vectors and right singular vectors. The SVD theorem states:

$$\mathbf{A}_{nxd} = \mathbf{U}_{nxn}\mathbf{S}_{nxd}\mathbf{V}^T_{dxd} \tag{11}$$

where

$$\mathbf{U}^T\mathbf{U} = \mathbf{I}_{nxn} \tag{12}$$

$$\mathbf{V}^T\mathbf{V} = \mathbf{I}_{dxd} \tag{13}$$

Calculating the SVD consists of finding the eigenvalues and eigenvectors of $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$:

(a) The eigenvectors of $\mathbf{A}^T\mathbf{A}$ make up the columns of $\mathbf{V}$,

(b) The eigenvectors of $\mathbf{AA}^T$ make up the columns of $\mathbf{U}$.

(c) The singular values in $\mathbf{S}$ are square roots of eigenvalues from $\mathbf{AA}^T$ or $\mathbf{A}^T\mathbf{A}$.

The singular values of matrix $\mathbf{S}$ is a diagonal matrix and arranged in descending order. The singular values are always real numbers. If the matrix $\mathbf{A}$ is a real matrix, then $\mathbf{U}$ and $\mathbf{V}$ are also real.

7. Choose only K eigenvectors corresponding to the K largest eigenvalues and project into eigenspace.

2.2 Recognition Procedure

Face recognition can be done by projecting a new facial image onto eigenspace by following formula:

$$\omega_i = \mathbf{u_i}^T(\mathbf{M_{new}} - \mu_\mathbf{t}) \tag{14}$$

where $i = 1, 2, 3, .., K$ and u_i is the eigenvectors corresponding with K largest eigenvalues. The last step of PCA feature extraction is to form feature vector:

$$\mathbf{\Omega_t} = [\omega_1 \quad \omega_2 \quad \omega_3 \quad \dots \quad \omega_K]^T \tag{15}$$

Finally, the recognition is done by calculating the similarity score between two feature vectors and comparing two thresholds.

3 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) based face recognition was introduced by Belhumeur *et al.* [2]. The LDA aim is to find out the best projection of original data matrix on a lower dimensional space by maximizing the ratio of the within class and the between class variances. The technique is to model the space and make it feasible using PCA, then transform the space using LDA in such away that the components are ordered with respective discriminative properties. The feature reduction is done by discarding the least discriminative components. The figure 1 shows how LDA works by projecting the dataset into two rotates axes. Projection to the lower right axis achieves the maximum separation between the categories and projection to the lower left axis yields the worst separation. There are some publications [7][9][10][11] that present the detail tutorial of LDA and its implementation for face recognition. They also described some weaknesses of LDA and introduced their solutions.

PCA-LDA can be implemented using SVD decomposition[3][5][12]. The sub section below will discuss how PCA and LDA produce a transformation matrix to obtain a feature vector.

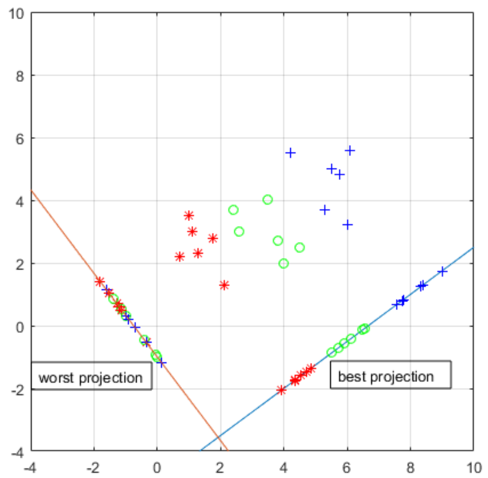

Figure 1: LDA projection

3.1 PCA Transformation

Suppose we have a training set of data consist of n face images. Then, the n sample vectors $\mathbf{m_i}$ of the face images with dimension d are ordered in an $n \times d$ matrix $\mathbf{M}$. If the zero mean of matrix $\mathbf{M}$ is $\mathbf{Z}$ and the mean of $\mathbf{M}$ is $\mu_\mathbf{t}$, then the covariance matrix of $\mathbf{Z}$ is defined by:

$$\mathbf{C_t} = \frac{1}{n-1}\mathbf{Z}\mathbf{Z}^T \tag{16}$$

The SVD decomposition of $\mathbf{Z}$ is:

$$\mathbf{Z} = \mathbf{U_t}\mathbf{S_t}\mathbf{V_t}^T \tag{17}$$

Where $\mathbf{U_t}$ and $\mathbf{V_t}$ are the left and right singular vectors (i.e. the eigenvectors of $\mathbf{Z}\mathbf{Z}^T$ and $\mathbf{Z}^T\mathbf{Z}$ respectively) and unitary matrices (i.e. $\mathbf{U_t}\mathbf{U_t}^T = \mathbf{I}$ and $\mathbf{V_t}\mathbf{V_t}^T = \mathbf{I}$). $\mathbf{S_t}$ is a diagonal matrix with the singular values of $\mathbf{Z}$, which are the square roots of the eigenvalues of $\mathbf{Z}\mathbf{Z}^T$. Then, the equation of $\mathbf{C_t}$ becomes:

$$\mathbf{C_t} = \frac{1}{n-1}\mathbf{Z}\mathbf{Z}^T = \mathbf{U_t}\frac{\mathbf{S_t}\mathbf{S_t}}{n-1}\mathbf{U_t}^T = \mathbf{U_t}\mathbf{\Sigma_t}\mathbf{U_t}^T \tag{18}$$

So, the transformation that whitens the total distribution is:

$$\mathbf{T_1} = \sqrt{n-1}\mathbf{S_t}^{-1}\mathbf{U_t}^T \tag{19}$$

Because:

$$\mathbf{T_1}\mathbf{C_t}\mathbf{T_1}^T = \mathbf{I} \tag{20}$$

The next step is to transform all sample vectors $\mathbf{m_i}$ by $\mathbf{T_1}$.

$$\mathbf{m_i'} = \mathbf{T_1}(\mathbf{m_i} - \mu_\mathbf{t}) \tag{21}$$

3.2 LDA Transformation

We assume that the within distribution of all subjects is normal with the same within class covariance $\mathbf{C_w}$, but different means and that the total distribution of all faces is normally distributed with total covariance $\mathbf{C_t}$. The within class covariance $\mathbf{C_w}$ are then calculated from the transformed data. If there are various classes c in the sample vectors and each class has n_c samples, so the mean of each class can be obtained by summing over the sample vectors of the specic class and dividing by the number of samples of the specic class:

$$\mu_{\mathbf{ci}} = \frac{1}{n_c} \sum_{i:m_i \in c} \mathbf{m'_i} \tag{22}$$

So, the zero mean of vector $\mathbf{m'_i}$ is:

$$\mathbf{z_{ci}} = \mathbf{m'_i} - \mu_{\mathbf{ci}} \tag{23}$$

The within class covariance $\mathbf{C_w}$ can be estimated by ordering the class zero mean vector $\mathbf{z_{ci}}$ into an $n \times d$ matrix $\mathbf{Z_c}$:

$$\mathbf{C_w} = \frac{1}{n-1} \mathbf{Z_c} \mathbf{Z_c}^T \tag{24}$$

Using SVD decomposition, we can obtain:

$$\mathbf{C_w} = \frac{1}{n-1} \mathbf{Z_c} \mathbf{Z_c}^T = \mathbf{U_w} \frac{\mathbf{S_w} \mathbf{S_w}}{n-1} \mathbf{U_w}^T = \mathbf{U_w} \mathbf{\Sigma_w} \mathbf{U_w}^T \tag{25}$$

And the transformation that decorrelates the within distribution is:

$$\mathbf{T_2} = \mathbf{U_w}^T \tag{26}$$

Finally, the total transformation $\mathbf{T}$ is the product of the two transformations:

$$\mathbf{T} = \mathbf{T_2} \mathbf{T_1} = \mathbf{U_w}^T \sqrt{n-1} \mathbf{S_t}^{-1} \mathbf{U_t}^T \tag{27}$$

The best discrimination in LDA is obtained by projecting the vectors on the subspace with the ℓ smallest eigenvalues. If the dimensionality of first transformation ($\mathbf{T_1}$) is reduced to p, then the dimensionality of second transformation ($\mathbf{T_2}$) is reduced to ℓ. This means only the smallest ℓ eigenvalues of the p remaining eigenvalues and corresponding eigenvectors are used resulting the final transformation ($\mathbf{T}$). So, the optimum performance is obtained by seeking the ℓ smallest eigenvalues from the best p largest eigenvalues or we can write:

$$\mathbf{T_{p\ell}} = \mathbf{T_{2\ell}} \mathbf{T_{1p}} = \mathbf{U_{w\ell}}^T \sqrt{n-1} \mathbf{S_{tp}}^{-1} \mathbf{U_{tp}}^T \tag{28}$$

Where $\mathbf{T_{p\ell}}$ is a $d \times \ell$ transformation matrix. Thus, the features can be extracted using this transformation matrix.

4 Similarity Scores

There are so many methods to calculate similarity scores. This paper presents the use of Euclidean Distance, Cosine Similarity, and Likelihood Ratio as classifiers.

4.1 Euclidean Distance

In mathematics, Euclidean Distance is a distance between two points in the straight line. The calculation refers to the old literature as *Pythagorean* metric. So, the distance between two vectors **r** and **x** can be written as:

$$d(\mathbf{r},\mathbf{x}) = \sqrt{\sum_i (r_i - x_i)^2} \tag{29}$$

4.2 Cosine Similarity

The cosine similarity measures the cosine of the angle between two vectors or points. The cosine similarity is defined as:

$$d(\mathbf{r},\mathbf{x}) = \frac{\mathbf{r}^T\mathbf{x}}{||\mathbf{r}||||\mathbf{x}||} \tag{30}$$

4.3 Likelihood Ratio

The likelihood ratio can be regarded as a score and the decision if two facial images are of the same subject is taken by comparing this score to a threshold [3][5]. A simple expression for the log of the likelihood ratio can be calculated by first applying a transformation **T** that de-correlates and scales the total distribution such that it becomes white and simultaneously de-correlates the within distribution. This transformation is obtained using the singular values and vectors of $\mathbf{C_t}$ and $\mathbf{C_w}$. The derivation of the likelihood ratio, see [5], then becomes:

$$LR(\mathbf{x},\mathbf{y}) = \mathbf{x}^T\mathbf{\Lambda}\mathbf{x} + \mathbf{y}^T\mathbf{\Lambda}\mathbf{y} + (\mathbf{x}+\mathbf{y})^T\mathbf{\Gamma}(\mathbf{x}+\mathbf{y}) \tag{31}$$

Where:

$$\mathbf{\Lambda} = \mathbf{I} - \mathbf{\Sigma_w}^{-1} \tag{32}$$

$$\mathbf{\Gamma} = \mathbf{\Sigma_w}^{-1}[2\mathbf{\Sigma_w}^{-1} + \mathbf{\Sigma_b}^{-1}]^{-1}\mathbf{\Sigma_w}^{-1} \tag{33}$$

$$\mathbf{\Sigma_b} = \mathbf{I} - \mathbf{\Sigma_w} \tag{34}$$

All required parameters above are obtained from previous equation, i.e. $\mathbf{\Sigma_w}$ derivation in equation 25, see Linear Discriminant Analysis section.

5 Experiments

The experiments are conducted using FRGC v2 dataset. The FRGC consists of controlled and uncontrolled images. In these experiments, we used only controlled images. Firstly, the registration is applied to align the faces in the FRGC v2 dataset. It is an important preprocessing step for face recognition. The face properties, i.e. eyes, nose, mouth, and face, are detected automatically using Viola-Jones cascade detectors[13]. The registration refers to the eyes position and transform the face image using geometric transformation and linear interpolation. Moreover, the ellipse mask is applied to focus the recognition on face area without hair and ears. Then, the histogram equalization is performed to overcome the illumination problem[14]. The figure 2 shows the result of each face registration step.

The equal-error rate is used as verification rate for performance measurement. The four types of experiments were performed in different conditions and situations:

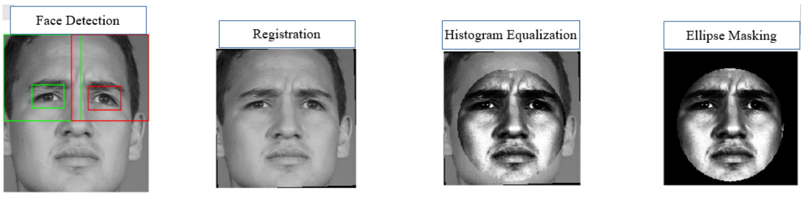

Figure 2: Face Registration

1. The Impact of Number of Subjects to Recognition Performance
   This experiment is used to know the impact of number of subjects to recognition performance, in relation with the behavior on three classifiers: Euclidean Distance, Cosine Similarity, and Likelihood Ratio. The dataset was separated into 3000 samples for training and 500 samples for testing with image size 50x50 pixels. First, we took 500 samples from 25 subjects to train the system. Then, the classifier parameters and the transformation matrix was calculated from training set and applied into testing set. So, the similarity scores were obtained from testing set. The next step was to extend the number of subjects on training set, i.e. 1000, 2000, and 3000 samples. The behavior of PCA-LDA was obtained by analyzing p and ℓ coefficients on three classifiers.

2. The Impact of Number of Images per Subject
   The training set was divided into 4 subsets to analyze the impact of number of images per subject. We took 25 subjects with various numbers of images per subject, i.e. 2, 5, 10, and 20 images per subject respectively. The recognition performance was obtained from testing set with 500 samples from 25 subjects. The likelihood ratio classifier was used in this experiment. Based on the results of experiment 1 on 500 training set and likelihood classifier, the p and ℓ coefficients is defined to 28 and 23.

3. The Impact of Number of Images per subject and Number of Subjects trade off
   The third experiment was conducted to find a trade off between number of images per subject and number of subjects. In this experiment, we used 400 training set from various subjects(i.e. 200, 80, 40, and 20) and various images number per subject(i.e. 2, 5, 10, 20). The p coefficient is defined to 200 and the observation is done along ℓ axis.

4. The Impact of Image Size
   An image contains information as much as its resolution or size. The aim of this experiment is to know the impact of image size, i.e. 200x200, 100x100, 50x50, 40x40, 30x30, 25x25, 20x20, 15x15, 12x12, and 10x10, to recognition performance. The 2000 training set from 100 subjects and the 500 testing set from 25 subjects were used to verify the performance. The PCA-LDA coefficients are defined as: p=100 and ℓ=25.

6 Results

6.1 The Impact of Number of Subjects to Recognition Performance

The results showed a valley, as an optimal area, along ℓ coefficient axis at $\ell \leq s-1$ on Euclidean and Cosine classifier. In this case, s represents the number of subjects. The valley became smaller for bigger training set and disappeared for training set more

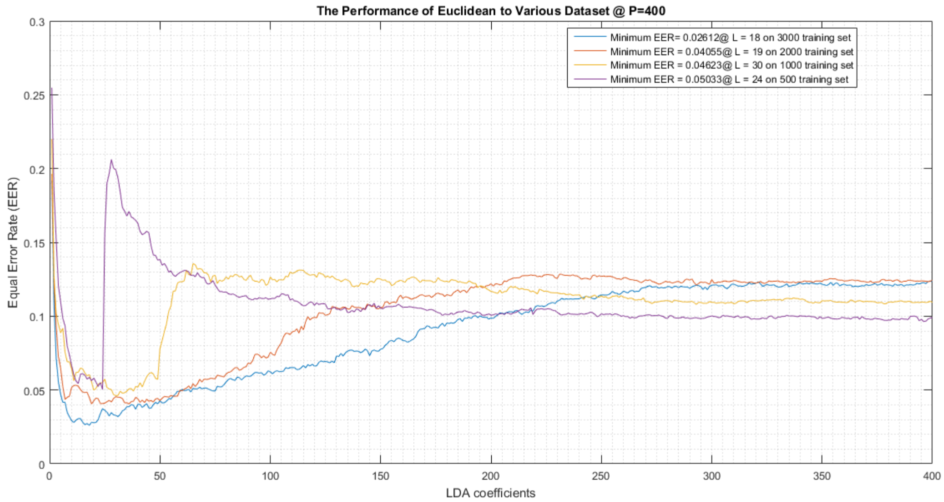

Figure 3: LDA coefficients using Euclidean Distance on various dataset at p=400

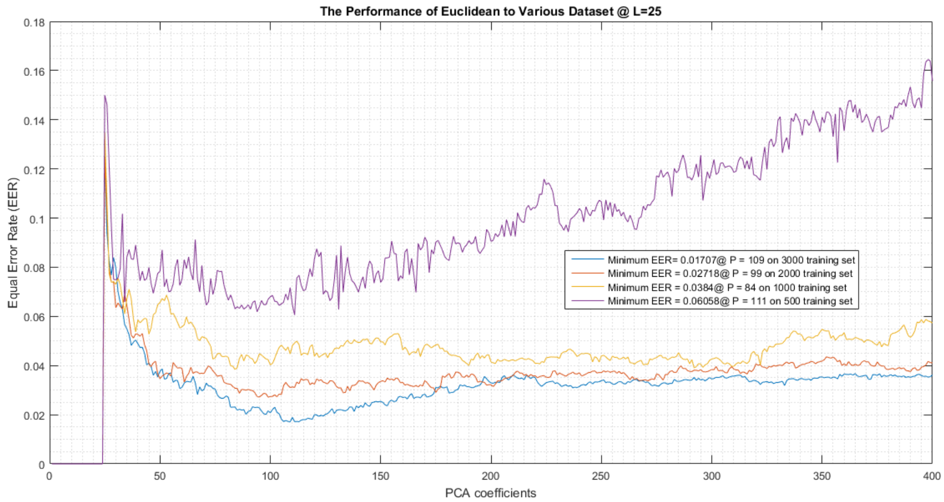

Figure 4: LDA coefficients using Euclidean Distance on various dataset at ℓ=25

than 2000. Similarly, PCA-LDA with likelihood ratio has optimal area at $\ell \leq s-1$. After that point, the performance line becomes flat or there are no contribution to recognition. It is happened because PCA-LDA maximizes the ratio of the between class and the within class which has the optimum dimension at $d \leq (p, s-1)$ [3]. The detail explanation for each classifier can be seen in the sub section below.

6.1.1 Euclidean Distance

Based on the performance observation on all parameters, we obtained the behavior of PCA-LDA on Euclidean Distance classifier. The results showed a valley with a bump at LDA coefficient $\ell = s-1$. figure 3 showed the bump position at ℓ coefficient on dataset with various number of subjects. The bump size is vice versa with the number of subjects. The bigger the number of subject, the smaller the bump size. But, the

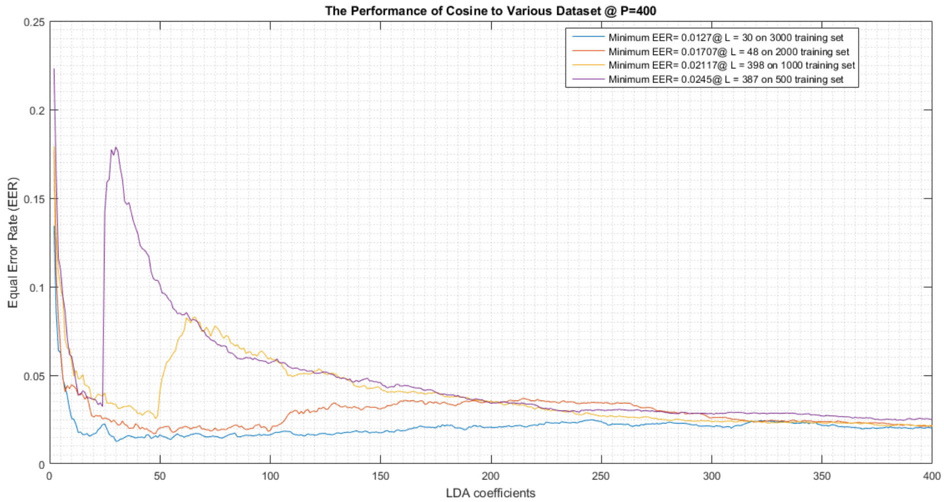

Figure 5: LDA coefficients using Cosine Similarity on various dataset at p=400

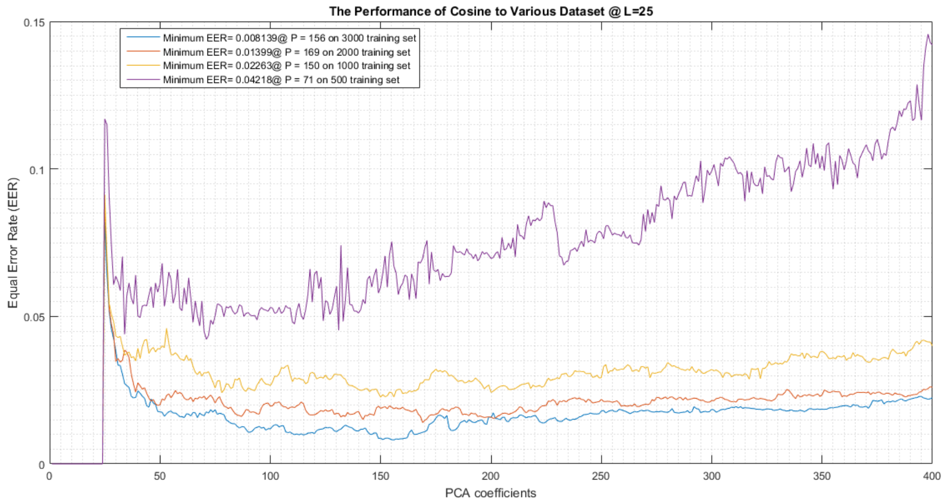

Figure 6: LDA coefficients using Cosine Similarity on various dataset at ℓ=25

bump disappeared on 3000 training set. It means that we have to provide minimum 2000 training set to get a representative training set. However, the optimal performance was always in the range of ℓ =10 to 50.

The next observation for p coefficient, we took ℓ=25 and observed the performance for all p coefficients. As shown in figure 4, the graphics showed that the trend of optimal performance was in the range p=80 to 120. The number of subjects also gave an impact to recognition performance. The 500 training set produced high fluctuation and the worst performance. Otherwise, the larger the training set, the better and the more stable the performance.

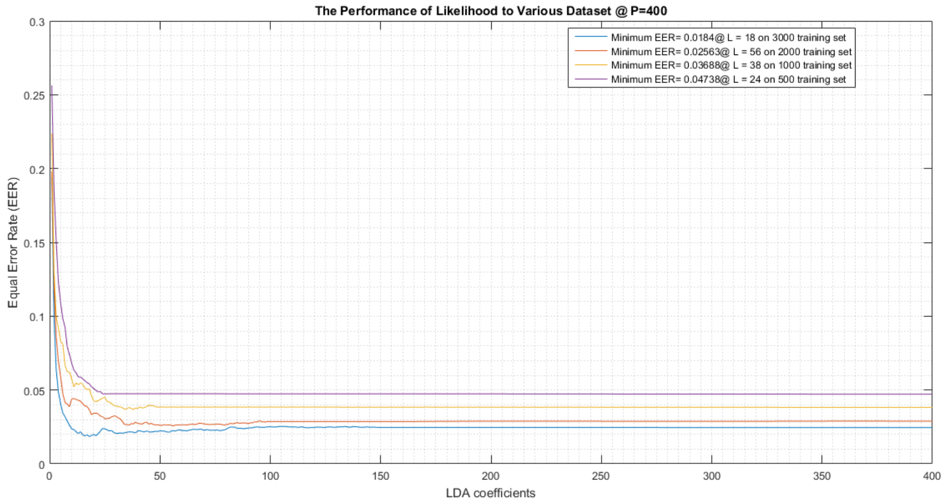

Figure 7: LDA coefficients using Likelihood Ratio on various dataset at p=400

6.1.2 Cosine Similarity

Similar with the Euclidean Distance, The LDA coefficient observation showed that Cosine has a bump at LDA coefficient $\ell = s - 1$ and the bump size becomes smaller on bigger dataset. Whereas, the ℓ observation at p=400 is shown in the figure 5. The optimal area was in the range ℓ=10 to 100. But, the trend on all dataset went down to the minimum EER after the bump. For example, the minimum EER on 500 dataset was reached at ℓ=387 and 1000 dataset at ℓ=398. It was possible for all dataset to have lower EER at higher ℓ coefficient, because the limitation of our observation was 400 coefficients.

Regarding the observation on p coefficient, as shown in the figure 6, the optimal performance for Cosine was reached at p=70 to 200. The graphics showed the performance of Cosine on all p coefficient at ℓ=25. The trends for all dataset were similar, except for the dataset 500. It had high fluctuation and the EER raised rapidly.

6.1.3 Likelihood Ratio

The observation of LDA performance on Likelihood classifier can be seen in the figures 7 and 8. Figure 7 showed that the ℓ coefficients become flat at $\ell > s - 1$. As the proof in [3], a higher dimensionality $\ell > subject - 1$ did not contribute to the Likelihood ratio. So, the optimal area of Likelihood ratio was in the dimension $d \leq (p, s - 1)$. On the other hand, the optimal performance on p observation, as shown in figure 8, was in the range p=50 to 200 for dataset larger than 1000.

6.2 The Impact of Number of Images per Subject

Table 1 showed the impact of images per subject to recognition performance. The EER was growing down while the number of images per subject was increased for all classifiers. It means that the addition of number of images per subject will produce the better performance. However, the different of EER between them becomes smaller with double increase in the number of images. So, the more addition of images per subject will not give contribution significantly to recognition, but only give the risk of slower computation.

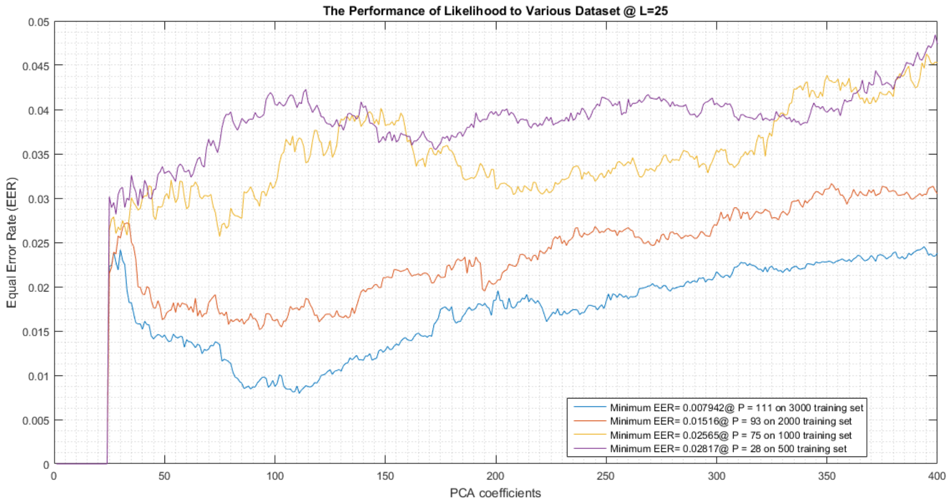

Figure 8: LDA coefficients using Likelihood Ratio on various dataset at ℓ=25

Table 1: Impact of Images per subject

Subjects	Images per	Equal Error Rate(EER)		
	Subject	Likelihood	Cosine	Euclidean
25	2	0.2890	0.1724	0.1946
25	5	0.1113	0.1512	0.1751
25	10	0.1019	0.1264	0.1610
25	20	0.0974	0.1224	0.1495

6.3 The Impact of Number of Images per subject and Number of Subjects trade off

Table 2: Number of images per subject and number of subjects trade off

Subjects	Images per	Minimum Equal Error Rate(EER)		
	Subject	Likelihood	Cosine	Euclidean
200	2	0.3567	0.0573	0.172
80	5	0.0969	0.0556	0.112
40	10	0.0858	0.0504	0.101
20	20	0.0988	0.0758	0.112

The analysis of recognition performance on ℓ coefficients at $p = 200$ was shown in the figures 9,10, and 11. The optimal performance on three classifiers indicated the same characteristics: the 200x2 training set produced the worst performance, the 80x5 and 20x20 training set produced the similar performance, and the 40x10 yielded the best performance. The minimum Equal Error Rate on three classifiers, as shown in the table 2, also indicated that the best combination was the training set from 40 subjects with 10 images per subject. So, both of the number of subjects and the number of images per subject give contribution to recognition.

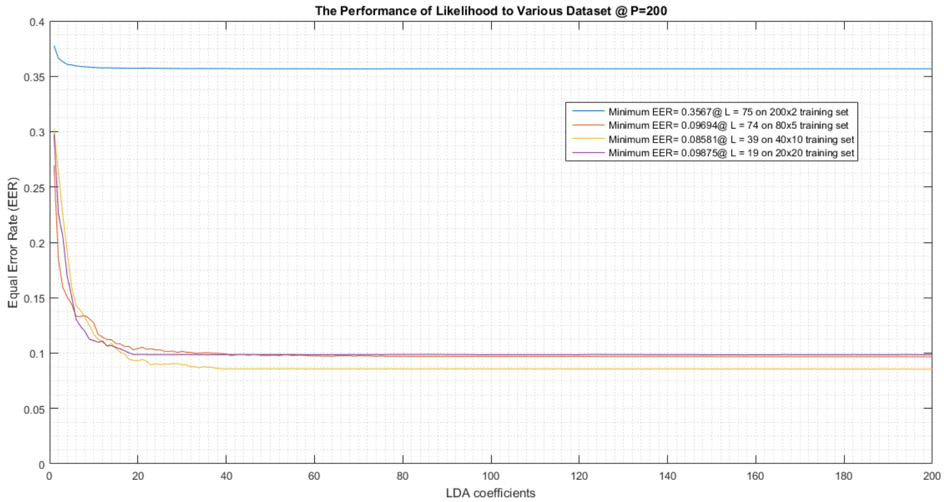

Figure 9: The impact of number of images per subject and number of subjects trade off on Likelihood classifier

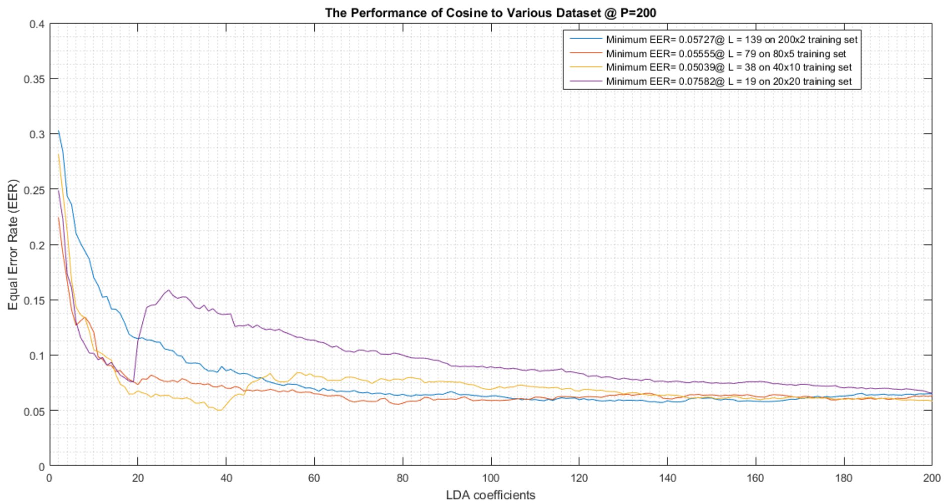

Figure 10: The impact of number of images per subject and number of subjects trade off on Cosine classifier

6.4 The Impact of Image Size

The results of this experiment were shown in the table 3. They showed that PCA-LDA can perform face recognition on very low resolution images up to 15x15 pixels. But, the performance will break down for the resolution less than 15x15 pixels. It was caused by the less information in the very small image. So, the PCA-LDA can not discriminate it well.

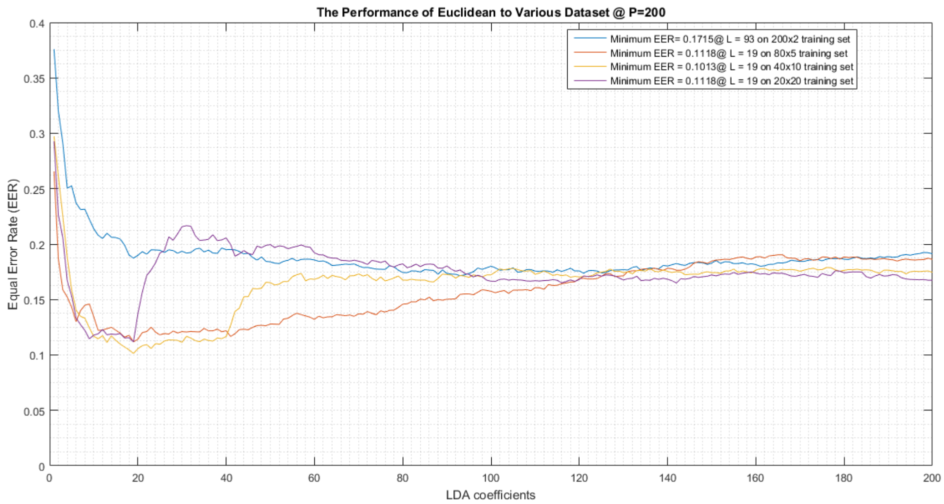

Figure 11: The impact of number of images per subject and number of subjects trade off on Euclidean classifier

Table 3: The Impact of Image Size

Image Size or	Equal Error Rate(EER)		
Resolution	Likelihood	Cosine	Euclidean
200 x 200	0.0178	0.0190	0.0324
100 x 100	0.0181	0.0198	0.0355
50 x 50	0.0176	0.0212	0.0295
40 x 40	0.0157	0.0209	0.0306
30 x 30	0.0166	0.0202	0.0324
25 x 25	0.0142	0.0177	0.0298
20 x 20	0.0163	0.0164	0.0372
15 x 15	0.0184	0.0165	0.0379
12 x 12	0.0259	0.0219	0.0443
10 x 10	0.1401	0.1470	0.1432

Conclusion

The behaviors of PCA-LDA on three classifiers were presented in this paper. For all classifier, the larger number of subjects will give the better performance. The number of images per subject also produced the performance as well as the number of subjects. The higher number of images per subject the better performance we obtained. Regarding the number of images per subject and number of subjects trade off, it indicated that both of them have contribution to recognition.

Regarding the p and ℓ coefficients, PCA-LDA has different behavior for each classifier. Firstly, Likelihood ratio had optimal performance at dimension $d \leq (p, s-1)$ and p in the range 50 to 200 for large training set, i.e. 2000 samples or more than 2000 samples. For the small training set,i.e. less than 2000, p coefficient was close to $s-1$ with $p > s-1$. The next classifiers, Euclidean and Cosine, have a bump at $\ell = s-1$ and form a valley at $\ell < s-1$ as optimal areas. The Cosine has optimal areas at ℓ= 10 to 100. The trend of cosine line is growing down after the bump and it is possible to reach the minimum EER at higher ℓ coefficients. The Euclidean has different behavior with Cosine, the ℓ coefficients have optimal area at 10 to 50. ℓ were saturated after the bump for $\ell > s - 1$. Otherwise, the optimal range of p coefficients is p=80 to 120 for Euclidean and p=70 to 200 for Cosine classifier.

The PCA-LDA also has good performance on low resolution images. It can perform well up to 15x15 pixels resolution, but the performance will break down afterward.

Acknowledgements

The research described in this paper was supported by Research and Innovation in Science and Technology Project (RISET-Pro) of Ministry of Research, Technology, and Higher Education of Republic Indonesia (World Bank Loan No.8245-ID).

References

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991, pMID: 23964806. [Online]. Available: https://doi.org/10.1162/jocn.1991.3.1.71

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, Jul 1997.

[3] R. Veldhuis, A. Bazen, W. Booij, and A. Hendrikse, *Hand-Geometry Recognition Based on Contour Parameters*. SPIE – The Int. Society for Optical Engineering, 3 2005, pp. 344–353, imported from DIES.

[4] R. Veldhuis and A. Bazen, *One-to-template and ono-to-one verification in the single- and multi-user case*. Netherlands: Werkgemeenschap voor Informatie- en Communicatietheorie (WIC), 5 2005, pp. 39–46, imported from DIES.

[5] L. Spreeuwers, "Derivation of lda log likelihood ratio one-to-one classifier," vol. 2014, no. 1, p. 5, 2014.

[6] A. Sharma and K. K. Paliwal, "A two-stage linear discriminant analysis for face-recognition," *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1157 – 1162, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865512000335

[7] S. Ioffe, "Probabilistic linear discriminant analysis," in *Computer Vision – ECCV 2006*, A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 531–542.

[8] N. Zanetti, A. Kumar, L. V. Huynh, and C. N. Teixidor, "Face recognition performance analysis: Impact of number of training samples and trade-off between number of individuals and images per individual," *University of Twente Students Journal of Biometrics and Computer Vision*, 2015. [Online]. Available: https://ojs.utwente.nl/index.php/UTSjBCV/article/view/4

[9] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," vol. 30, pp. 169–190,, 05 2017.

[10] F. Z. Chelali, A. Djeradi, and R. Djeradi, "Linear discriminant analysis for face recognition," in *2009 International Conference on Multimedia Computing and Systems*, April 2009, pp. 1–10.

[11] C. Zhang, Q. Ruan, and X. Pan, “Local fisher discriminant embedding for face recognition,” in *2008 9th International Conference on Signal Processing*, Oct 2008, pp. 1660–1663.

[12] L. Spreeuwers, R. Veldhuis, S. Sultanali, and J. Diephuis, *Fixed FAR Vote Fusion of regional Facial Classifiers.* Gesellschaft fr Informatik, 9 2014, pp. 1–4.

[13] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. I–511–I–518 vol.1.

[14] C. Tosik, A. Eleyan, and M. S. Salman, “Illumination invariant face recognition system,” in *2013 21st Signal Processing and Communications Applications Conference (SIU)*, April 2013, pp. 1–4.