

## Face Recognition Using Holistic Features and Linear Discriminant Analysis Simplification

I Gede Pasek Suta Wijaya<sup>\*1</sup>, Keiichi Uchimura<sup>2</sup>, and Gou Koutaki<sup>2</sup>

<sup>1</sup>Electrical Engineering Dept., Faculty of Engineering, Mataram University  
Jl. Majapahit 62 Mataram, Nusa Tenggara Barat, Indonesia

<sup>2</sup>Graduate School of Science and Technology, Kumamoto University  
Kumamoto Shi, Kurokami 2-39-1, Japan

e-mail: gpsutawijaya@gmail.com<sup>\*1</sup>, {uchimura, koutaki}@cs.kumamoto-u.ac.jp<sup>2</sup>

### Abstrak

Pada makalah ini, suatu metode yang berbasis ciri-ciri holistic (menyeluruh) dari citra wajah and LDA (linear discriminant analysis) yang disederhanakan diusulkan sebagai algoritma alternative dari pengenalan wajah. Metode ini diusulkan untuk mengatasi masalah waktu yang dibutuhkan dalam proses pelatihan kembali dari metode LDA konvensional ketika data baru diregistrasi ke dalam data pelatihan. Ciri-ciri holistic (holistic features) wajah diusulkan untuk mengurangi dimensi ukuran data citra wajah. Sedangkan LDA yang disederhanakan (simplified LDA) yang merupakan suatu penentuan kembali scatter antar klas menggunakan suatu konstanta rerata global untuk mengurangi kompleksitas waktu dari proses pelatihan kembali. Untuk mengetahui keberhasilan metode ini, beberapa percobaan dilakukan menggunakan beberapa database wajah yang memiliki kerumitan yang tinggi seperti database ORL, YALE, ITS-LAB, INDIA dan FERET. Beberapa variasi metode dari PCA dan LDA seperti DLDA, 2DLDA, (2D)<sup>2</sup>DLDA, 2DPCA, dan (2D)<sup>2</sup>2DPCA digunakan sebagai metode pembandingan. Data percobaan menunjukkan bahwa metode yang diusulkan dapat menyelesaikan permasalahan pelatihan kembali yang ditunjukkan dengan waktu pelatihan yang pendek dan tingkat pengenalan yang stabil.

**Kata kunci:** PCA, LDA, ciri-ciri holistic, ekstraksi ciri-ciri, pengenalan wajah.

### Abstract

This paper proposes an alternative approach to face recognition algorithm that is based on global/holistic features of face image and simplified linear discriminant analysis (LDA). The proposed method can overcome main problems of the conventional LDA in terms of large processing time for retraining when a new class data is registered into the training data set. The holistic features of face image are proposed as dimensional reduction of raw face image. While, the simplified LDA which is the redefinition of between class scatter using constant global mean assignment is proposed to decrease time complexity of retraining process. To know the performance of the proposed method, several experiments were performed using several challenging face databases: ORL, YALE, ITS-Lab, INDIA, and FERET database. Furthermore, we compared the developed algorithm experimental results to the best traditional subspace methods such as DLDA, 2DLDA, (2D)<sup>2</sup>DLDA, 2DPCA, and (2D)<sup>2</sup>2DPCA. The experimental results show that the proposed method can be solve the retraining problem of the conventional LDA indicated by requiring shorted retraining time and stable recognition rate.

**Keywords:** PCA, LDA, holistic features, features extraction, face recognition

Copyright © 2012 Universitas Ahmad Dahlan. All rights reserved.

### 1. Introduction

The published methods of face recognition can be categorized in to three groups[1]. Firstly holistic matching method, which uses the whole face region as raw input to the recognition system; secondly features based (structural) matching methods which use the local features such as eyes, noses, and mouth, and local statistics (geometrics and or appearance) as data input to the recognition system; and thirdly hybrid methods which use both whole face

and local features as data input to the system. The method that involves to the first category: principle component analysis (PCA), linear component analysis (LDA), probabilistic decisions based neural network, and their variations or combinations. The methods included in the second category are Dynamic link architecture, hidden Markov model (HMM), and convolution neural network (CNN). The methods included in the third groups are, modular eigenfaces, hybrid local feature method, flexible appearance model, and Face regions and components. Among them, the PCA-based and LDA-based methods of face recognition are well known and encouraging results have been achieved. However, both of them have their limitations: large computational costs, high memory space requirement, and retraining problems. Those systems have to retrain all face image classes to get optimal projection when the new classes are added into the system. In addition, the main disadvantage of the PCA projected features is lack of discrimination power because it removes the null spaces of data scatter that have useful discriminant information. The LDA works with assumption that the data samples have Gaussian distribution.

This paper proposes an alternative approach to face recognition algorithm that is based on holistic features (HF) of face image and simplified LDA (sLDA) to overcome large computational costs and retraining problem of the conventional LDA (CLDA). There are three main objectives of the proposed method: firstly, to prove that the global features of face image contains most of face classification information; secondly, to redefine the predictive between class scatter ( $S_b$ ) which have the same structure as that of CLDA but it has less computational complexity; and finally, to know the effectiveness of sLDA for face recognition compared with the best traditional subspace methods.

This paper is organized as follows: section 2 describes the previous work which related to the proposed system; section 3 explains how to implement the DCT analysis as a main element of global feature extraction, section 4 explains briefly the algorithm of classical LDA, the simplified LDA, and their advantages and disadvantages, section 5 describes the implementation for face recognition methods, section 6 presents the experimental results and the results discussion, and the rest concludes the paper.

## 2. Previous Work

The related works to our approach are the face recognition based on holistic or global matching methods as described in Refs. [1-15]. Ref. [2] describes the evaluations of large-scale Chinese face database, which consisting of 99594 images of 1040 subjects that have large diversity of the variations, including pose, expression, accessory, lighting, time, backgrounds, distance of subjects to camera, and the combined variations, using several well know methods: PCA, PCA+LDA, Gabor-PCA+LDA, and Local Gabor Binary Pattern Histogram Sequence (LGBPHS). The evaluation results shows that the algorithm based on LGBPHS outperform over other three algorithms.

In terms of frequency based methods analysis, Ref. [3] describes face recognition based on a combination of the DCT analysis and face localization techniques to acquire the global information of face image, but it requires eyes coordinate which have to input manually, to perform geometrical normalization. The global face information was created by keeping a small number of the DCT coefficients that had large magnitude values. Ref. [4] describes face recognition based on wavelet-packet tree analysis for frontal view of human faces under roughly constant illumination. The facial feature was built by implementing wavelet-packet tree analysis of bounding box face and then calculating the mean and variance of sixteen matrices wavelet coefficients. That approach does not work for non-frontal faces view and needs constant illumination to make the bounding box face. Furthermore, the combination of the DCT or the DWT with simple modified LDA, APCA or weight LDA had been implemented to recognize pose invariant face image and gave better performances than stand alone PCA or LDA, as described in Refs [12-14]. Moreover, Ref. [5] proved that the PCA and LDA could be directly implemented in DCT domain with the same result as the one obtained from spatial domains. However, they still lack of the recognition rate. Even though, the Ref. [12] could solve the retraining problem of LDA, however its performance will decrease significantly for large sample size.

The most related approach to our system is face recognition based on PCA and/or LDA and their variations as described in Refs. [5-14]. In term of data dimensional, the PCA and LDA can be classified in to one-dimensional (1D-PCA/LDA) [5,12-14] which is vector-based analysis

and two-dimensional (2D-PCA/LDA) [6-11] which is matrix-based analysis. The 2D-PCA outperforms over the 1D-PCA. However, the 2D-PCA requires more coefficients for image representation than 1D-PCA. The two-directional 2D-PCA [8], which works in both the row and column direction of image has been proposed in order to achieve higher and more stable accuracy as well as fewer coefficients is required for image representation. However, they still have retraining problem.

Like 1D-PCA, the 1D-LDA is essentially working on vector based for features clustering. Generally, the 1D-LDA provides better performance over the PCA, because the LDA discriminate the data using both between class scatter and within class scatter information. The Ref. [16] shows that the LDA provides higher discrimination power than the PCA and the classification information of the PCA spread to all over to principle components while the LDA's concentrate in top few discriminant vectors. However, the 1D-LDA has several difficulties: computational cost and singularity problems. Due to those problems, direct LDA has been proposed, as described in Ref. [9]. In order to keep the two-dimensional structure of face image the 2D-PCA and 2D-LDA has been developed which the 2D-LDA gave better performance than the 2D-PCA. In order to get more compact features, the two-directional and two-dimensional PCA ((2D)<sup>2</sup>PCA)[8] and PCALDA ((2D)<sup>2</sup>PCALDA)[11] have been developed. The (2D)<sup>2</sup>PCALDA out performs over the others algorithms. Therefore, we will compare our proposed methods to the (2D)<sup>2</sup>PCALDA methods. The strength of 1D-PCA and 1D-LDA for face recognition with DCT based holistic features as raw input on the system has been presented in Refs. [13-14]. Moreover, we also developed an alternative 1D-PCA [14] that improve the global mean dependent of the covariance matrix and weight 1D-LDA[13] that improve the class separable. However, the LDA and its variations as described previously still have retraining problem because the between class scatter depend on the global mean. The Ref. [15] proposed a new formulation of scatter matrices to extend the two-class non-parametric discriminant analysis and multi-classifier integration to address Gaussian distribution assumption of the LDA-based methods. However, the proposed formulation of non-parametric between class scatter matrices required larger time computational cost than that of the conventional LDA (by about three times of the LDA-based).

To address large computational cost and retraining problems, we propose a new formulation of  $S_b$  that is based on constant global means. The proposed formulation of  $S_b$  has the same characteristic as the original one in terms of its symmetric.

### 3. Holistic Features Extraction

In order to get a global/holistic compact feature of face image called as holistic features (HF), we apply frequency and moment analysis as pre-processing face image. From this processing, we keep a small size of dominant frequency contents and invariant moment coefficients as HF. The frequency analysis (i.e. DCT), which has good energy compactness, is utilized to obtain the dominant frequency content [3-5]. In addition, the DCT decomposes the entire face image without geometrical normalization, bounding box, and blocking processing. From the DCT decomposition output, the dominant frequency content is created by three steps: firstly, convert the DCT coefficients to a vector using row ordering technique; secondly, sort the vector descending using quick sort algorithm, and finally truncate  $m$  first vector elements (i.e., less than 100 elements). Those processes are performed on both training and query (probe) face images. However, in the training process, those are performed one time.

Consider the dominant frequency content, if they are reconstructed into the face images, the reconstructed face images will be different. However, we can still understand that they are the face images, as shown in Figure 1(a). It means that the dominant frequency content existing in low-frequency components is sufficient for face image representation. In other word, if an image is transformed to the frequency domain and then the high frequency components are removed, the reconstructed image will lose a little significant information. Furthermore, if the difference between the reconstructed image and the original is determined by root mean square error (RMSE), we will get results as shown in Table 1. It shows that the dominant frequency of DCT-based tends to contain most information of face image than that of the wavelet (DWT)-based.

In order to get robust HF of any face pose variations, the moment information that provides invariant measure of face images shape is considered. The moment information is

obtained using invariant moment analysis, which is derived from central moment analysis[17]. The invariant moment set is invariant to translation, scale change, and rotation; therefore, this concept can be used to get the holistic information of any face pose variations. Next, the invariant moment analysis was determined just in the intensity component of color images. Finally, from both the selected frequency coefficients ( $f$ ) and the moment invariant set ( $g$ ), we construct the HF vector,  $x_i = [f_i^{DCT}, g_i]$ , where  $i$  is  $i$ -th class of face image. The dimensional of  $x$  is  $m+n$ , where  $m$  is number of selected frequency coefficients and  $n$  is number of selected invariant moments.

Table 1. The RMSE comparison of reconstructed image

Freq. Analysis	RMSE as function features size						
	16	25	36	49	64	81	100
DCT	2796.2	2581.2	1120.1	1007.8	847.3	1317.7	1190.1
DWT	6280.2	3964.0	3301.6	3300.9	1874.6	1468.2	1257.4

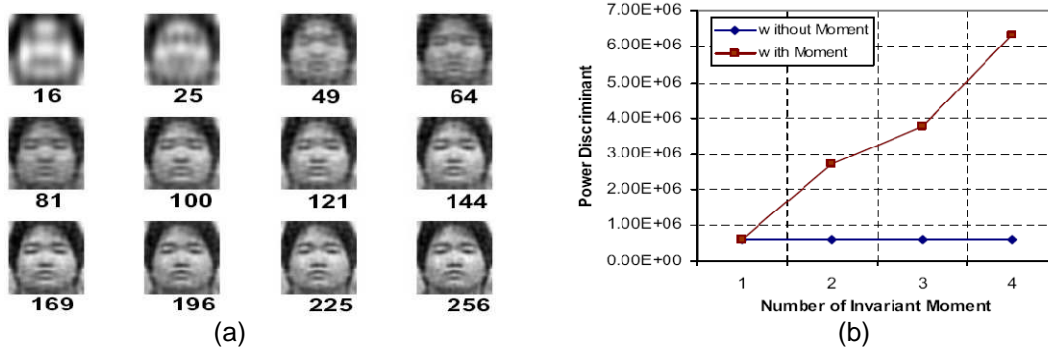


Figure 1. (a) The reconstructed images of DCT-based with size from 16 (top-left) to 256 (bottom-right) elements and (b) the discrimination power as function number of invariant moment

The strength of the proposed HF: it is such a compact size that have higher discrimination power than that of without moment information, as shown in Figure 1(b) and it has almost similar features of small pose variations of face image which has been proved in Ref. [12]. It means that the HF can overcome large computational cost of the CPCA and CLDA based face recognition algorithms and the large variability of pose in single face. In this case, the discrimination power was determined by Eq. (11) and the invariant moments, which are higher than 4, are not included because they make the within class scatter matrix ( $S_w$ ) be singularity. This problem comes because those invariant moment's values are close to or event zero.

#### 4. Features Clusters

##### 4.1. Classical LDA

The aim of LDA is to find a transformation data such that feature clusters are most separable after the transformation. The LDA determines the optimum projection matrix from both the  $S_b$  and the within-class scatter matrix ( $S_w$ ), as explained in Refs. [5,9,13].

Suppose we have  $\{(X_1^1, x_1^1, C_1), \dots, (X_{N_1}^1, x_{N_1}^1, C_1); \dots; (X_1^L, x_1^L, C_L), \dots, (X_{N_L}^L, x_{N_L}^L, C_L)\}$ , are images sample form  $L$  classes, each class has  $N$  samples,  $X_i^k$  represents image matrix of  $i$ -th samples of  $k$ -th class,  $x_i^k$  represents features vector of  $i$ -th samples of  $k$ -th class, where  $i=1, \dots, N_k$ , and  $N_k$  is number of training samples of class  $C_k$ . Let  $N = \sum_{j=1}^L N_j$  is total samples size. Let define  $\mu_k$  as mean features vector of class  $C_k$ , and  $\mu_a$  as mean features vector of all samples:  $\mu_k = 1/N_k \sum_{i=1}^{N_k} x_i^k$  and  $\mu_a = \sum_{k=1}^L N_k / N \cdot \mu_k$ . From the data samples, both the  $S_b$  and the  $S_w$  can be determined as follows.

$$S_b = \sum_{k=1}^L P(x^k) (\mu_k - \mu_a) (\mu_k - \mu_a)^T \tag{1}$$

$$S_w = \frac{1}{N} \sum_{k=1}^L \sum_{i=1}^{N_k} (x_i^k - \mu_k) (x_i^k - \mu_k)^T \tag{2}$$

where the  $P(x^k)=N_k/L$ . From both the  $S_b$  and  $S_w$ , the optimum projection matrix of LDA,  $W$ , can be determined which has to satisfy the following criterion.

$$J_{LDA}(W) = \arg \max_W \frac{|W^T S_b W|}{|W^T S_w W|} \tag{3}$$

The  $W=[w_1, w_2, w_3, \dots, w_p]$  which satisfy the Eq. (3) can be obtained by solving eigen problem of inverse  $S_b$  times  $S_w$  and then select  $p$  orthonormal eigenvectors corresponding to the largest eigenvalues (i.e.  $p < q$ ,  $q$  is the dimensional size of vector input,  $x$ ). Finally, the input vector is projected into new space using the following equation.

$$y_i^k = W^T x_i^k \tag{4}$$

The main problem of the LDA has the singularity problem of scatter matrix due to the high data dimensional and small number of training samples so called as small size problem (SSS). Furthermore, they require retraining all samples to obtain the most favorable projection matrix when new data registered into the system. It comes because all terms on summation of  $S_b$  depend on global mean and summation should be recalculated from scratch (see Eq. 1). Regarding to the SSS problem of LDA, some methods have been proposed to solve that problem such as DLDA, RLDA, and PCA+LDA. However, those methods still require large computational costs and memory space requirement, and have retraining problems.

**4.2. Simplified LDA**

In order to decrease the retraining computational load of the LDA algorithm, we simplified the LDA using a constant global mean assignment for all samples. Suppose we have the data cluster of two classes in three-dimensional which is normalized in the range [0-1], shown in Figure 2(a).

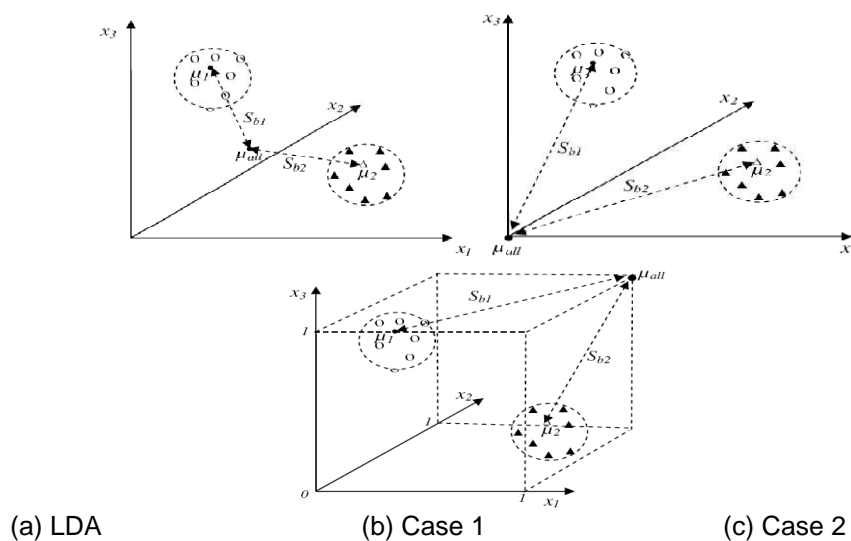


Figure 2. The illustration two classes clustering of three-dimensional data

Using the Eq. (1), the  $S_b$  of this case can be rewritten as follows.

$$S_b^{org} = P(x^1)(\mu_1 - \mu_a)(\mu_1 - \mu_a)^T + P(x^2)(\mu_2 - \mu_a)(\mu_2 - \mu_a)^T \quad (5)$$

From the Figure 2(b), when the  $\mu_a$  is moved in the origin point ( $\mu_a$  is equal to null vector,  $\mu_a = [0 \ 0 \ 0]^T$ ), the  $S_b$  of this case can be simplified as

$$S_b^{c_1} = P(x^1)\mu_1\mu_1^T + P(x^2)\mu_2\mu_2^T \quad (6)$$

In this case, the  $S_b^{c_1}$  has less computational complexity, because it does not require to recalculate global mean and it has the same characteristic as the original one in term of separability and symmetrical matrix. The separability means that  $|W^T S_b^{c_1} W|$  is probably higher or the same than or as the  $|W^T S_b^{org} W|$ , which make the data cluster be more or the same separable than or as that of the original one. This condition makes the discrimination power of the data projection be higher or the same than or as that of the original the data projection.

If the  $\mu_a$  is moved to the maximum value of the range ( $\mu_a = [1 \ 1 \ 1]^T$ ), as shown in Figure (c), the  $S_b$  of this case can be written as

$$S_b^{c_2} = P(x^1)(\mu_1 - [1])(\mu_1 - [1])^T + P(x^2)(\mu_2 - [1])(\mu_2 - [1])^T \quad (7)$$

Even though the predictive  $S_b^{c_2}$  has the same computational complexity as that of the Eq.(5) but it does not require to recalculate global mean. Logically, the  $S_b^{c_1}$  and  $S_b^{c_2}$  have more separable scatter than that of the original one because the original global mean of the training data is placed in between  $[0]$  and  $[1]$  while the constant global mean is placed in the maximum  $(1)$  and in the minimum value  $(0)$ . Therefore the distance between the mean of each class to the original global mean is probably longer or the same than or as that of each class to the constant global mean. In other words, the  $|W^T S_b^{c_1} W|$  and  $|W^T S_b^{c_2} W|$  are probably higher or the same than or as the  $|W^T S_b^{org} W|$ .

For  $n$ -dimensional data, all cases of approximation  $S_b$  can be generalized as the following equation.

$$S_b^c = \sum_{k=1}^L P(x^k)(\mu_k - \mu_p)(\mu_k - \mu_p)^T \quad (8)$$

where the  $\mu_p$  is a constant of global mean  $\mu_a$ . When a new class,  $x^{new}$  comes into the system, the Eq. (8) can be updated as following equation.

$$\begin{aligned} S_b^{c_u} &= \sum_{k=1}^L P(x^k)(\mu_k - \mu_p)(\mu_k - \mu_p)^T + P(x^{new})(\mu_{new} - \mu_p)(\mu_{new} - \mu_p)^T \\ &= S_b^c + P(x^{new})(\mu_{new} - \mu_p)(\mu_{new} - \mu_p)^T \\ &= S_{old}^c + S_{new}^c \end{aligned} \quad (9)$$

In order to decrease much more retraining time complexity, the updating of the  $S_w$  also can be simplified as the following equation.

$$\begin{aligned} S_w^u &= \frac{1}{N} \sum_{k=1}^L \sum_{i=1}^{N_k} (x_i^k - \mu_k)(x_i^k - \mu_k)^T + \frac{1}{N_{new}} \sum_{i=1}^{N_{new}} (x_{new}^k - \mu_k)(x_{new}^k - \mu_k)^T \\ &= \frac{1}{N} \sum_{k=1}^L S_w^k + \frac{1}{N_{new}} S_w^{new} \\ &= \frac{1}{N^u} (S_w^{old} + S_w^{new}) \end{aligned} \quad (10)$$

where  $S_w^k = \sum_{i=1}^{N_k} (x_i^k - \mu_k)(x_i^k - \mu_k)^T$  and  $N^u = N + N_{new}$ .

Let us compare the Eq. (9) with Eq. (1) and Eq. (10) with Eq. (2), both the Eq. (9) and Eq. (10) seem to have much less complexity than that of original one (Eq. (1) and Eq. (2)). The time complexity of recalculating  $S_b$  using Eq. (9), it requires:  $n^2$  multiplication operations for case 1,  $n^2$  multiplication and  $n^2$  addition operations for case 2, while the updating of  $S_w$  using Eq. (10) also need  $n^2$  multiplication and  $n^2$  addition operations.

By substituting the original  $S_b$  and  $S_w$  with updated ones (Eq. (9) and Eq. (10)) of classical LDA eigen analysis, the optimum  $W$  called as sLDA projection matrix is obtained. In case of updating  $S_b$  using case 1 and case 2 model of constant global mean, the sLDA is called as sLDA\_C1 and sLDA\_C2, respectively. The optimum projection matrices are constructed by selecting small number ( $m$ ) of eigenvectors, which correspond to the largest eigenvalues. By using these optimum projection matrices, the projected features of the both training and querying data set can be performed as done by the classical LDA using Eq. (4).

**4.3. The Strength of sLDA**

In order to prove whether the approximation  $S_b$  provide the same separability as the original one, the evaluation using discrimination power (DP) [16], which represent the ability of features separation is performed by the following equation.

$$J(W) = sep(W) = trace(S_w^{-1}S_b) \tag{11}$$

Where  $sep$  means separation. As reported in Ref. [16], the DP parameter had been successfully used to analysis for knowing which parts of the face have large discriminant information.

In this case, the DP is examined in of all cases of approximation  $S_b$  using data from well-known ORL database, which is performed using the following procedure. In this case, we also compare the DP of the proposed methods with original LDA (LDA).

- (i) Obtaining the optimum  $W$  of simplified LDA using mentioned procedure.
- (ii) Transforming the  $x_i^k$  to the projected data ( $y_i^k$ ) using the Eq. (4).
- (iii) Determining the within and between class scatter of projected data ( $y_i^k$ ) called as  $S_b^{Pro}$  and  $S_w^{Pro}$  using Eq. (1) and Eq. (2) respectively.
- (iv) Calculating the DP of the projected data by substituting  $S_b$  and  $S_w$  of Eq. (11) with  $S_b^{Pro}$  and  $S_w^{Pro}$ , respectively.

The examination result shows that all cases of approximation  $S_b$  have closely the same DP as that of the original LDA algorithm (see Figure 3(a)). It means that the proposed methods provide the ability in terms of feature cluster with the original LDA but requiring less time processing. In order words, all cases of approximation  $S_b$  have been proved that the  $|W^T S_b^{c1} W|$  and  $|W^T S_b^{c2} W|$  have much the same value as the  $|W^T S_b^{org} W|$  which make the data cluster of the simplified be the same separable as that of the original one.

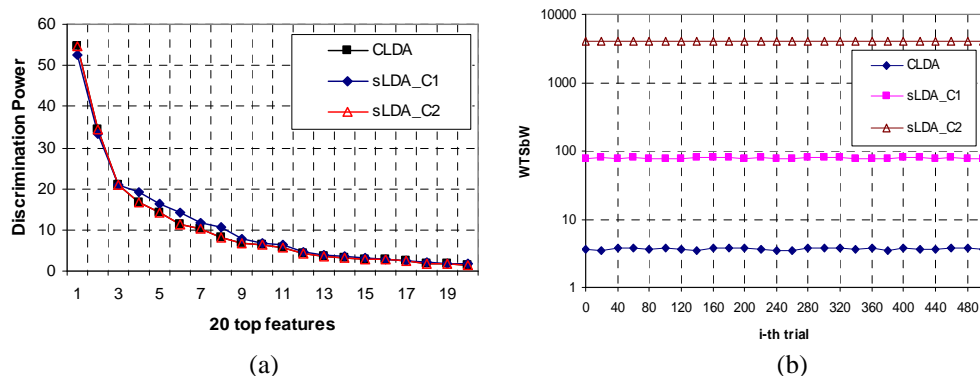


Figure 3. (a) The discrimination power of our proposed methods compared to established LDA and (b). the comparison of the  $|W^T S_b^{org} W|$  of the proposed methods with that of CLDA

In addition, all cases of approximation  $S_b$  tend to provide almost the same achievements as those of the original one when they are implemented as features classification such as face recognition.

In addition, to prove that  $|W^T S_b^{c1} W|$  and  $|W^T S_b^{c2} W|$  are probably higher or the same than or as the  $|W^T S_b^{org} W|$ , we did further experiment as follows. Suppose a training set consists of 100 classes, each class has at least five members and each vector of each class is created by random value. From this data set, we tried 500 experiments and each  $i$ -th ( $i=1, \dots, 500$ ) process  $|W^T S_b^{c1} W|$  and  $|W^T S_b^{c2} W|$  were determined by trace their most significant eigen values. The results show that the scatter of case 2 provide higher than that of case 1 and that LDA, as shown in Figure 3(b). It means that the case 1 and 2 possibly has better class separable than that of the LDA. Therefore, the case 1 and 2 provide little bit higher recognition rate than that of the CLDA (see Figure 7). This result also match/inline with the result of power discriminant analysis, as shown in Figure 3.

### 5. The Implementation for Face Recognition

The diagram block of the proposed face recognition is presented in Figure 4 consisting of three main processing: feature extraction, retraining process, and recognition process. The function of HF is to find out the most useful information from face images using the algorithm as presented in section 3. The retraining process is done by the sLDA as presented in section 4.2. Each retraining step, the matrix  $S_b$  and  $S_w$  have to be save for performing next retraining.

In recognition process, Each face image is extracted into HF and then the HF is transformed into new space data called as sLDA projected data. The sLDA projected data of the training samples is matched with the querying projected data using nearest neighbor rule. It means the smallest score is concluded as the best matching face.

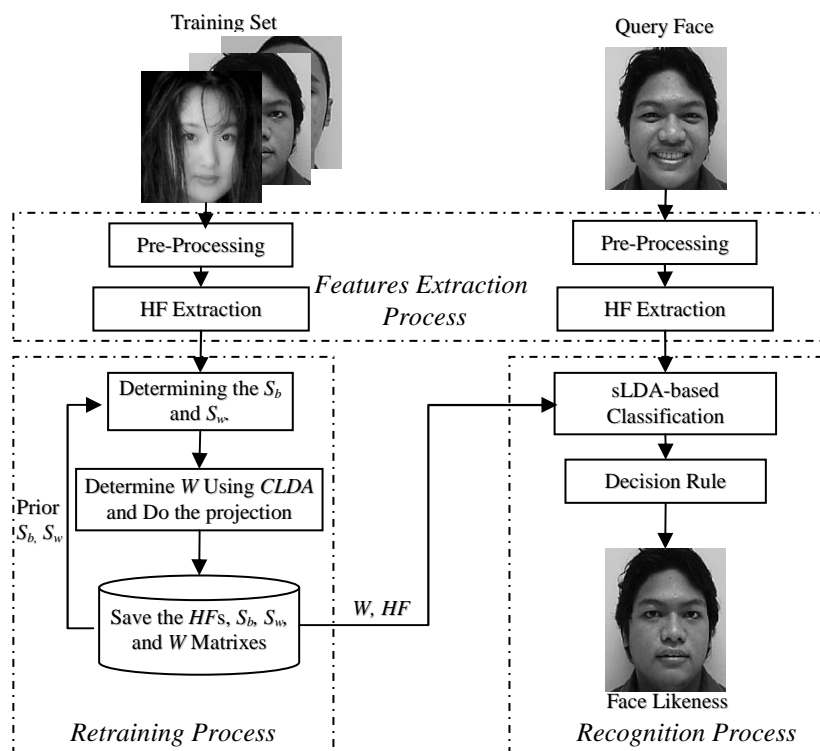


Figure 4. Diagram block of sLDA based face recognition



## 6. Experimental and Results

The experiments are carried out using several challenging face databases: the ORL database [19], YALE database [22], ITS-Lab. Kumamoto University database [12], INDIA database [21], and FERET database [18]. Each database has special characteristics, which are described below.

The ORL database was taken at different times, under varying lighting conditions with different facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). All of the images were taken against a dark homogeneous background. The faces of the subjects are in an upright, frontal position (with tolerance for some side movement). The ORL database is a grayscale face database that consists of 40 people, mainly male. Total face images are 400 samples. The example of face pose variations of ORL database is shown in Figure 4(a). The ITS-Lab database consists of 48 people and each person has 10 pose orientations as shown on Figure 4(b). Total face images are 480 samples. The face images were taken by Konica Minolta camera series VIVID 900 under varying light condition. The YALE database is grayscale database, which consists of 15 people, and each person consists of 11 differential facial expression, illumination, and small occlusion (by glass). Therefore, there are 165 face images in the database. On of the person is shown on Figure 4(c). The INDIA database consists of 61 people (22 women and 39 men) with each person having eleven pose orientations as shown on Figure 4(d): looking front, looking left, looking right, looking up, looking up towards left, looking up towards right, and looking down. The Indian database also included some emotions: neutral, smile, laughter, sad/disgust. Total face images are 671 samples. From FERET database, 2032 images from 508 individuals were selected which correspond to four different sets, namely (*fa*, *fb*, *ql*, *qr*). The example of face pose variations of some database is shown in Figure 4(e).

All the experiments are performed in those databases with the following assumptions:

- (i) The face image size was 128 x128 pixels (28 pixels/cm) represented using 24 bits for color image and 8 bit of for grayscale image per pixel.

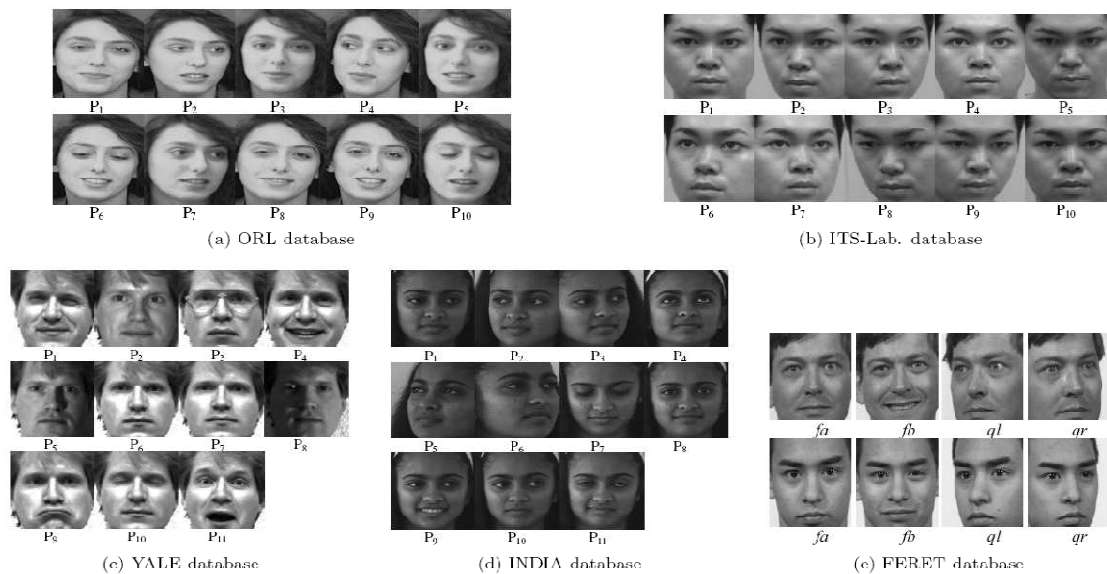


Figure 5. Example of face pose variation of single subject

- (ii) A half of samples of each class/subject members were selected as training faces, and testing faces overlapped with the training faces.
- (iii) The training face was included in the query process because there was possibility that the query faces were almost or exact the same as the training face in the real time system.
- (iv) The dimensional size of HF was set-up 53 elements of original size ( $128^2$  pixels).

In order to know whether our proposed methods can work for with and without HF (no-dimensional reduction), we did experiments on small size database (ORL and ITS databases)

and large size database (FERET database). In addition, to handle the out of memory problem on the LDA and our proposed methods due to large dimensional size of input image (128x128 pixels), we resized the original face image into 32 x 32 pixels. This size was chosen because some researchers used this dimensional size when they directly implemented the LDA and its variations for data cluster without dimensional reduction. The results show that our proposed methods can works well (see Figure 7) which are indicated by high enough recognition rate by about more than 89% and more than 96% for without (base line) and with HF, respectively. In addition, the results also show that the recognition rate of the proposed methods (sLDA\_C1 and sLDA\_C2) are almost the same as that of the CLDA for both using HF and without HF, which means the propose method has the same ability as the CLDA in terms of features clustering. The HF can provide better recognition rate than that of without HF because it has ability to extract the most significant and remove the noise information of face image. In addition, the HF also provides higher discriminatory power than that of without HF as reported in Wijaya et al (2008) which means the HF already contain most useful discriminant information.

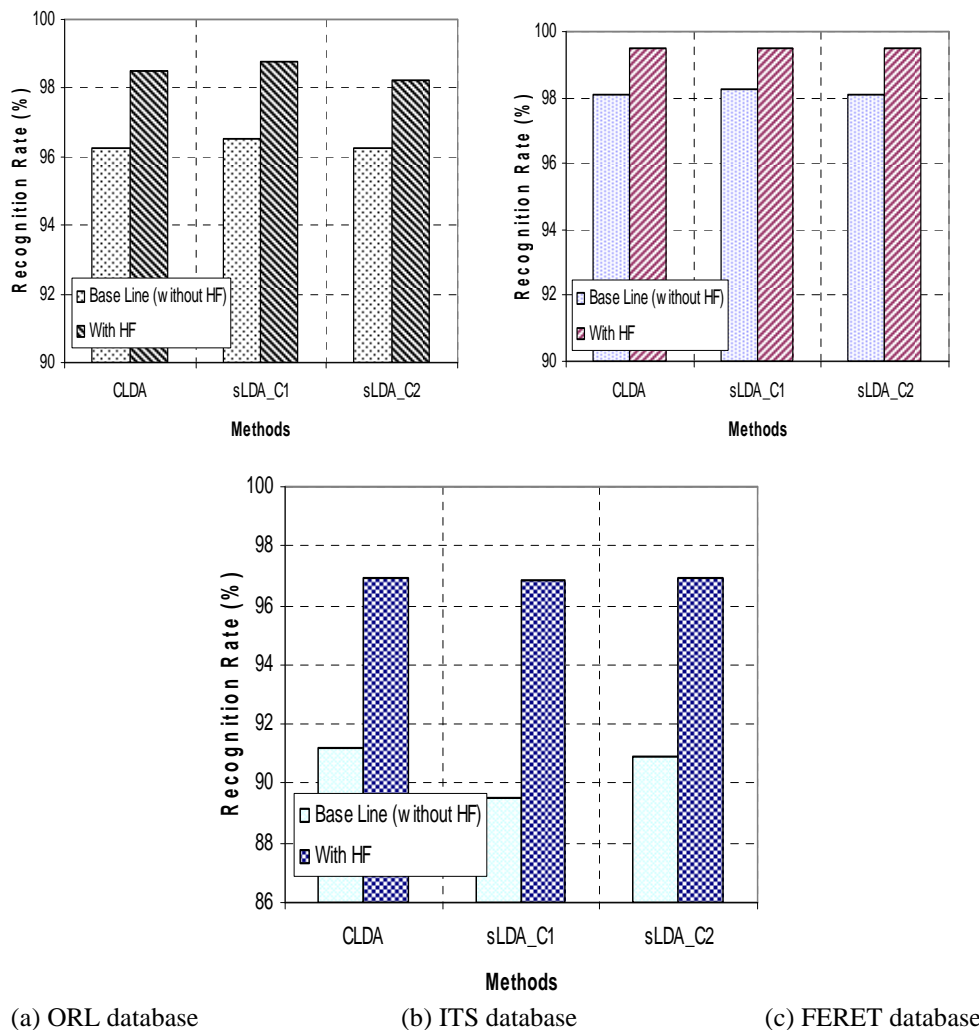


Figure 6. The recognition rate of CLDA and our proposed methods (sLDA\_C1 and sLDA\_C2) for with and without HF

The second experiment, which was performed in all the mentioned databases, investigates the robustness of our proposed methods over the established methods. From the experimental results, all variations of sLDA tend to provide the same recognition rate as the

CLDA but outperform over the established methods in all databases, as shown in Figure 7. It means that our proposed methods give robust recognition rate in all tested databases. It can be achieved because all of the sLDA variations almost have the same discrimination power as the CLDA, see Figure 3(a). It proves that the approximation  $S_b$  of sLDA has the same structure and can satisfy optimum criterion as that of the CLDA.

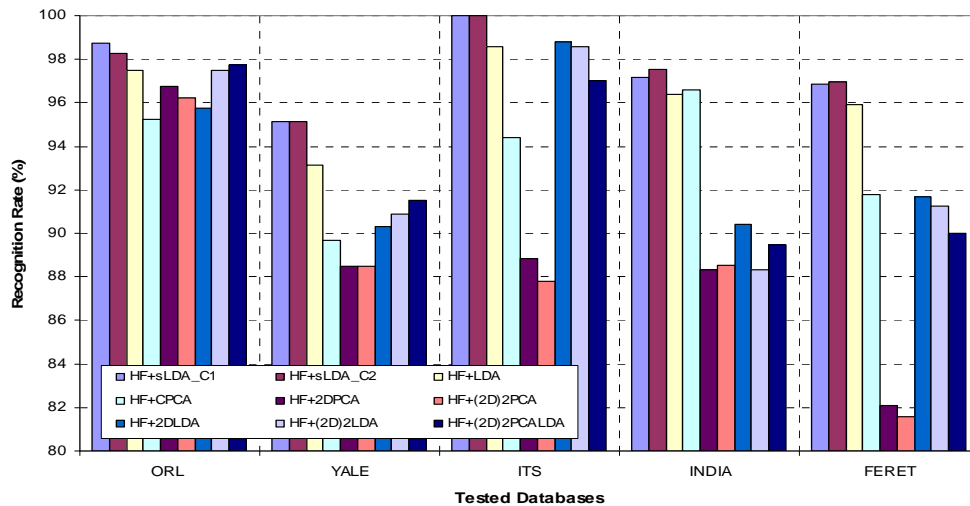


Figure 7. The comparison of the recognition rate of the proposed to established algorithms

In order to show that the proposed method required less retraining time for new insertion data, the next experiment was performed. It was done in FERET face database performing the retraining gradually: firstly, it was trained 208 face classes as initial data and then added gradually 20 new face classes to the system until reaching 508 face classes. The experimental results were plotted in Figure 8(a). It shows that all variations of the sLDA require less retraining time than that of CLDA. Even though, the retraining times of the sLDA increase but its average of increment slope is less than half of that of CLDA. Moreover, the larger number of classes are, the larger increment of training time of the CLDA will be, while the sLDA almost require constantly retraining time. This result proves that the sLDA requires less retraining time, which means the sLDA, can solve the retraining problem of CLDA. It can be achieved because the sLDA has simpler computation complexity than that of CLDA ( $d+d^2$  multiplication for sLDA\_C1,  $d+d^2$  multiplication and  $d$  addition for sLDA\_C2, and  $(L+1)d+d^2$  multiplication and  $(L+1)d$  addition for CLDA), as described in section (4.3).

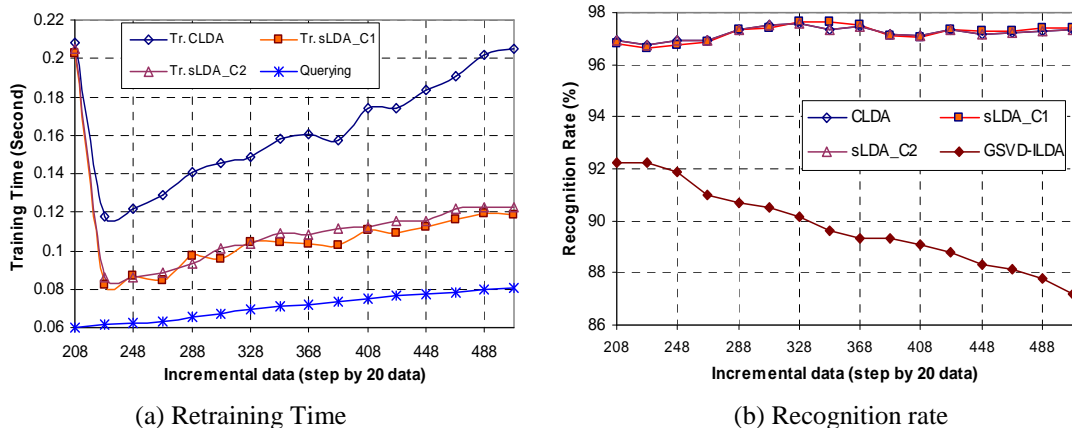


Figure 8. (a) Retraining and querying time, and (b) Recognition stability for incremental data.

Regarding to the recognition rate stability of our proposed methods, they give almost the same stability recognition rate as that of the CLDA but more stable than that of GSVD-ILDA which is one of the recent variation of LDA for solving retraining problem [20]. The results are in-line with the previous experimental results that all of the sLDA variations have the same structure as the CLDA but they have simpler computation complexity. Therefore, all cases of sLDA are alternative algorithm for features cluster of large sample size databases, which requires much retraining processing.

## 7. Conclusion and Future Work

From all of the experimental results have been prove our hypothesis as describe below. Firstly, the HF as a dimensional reduction of raw image has been successfully implemented with good achievement when the one-dimensional CLDA, and sLDA as features clustering. Secondly, all variation of the sLDA-base face recognition has been proved that they require less time for retraining. The rest, the proposed method (sLDA) outperforms over the recent sub-space methods (2DPCA and 2DLDA based methods) and the sLDA can used as alternative solution to avoid recalculating the  $S_b$  and global mean on the retraining process of LDA. In addition, the sLDA is alternative solution for large size data clustering because it does not depend on the global means. However, All of the sLDA just work for updating the retraining data belonging to the new class.

The new strategy using statistical prediction will be considered to avoid this problem. In addition, to get more precise verification result, we will consider more local features analysis involving eyes, nose, mouth, and context information of the face image.

## References

- [1] W Zhao, R Chellappa, and A Rosenfeld. Face Recognition: a Literature Survey. *ACM Computing Surveys*. December 2003, 35: 399-458.
- [2] W Gao, B Chao, S Shan, X Chen, D Zhou, X Zhang and D Zhao. The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. *IEEE Transactions on System, Man, and Cybernetics-Part A: Systems and Human*. January 2008; 38(1): 149-161.
- [3] ZM Hafed and MD Levine. Face Recognition Using the Discrete Cosine Transforms. *International Journal of Computer Vision*. 2001; 43(3): 167-188.
- [4] D-Q Dai and PC Yuen. Wavelet Based Discriminant Analysis for Face Recognition. *Applied Mathematics and Computation*. 2006; 175: 307-318.
- [5] W Chen, J-E Meng, and S Wu. PCA and LDA in DCT Domain. *Pattern Recognition Letter*. 2005; 26: 2474-2482.
- [6] J Yang, D Zhang, AF Frange, and J-Y Yang. Two Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition, *IEEE Transaction on Pattern Analysis and Machine Intelligence*. 2004; 26(1): 131-1374.
- [7] Y-G. Kim, Y-J. Song, U-D. Chang, D-W, Kim, T-S. Yun, and J-H. Ahn. Face Recognition Using a Fusion Method Based on Bidirectional 2DPCA. *Applied Mathematics and Computation*. 2008: 205, 601-607.
- [8] D. Zhang and Z-H Zhou. (2D) 2PCA: Two-Directional Two-Diemnsional PCA for Efficient Face Representation and Recognitio. *Neurocomputing*. 2005; 69: 224-231.
- [9] H Yu and J Yang. A Direct LDA Algorithmn for High-Dimensional Data-with Applicaton to Face Recognition. *Pattern Recognition*. 2001; 34: 2067-2070.
- [10] J Wang, W Yang, Y Lin, and J Yang. Two-Directional Maximum Scatter Difference Discriminant Analysis for Face Recognition. *Neurocomputing*. 2008; 72: 352-358.
- [11] S Nousath, GH Kumar, and P Shivakumara. (2D) 2LDA: An Efficient Approach for Face Recognition. *Pattern Recognition*. 2006; 39: 1396-1400.
- [12] IGPS Wijaya, K Uchimura, and Z Hu. Multipose Face Recognition Based on Frequency Analysis and Modified LDA. *The Journal of the IIEEJ*. 2008; 37(3): 231-243.
- [13] IGPS. Wijaya, K. Uchimura, and Z. Hu. *Pose Invariant Color Face Recognition Based on Frequency Analysis and DLDA with Weight Score Classification*. The 2009 World Congress on Computer Science and Information Engineering (CSIE 2009). USA: Los Angeles. March 2009.
- [14] IGPS Wijaya, K Uchimura, and Z Hu. *Why the Alternative PCA Provide Better Performance for Face Recognition*, The International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2009), London UK. May 2009: 169-172.
- [15] Z Li, D Lin, and X Tang. *Non Parametric Discriminant Analysis for Face Recognition*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2009; 31(4): 755-761.

- [16] K Etemad, R Chellappa. Discriminant Analysis for Recognition of Human Face Images. *J. Opt. Soc. Am. A*. 1997; 14(8): 1724-1733.
- [17] RC Gonzalez and RE Woods. Digital Image Processing, Third Edition. USA: Pearson Prentice Hall. 2008: 839-842.
- [18] PJ Philips, H Moon, SA Risvi, and PJ Rauss. The FERET Evaluation Methodology for Face Recognition Algorithms. *IEEE Trans. Pattern Anal. Machine Intell.* 2000; 22(10): 1090-1104.
- [19] F Samaria and A. Harter. *Parameterization of a Stochastic Model for Human Face Identification*. 2nd IEEE Workshop on Applications of Computer Vision. Sarasota Florida. 1994: 138-142.
- [20] H Zhao, P Yuen. Incremental Linear Discriminant Analysis for Face Recognition. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*. 2008; 38(1): 210-221.
- [21] Indian Face Database, <http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/>
- [22] The Yale Face Database, <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>