Face Recognition Using Holistic Features and Within Class Scatter-Based PCA

I Gede Pasek Suta Wijaya, *Non-Member, IEEE,* Keiichi Uchimura, *Non-Member, IEEE,*
Gou Koutaki, *Non-Member, IEEE*

Abstract—**The Principle Component Analysis (PCA) and its variations are the most popular approach for features clustering, which is mostly implemented for face recognition. The optimum projection matrix of the PCA is typically obtained by eigen-analysis of global covariance matrix. However, the projection data using the PCA are lack of discriminatory power. This problem is caused by removing the null space of data scatter that contains much discriminant information. To solve this problem, we present alternative strategy to the PCA called alternative PCA, which obtains the optimum projection matrix from within class scatter instead of global covariance matrix. This algorithm not only provides better features clustering than that of common PCA (CPCA) but also can overcome the retraining problem of the CPCA. In this paper, this algorithm is applied for face recognition with the holistic features of face image, which has compact size and powerful energy compactness as dimensional reduction of the raw face image. From the experimental results, the proposed method provides better performance for both recognition rate and accuracy parameters than those of CPCA and its variations when the tests were carried out using data from several databases such as ITS-LAB., INDIA, ORL, and FERET.**

Index Terms—**Holistic features, within class scatter, sub-space, LDA, PCA, and face recognition.**

I. INTRODUCTION

The Principle Component Analysis (PCA) and its combinations become most popular approach for features clustering, because of its uncomplicated processing. The optimum projection matrix (W) of PCA is typically obtained by eigen analysis of global covariance matrix [1], [2], [3], [4], which remove the null space of data scatter. This removing make the projection data using PCA be lack of discriminatory power because the null space of data scatter containing much discriminant information is not included. To keep the null space of data scatter, an alternative approach is proposed to obtain the optimum W using eigen analysis of within class scatter (S_w) of trained data. The advantage of this algorithm not only tends to provide better separable projection data but also can solve the retraining problem of the common PCA (CPCA). The retraining problem always comes when the PCA is applied for incremental data, because it has to recalculate the global mean and covariance when new data is registered into the system. Consequently, it makes the CPCA algorithms require large

IGPS. Wijaya is with the Department of Informatics Engineering and Electrical Engineering, Mataram University, Indonesia, and the Department of Electrical and Computer Engineering, Kumamoto University, JAPAN, e-mail: gpsutawijaya@te.ftunram.ac.id

K. Uchimura and G. Koutaki are with the Department of Electrical and Computer Engineering, Kumamoto University, JAPAN, e-mail: {uchimura,koutaki}@cs.kumamoto-u.ac.jp

computational costs and high memory space requirements. In addition, the large computational costs and high memory space requirements are also caused by implementing directly the raw face image as the input of the PCA algorithm. In order to overcome the large computational costs and high memory space requirement problems, the compact holistic features (HF) of face image is applied as dimensional reductions of the raw face image, as proposed in Refs. [4], [5].

II. PREVIOUS WORK

The most related approach to our proposed method is face recognition based on PCA and its variations as described in some papers [2], [3], [4], [7], [8], [9]. From those references, the PCA can be classified into one-dimensional PCA (1D-PCA)[2], [4], [7] which is vector-based analysis and two-dimensional PCA (2D-PCA) [3], [8], [9], [10] which is matrix-based analysis. The 2D-PCA methods were proposed in order to keep the two-dimensional structure of data input, such as face image. In addition, the 2D-PCA gave better achievements than that of the 1D-PCA. However, the 2D-PCA requires more coefficients for image representation than that of 1D-PCA.

The 2D-PCA has several variations, such as the two-directional and two-dimensional PCA ($(2D)^2$PCA)[9] and the two-directional and two-dimensional PCALDA ($(2D)^2$PCALDA)[11]. They work on both row and column directions of image to find out the optimum W, which the aims are not only to decrease coefficients requirement for image representation but also to increase image recognition achievements. The face recognition using the $(2D)^2$PCALDA outperformed over other algorithms. However, the $(2D)^2$PCALDA required large computational cost because it performed not only PCA but also LDA processes. In addition, the PCA-based face recognition and its variations still leave retraining problem for incremental data because they have to recalculate the global mean and global covariance to obtain most favourable W when new data are registered into the system. This problem makes them require long retraining time (large computational cost). Furthermore, they are also lack of discriminatory power because of not only removing the null space of data scatter but also spreading the discriminant information into all of the eigen vectors.

To address the retraining problems of the PCA, we propose a new formulation of PCA by defining the optimum W from S_w[6]. It can be realized by considering the fisher criterion and setting S_b as identity matrix. Then, this approach is integrated with Mahalanobis distance for face classifier with

the HF as dimensional reduction of the raw face image. In this paper, we provides further explanation about the strength and ability of the proposed algorithm to perform features clustering and carries out more experiments to prove the mathematical analysis of the within class scatter-based PCA which can solve the retraining problem and gives stable achievement for incremental data. In addition, we compare our proposed method, which work for one dimensional data, with established PCA-based methods such as common PCA, 2D-PCA, $(2D)^2$PCA, $(2D)^2$PCALDA, and GSVD-ILDA.

III. PCA VERSUS WITHIN CLASS SCATTER-BASED PCA

A. PCA Algorithms

The classical PCA (CPCA) works based on the following criterion[2], [7]:

$$J(W) = \arg\max_W (W^T C_g W) \qquad (1)$$

where C_g is the global covariance of training set consisting of i-th classes and each class has k-th elements. The C_g is determined using the following equation:

$$C_g = \frac{1}{N} \sum_{k=1}^{L} \sum_{i=1}^{N_k} (x_i^k - \mu_a)(x_i^k - \mu_a)^T \qquad (2)$$

where N and μ_a are total element and global mean of the training set.

The optimum projection matrix (W) which satisfies the Eq. 1 is obtained by performing eigen analysis of C_g and selecting m orthonormal eigen-vectors corresponding to the largest eigen-values (i.e. $m < n$, where n is dimensional of input vector, x_i^k), then placing them into $W = [w_1, w_2, w_3, ..., w_m]$. By this ($W$), the x_i^k is projected into PCA space (y_i^k) using equation below:

$$y_i^k = W^T x_i^k. \qquad (3)$$

The main problem of this method is lack of discriminatory power and has to retrain all of data samples to obtain the most favourable W for incremental data.

B. Within Class Scatter-based PCA

In order to get higher discriminatory power and solve retraining problem of CPCA, an alternative PCA (APCA) which is based S_w instead of global covariance (C_g, Eq. (2)) called as Within Class Scatter-based PCA. This method is derived from Fisher criterion Ref. [2], [4]. Let us call back the Fisher criterion without considering the between class scatter matrix, S_b, (in other word the S_b is set as identity matrix (I)), then it can be written as follows.

$$
\begin{aligned}
J(W) &= \arg\max_W \frac{|W^T S_b W|}{|W^T S_w W|} \\
&= \arg\max_W \frac{|W^T W|}{|W^T S_w W|}
\end{aligned} \qquad (4)
$$

Suppose the W is orthonormal ($W^T W = I$), therefore Eq. (4) can be simplified as

$$
\begin{aligned}
J(W) &= \arg\max_W \frac{|I|}{|W^T S_w W|} \\
&\cong \arg\min_W |W^T S_w W|
\end{aligned} \qquad (5)
$$

In order to satisfy this optimum criterion, the determinant of the denominator should be as small as possible or even close to zero. It can be satisfied by eigen analysis of S_w then selecting a small number of eigen-vectors (m) corresponding to the smallest eigen-values (keeping the null space). The same as PCA, the size of the selected eigen vectors (m) is much less than that of input vector (n).

This approach has two main advantages compare with the CPCA: it can solve retraining problem and its projection data are more separable, which is proved by higher discriminatory power than that of CPCA. The first advantage can be achieved because the updating of S_w matrix does not depend on the global mean of the training samples at all, as written below.

$$
\begin{aligned}
S_w &= \frac{1}{N} \sum_{k=1}^{L} \sum_{i=1}^{N_k} (x_i^k - \mu_k)(x_i^k - \mu_k)^T \\
&= \frac{1}{N} \sum_{k=1}^{L} S_w^k = \frac{1}{N} \left\{ \sum_{k=1}^{L-1} S_w^k + S_w^L \right\} \\
&= \frac{1}{N_{new}} \{ S_w^{old} + S_w^{new} \}
\end{aligned} \qquad (6)
$$

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

Let us compare the Eq. (2) to Eq. (6), the Eq. (2) depends on global mean (μ_a), but the Eq. (6) does not do at all. If the new class data come into the system, the Eq. (2) has to recalculate the global mean and global covariance from all trained classes. However, the Eq. (6) just calculates the newest class covariance, S_w^{new}, and then adds it to the S_w^{old}.

In terms of the discriminant of projection data, the APCA tends to provide more separable projection data than that of CPCA, because the $J(W)$ of APCA is higher than that of CPCA. From eigen analysis, it is easy to understand that the eigen values of the global covariance or the S_w data is diagonal matrix (λ^{APCA}), which the diagonal elements are the selected eigen values. By substituting the λ^{APCA} into Eq. (5), then we obtain:

$$
\begin{aligned}
J_{APCA}(W) &= \arg\max \frac{|I|}{|\lambda^{APCA}|} \\
&= \arg\max (|\lambda^{APCA}|^{-1}) \\
&= \arg\max \left(\prod_{i=1}^{m} \frac{1}{\lambda_i^{APCA}} \right)
\end{aligned} \qquad (7)
$$

Theoretically, the CPCA approach removes the null space (selecting m eigen vectors correspond to the largest eigen values), while the APCA approach keeps the null space of S_w (keeping m eigen vectors correspond to the smallest eigen values). It means one of the eigen values of global covariance or S_w is close to zero. Therefore, the determinant of covariance

of features projection using APCA becomes very large or close to infinity compared to trace of covariance of features projection using CPCA. From this evaluation, we can conclude that the projection data using APCA (i.e projection features vectors) are more separable than those of using CPCA.

Another optimum criterion can be used to justify the strength of APCA called as discriminatory power[12], as written below.

$$
\begin{aligned}
S^W &= S_w^{-1} S_b \\
J(W) &= sep(W) = trace(S^W) \quad (8)
\end{aligned}
$$

$J(W)$ is the discriminatory power of given projection data. The higher discriminatory power is the more separable projection data will be. Let us define the covariance matrix of projection features vector, as written below.

$$
\begin{aligned}
S_w^{proj} &= \sum_{k=1}^{L} E\left[(y^k - E[y^k])(y^k - E[y^k])^T\right] \\
&= \sum_{k=1}^{L} E\left[W^T(x^k - E[x^k])(x^k - E[x^k])^T W\right] \\
&= W^T\left\{\sum_{k=1}^{L} E\left[(x^k - E[x^k])(x^k - E[x^k])^T\right]\right\} W \\
&= W^T S_w W = [\lambda] \quad (9)
\end{aligned}
$$

By using Eq. (9), the discriminatory power of CPCA (Eq. (1)) can be derived as follows:

$$
\begin{aligned}
J_{CPCA}(W) &= \underset{W}{arg\,max}(W^T C_g W) \\
&= trace(C_g^{proj}) = \sum_{i=1}^{m} \lambda_i^{CPCA}. \quad (10)
\end{aligned}
$$

While the discriminatory power of APCA can be derived from Eq. (8) :

$$
\begin{aligned}
J_{APCA}(W) &= trace\left((S_w^{porj})^{-1} S_b\right) \\
&= trace\left((W^T S_w W)^{-1}\right) \\
&= \sum_{i=1}^{m} \frac{1}{\lambda_i^{APCA}} \quad (11)
\end{aligned}
$$

where $S_b = I$ (not considering S_b). As described previously, one of the eigen values of global covariance or S_w is close to zero, it makes the discriminatory power of APCA projection features vector become very large. Therefore, the Eq. (10) and Eq. (11) can be compared:

$$
J_{CPCA}(W) \ll J_{APCA}(W). \quad (12)
$$

Because both of optimum criteria are satisfied, It means that the APCA tends to provide better achievements for features clustering than that of the CPCA. This fact will be proved by experimental data in the next sections.
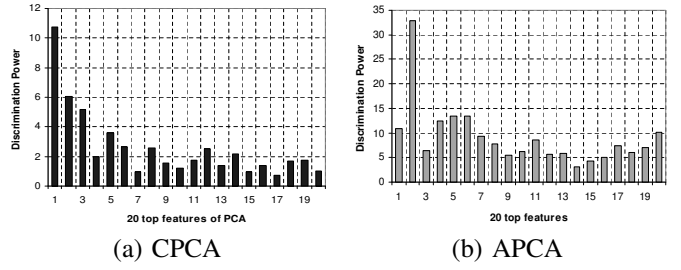

(a) CPCA   (b) APCA

Fig. 1. The discriminatory power of our proposed methods compared to CPCA

C. The discriminatory Power of CPCA and APCA

The discriminatory power (DP) parameter is used to know the features cluster ability of the CPCA and our proposed algorithm. This parameter has been successfully used to know which parts of the face have large discriminant information[12].

In this case, we examine the DP of projection data of the CPCA and APCA, which is determined using the following procedures:

1) Determining the projection data using the Eq. (3) under assumption the W is the optimum W of CPCA or APCA which is obtained using the CPCA or APCA algorithm with the HF as the data input.
2) Calculating the within and between class scatter of projection data (y_i^k) called as S_w^{Pro} using Eq. (9) and S_b^{Pro} using:

$$
S_b^{Pro} = \frac{1}{L} \sum_{k=1}^{L} P(y^k)(\mu_k^y - \mu_a^y)(\mu_k^y - \mu_a^y)^T \quad (13)
$$

where μ_k^y is the mean vector of projection data of k-th class and μ_a^y is the global mean vector of projection data.
3) Calculating the DP of the projection data by substituting S_w and S_b of Eq. (8) with S_w^{Pro} and S_b^{Pro} respectively.

In this case, we examine the DP using well-known ORL database[20]. The examination result shows that both of CPCA and APCA classification information of face image is not placed in few top discriminant vectors but spreading to all over features vectors, as shown in Fig. 1. However, the APCA has higher values of DP than CPCA which satisfies the optimum criterion, as written in the Eq (12). It means that the APCA requires less discriminant vectors than that of CPCA to get the same recognition rate as that of CPCA. Consequently, the more discriminant vectors are considered the larger the face features size will be, which affect to the memory space requirements. In other words, by using this result: the higher DP is, the higher recognition rate will be, which will be proved using experimental data in the next section.

IV. THE IMPLEMENTATION APCA AS FACE RECOGNITION

From DP point of view, the APCA provide better ability of features clustering than that of CPCA. In order to know its further effectiveness on features clustering, we apply the
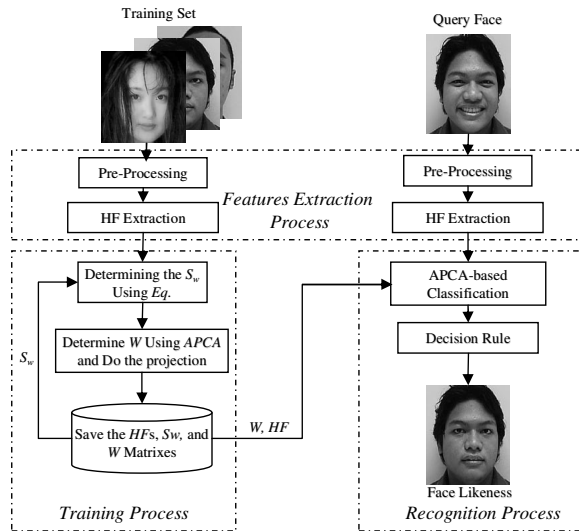
Fig. 2. The block diagram of APCA-based face recognition.

APCA-based features clustering as face recognition. APCA-based features clustering for face recognition consist of three components: pre-processing, features extraction, and classification as shown in Fig. 2.

The main function of pre-processing is the to normalize the large variability of face image due to lighting variations. Recently, the self-quotient image (SQI) and the modified local binary pattern (mLBP) approaches are the most suitable algorithm to this problem for Eigenface face recognition system [16]. In addition, the fast lighting normalization also has been proposed[18] which works well for large illumination face image such as YaleB database than the existed methods such as histogram equalization (HE) and mLBP. Therefore, the lighting normalization from Ref [18] is chosen to be implemented as pre-processing.

Next, if the CPCA and APCA algorithm are implemented directly for face recognition, it will require large computational cost and large memory space. Therefore, a compact HF of face image that is based on frequency and moment analysis of entire face is implemented as dimensional reduction of raw face image. The frequency analysis (i.e. DCT) is used to obtain the dominant frequency content of the face image. The DCT has good energy compactness, therefore original image can be represented by small number of DCTs coefficients[4], [5], [13], [14]. The HF extraction starts from DCT frequency analysis which is employed to get the dominant frequency content of the face image. 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, finally truncate p first vector elements (i.e., p is 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. Next, In order to get robust global HF of any face pose variations, the moment information that provides invariant measure of face images shape is considered. The invariant moment set is invariant to translation, scale change, and rotation therefore

this concept can be applied to get the holistic information of any face pose variations[15]. In this case, the invariant moment analysis is determined just in the intensity component of color images. Finally, from both the moment invariant set (I) and the selected frequency coefficients (f), we construct a HF vector, $x_i = [f_i^{DCT}, I_i]$, where i is i-th class of face image. The dimension of x_i is $p+q$, where p is number of selected frequency coefficients and q is number of selected invariant moments. This HF has been proved to provide higher discriminatory power than without moment information and sufficiently high recognition rate, as presented in Refs. [4], [5], [17].

The classification step consists of two main processes: training and recognition processes. On the training process, the system defines the optimum W using APCA based algorithm as described in sub section III-B with the HF as the raw input. Then, the extracted HF and the determined optimum W are saved into database for the next process. In the recognition process, the Mahalanobis distance based on nearest neighbour rule is implemented for face classification. The negative samples (non-training faces and non-face images) are used to define the threshold for face verification. If the minimum score is less than the defined threshold the input face image is verified as known face (positive face) and other wise is concluded as unknown face (negative face).

V. EXPERIMENTS AND RESULTS

In order to know the performance of the proposed face recognition, several experiments were carried out using data from several challenging face databases: ORL database [20], ITS-Lab. Kumamoto University database[4], INDIA database[22], and FERET database[19]. Each of the mentioned databases has special characteristics as described as follows.

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 Fig. 3(a). The ITS-Lab database consists of 98 people and each person has 10 pose orientations as shown on Fig. 3(b). Total face images of ITS-Lab database are 980 samples. The face images were taken by Konica Minolta camera series VIVID 900 under varying lighting condition. The INDIA database consists of 61 people (22 women and 39 men), which each person has eleven pose orientations as shown on Fig. 3(c): looking front, looking left, looking right, looking up, looking up towards left, looking up towards right, and looking down. The INDIA database also included some emotions: neutral, smile, laughter, sad/disgust. Total face images of INDIA database are 671 samples. From the FERET database, 2032 images of 508 individuals were selected which correspond to four different sets, namely (fa, fb, ql, qr). The example of face pose variations of the FERET database can be seen in Fig. 3(d).
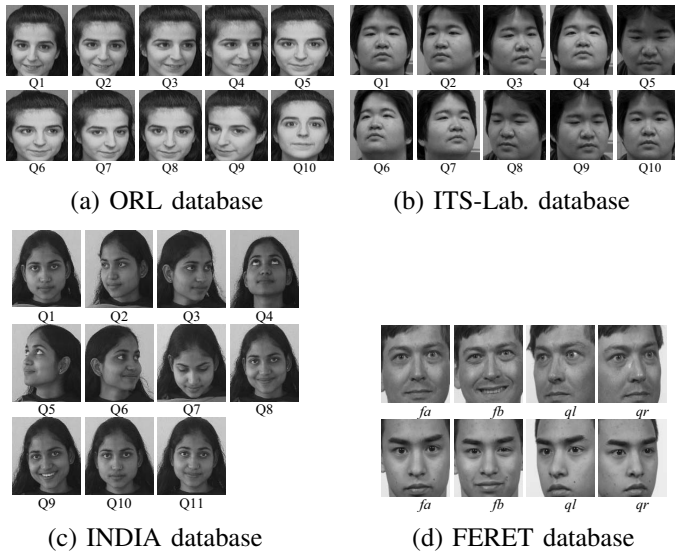
(a) ORL database   (b) ITS-Lab. database

(c) INDIA database   (d) FERET database

Fig. 3. Example of face pose variations of single subject of the tested databases.

(a) ITS. Lab. database   (b) INDIA database

(c) ORL. database   (d) FERET database

Fig. 4. CMS curve of proposed method compared to the CPCA.

TABLE II
THE OPTIMUM PROJECTION CRITERION $J(W)$ COMPARISON.

Method	ITS. LAB	INDIA	ORL	FERET
CPCA	0.1142	0.149	0.0656	0.171
APCA	3.33E+05	2.41E+06	9.36E+07	2.33E+06

TABLE I
THE COMPARISON OF THE RECOGNITION RATE OF THE PROPOSED
METHOD TO ESTABLISHED METHODS.

No	Methods	HF Dim.	Recognition Rate (%) of			
			ORL	ITS	INDIA	FERET
1	2DPCA	10x128	95.50	88.62	87.19	82.61
2	$(2D)^2$PCA	10x10	95.50	88.62	87.19	81.77
3	$(2D)^2$PCALDA	8x8	97.50	96.41	86.63	90.77
4	HF+2DPCA	8x12	96.75	88.82	88.32	82.10
5	HF+$(2D)^2$PCA	8x8	96.25	87.82	88.51	81.59
6	HF+CPCA	38	95.25	94.41	96.61	91.78
7	HF+APCA	38	**97.75**	**99.80**	**97.18**	**94.00**

A. The Performance of APCA vs. Established Methods

In the first experiments, the ORL, ITS-Lab., INDIA database, and FERET database were used to investigate the robustness of our proposed methods over the established PCA-based methods and Local Mean[18] as the lighting normalization. From the experimental results, as shown in the Table. I, the APCA tends to provide much better recognition rate that that of the established PCA-based methods with almost the same features dimension requirements for all tested databases. It can be achieved because the null space of the S_w which contains most discriminant information as shown by much higher discriminatory power of APCA than that of CPCA (see Fig. 1) is considered for face verification. These achievements confirm that the optimum criterion comparison as written in Eq. (12) is satisfied. In other words, it can be concluded that the more optimum W is, the more separable projection data will be obtained after the transformation.

The next experiment was performed to investigate deeply about the strength of the APCA against to CPCA and $(2D)^2$PCALDA. In this test, we used the cumulative matching score (CMS) parameter as the performance indicator and compared the APCA achievements with that of the $(2D)^2$PCALDA which is one of the best variation of PCA-based methods. The experimental results show that the APCA methods improve significantly the recognition rate of the
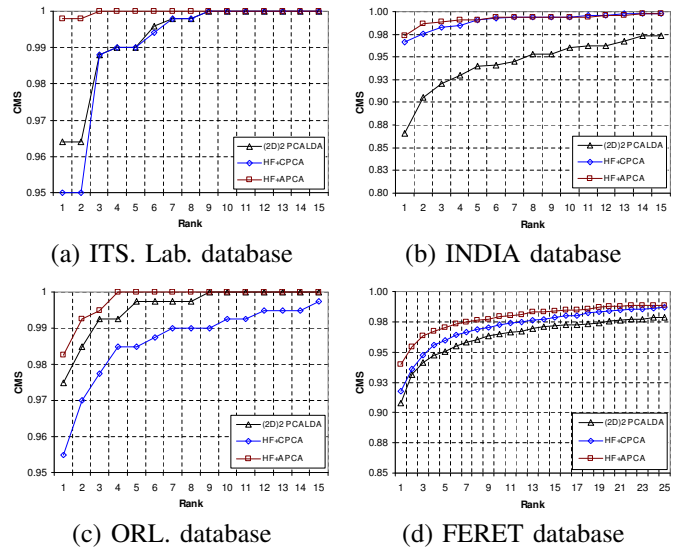
first rank of the CPCA and $(2D)^2$PCALDA for all tested databases, as shown in Fig. 4. This result explains and proves that the APCA method has better strength/ability in terms of features clustering than that of CPCA and its variation. These experimental results support the previous conclusion that the APCA has much higher optimum W than that of CPCA (see Table II), which is compatible with the mathematical analysis, as explained in sub section 3.B. Furthermore, these achievements also proves that optimum criterion ($J_{CPCA}(W) \ll J_{APCA}(W)$) is satisfied, as shown in Table II.

B. The Retraining Solution

In order to show that the proposed method can overcome the retraining problem, the last experiment was performed. It was carried out using data from FERET database, which consists of 508 face classes. The retraining was performed gradually. Firstly, 208 face classes were selected as initial training data and then incrementally 20 new face classes were added into the system until covered all of the face classes. In this case, we investigated the retraining time and recognition rate stability parameters and compared them with the CPCA and GSVD-ILDA[21]. The GSVD-ILDA-based recognition was developed to solve retraining problem of LDA for incremental data.

In terms of recognition rate, our proposed method gives higher and stable recognition rate than that of CPCA and the recent subspace method for incremental data (GSVD-ILDA), as shown in Fig. 5(a). This result supports our previous achievement, which proves that the APCA has better ability for features clustering with simpler computation complexity than that of CPCA. It is an alternative algorithm for features
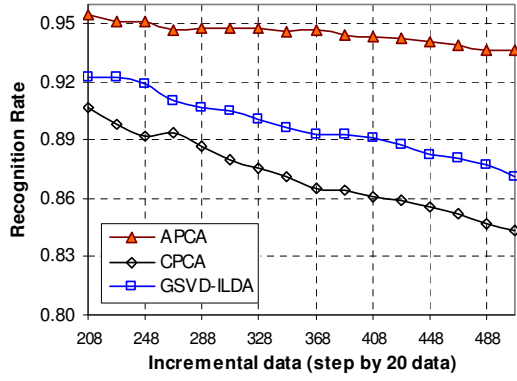
(a) Stability recognition rate.
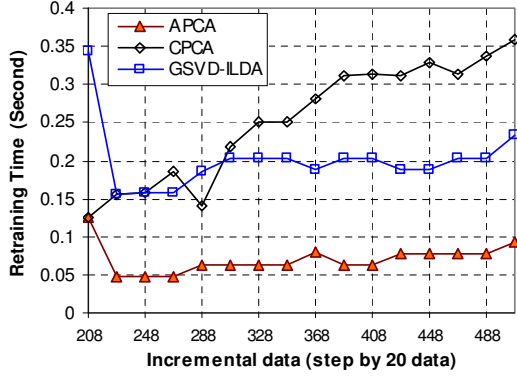
(b) Retraining time.

Fig. 5. The achievement of APCA compared to the CPCA and GSVD-ILDA.

clustering of incremental data, which requires much retraining process. In addition, the APCA also gives better and stable achievement than that of GSVD-ILDA because the optimum W of GSVD-ILDA is estimated by computing the best rank k-th approximation of the matrix $X = [A, B]$ for each incremental data B.

In terms of retraining process, our proposed method can solve the retraining problem of the CPCA method. The proposed method needs short retraining time when the training is performed gradually, as shown in Fig. 5(b). These results prove that APCA can crack the weakness of CPCA in terms of retraining problem. It can be done because the with-in class scatter matrix is obtained without considering the global means (mean of all samples) but just each class mean as described clearly in section 3.2. Compared to the GSVD-ILDA (see Fig. 5(b)), the APCA requires much less time for retraining because GSVD-ILDA has time complexity $O(nqk + n(c+p)t + q^2n + k^3)$, where t and k are number of selected leading principle sub matrix of SVD decomposition, as detail described in Ref. [21]. Even though the APCA has time complexity $O(n^3)$ for eigen analysis but the size of dimensional data input (n) is much less than number of data samples ($c+p$). In this test, the size of n is 53 elements while the number all classes is 508 and each class has 2 images, therefore $c+p=L$ is equal to 1016 images. It means the $n \ll c+p$ therefore the APCA needs shorter retraining time than GSVD-ILDA.

TABLE III
THE CONSUMPTION TIME COMPARISON.

Process	CPCA (s)		APCA (s)	
	L_2	MH	L_2	MH
Querying	0.0855	0.1058	0.0862	0.0925

Regarding computational cost for querying, the APCA requires shorter time than that of CPCA when the Mahalanobis distance is applied and almost the same time as that of PCA when Euclidean distance is applied for verification, as shown in Table III. It can be achieved because the Mahalanobis distance of APCA can be simplified, as follows.

Suppose x_T is features vector set of training samples, x_q is features vector of querying sample, and C_T is the global covariance of features vector set of the training samples.

$$d_i(\Delta x) = \Delta x C_T^{-1} \Delta x^T \tag{14}$$

where $\Delta x = x_i - x_q$, $x_i \in x_T$. The C_T of the APCA approach is diagonal ($W^T S_w W = \lambda$), therefore the Eq. (14) can be written as follows.

$$d_i(\Delta x) = \begin{bmatrix} \frac{\Delta x_1}{\lambda_1} & \frac{\Delta x_2}{\lambda_2} & \cdots & \frac{\Delta x_m}{\lambda_m} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \cdots \\ \Delta x_m \end{bmatrix}$$

$$= \sum_{i=1}^{m} \frac{(\Delta x_i)^2}{\lambda_i} \tag{15}$$

The computation complexity of Eq. (14) is almost the same as Euclidean distance because the Eq. (14) is the same as the normalized Euclidean distance. Moreover, the global covariance of APCA-based projection features is the same as eigen value of S_w of face features. Consequently, the APCA method does not need to recalculate it at all. Therefore, the computation time of the APCA with Mahalanobis distance is faster than that of CPCA. As notice, the experiments for time computation evaluation were performed on the PC with specification: Core-Duo Processor 1.7 GHz and 2 GB RAM.

VI. CONCLUSION AND FUTURE WORKS

From the mathematical analysis and experimental results, the within scatter-based PCA called as alternative PCA (APCA) has been proven that it gives better performance for both recognition rate and accuracy parameters than those of CPCA for all tested databases. Moreover, the APCA requires shorter processing time when it combines with Mahalanobis distance as verification and HF as dimensional reductions. In addition, the APCA can solve the retraining problem that is the main problem of the PCA-based face recognition for the incremental data. However, the APCA's performance in large data size is not good as in small data size because between-class projection features are worst separable after the transformation.

In order to get more precise verification result, we will consider more local features analysis involving eyes, nose, mouth, and context information of the face image. Because the APCA requires short processing time and can work for incremental data, therefore it will be potentially to be implemented as the real-time face image recognition.

REFERENCES

[1] W. Zhao, R. Chellappa, and A. Rosenfeld, "Face Recognition: A Literature Survey", ACM Computing Surveys, 35, pp.399-458, December 2003.

[2] W. Chen, E. Meng J., and S. Wu, "PCA and LDA in DCT Domain", Pattern Recognition Letter, 26, pp. 2474-2482, 2005.

[3] J. Meng, and W. Zhang, "Volume Measure in 2DPCA-Based Face Recognition", Pattern Recognition Letter, 28, pp. 103-108, 2007.

[4] IGPS. Wijaya, K. Uchimura, and Z. Hu, "Multipose Face Recognition Based on Frequency Analysis and Modified LDA", The journal of the IIEEJ, vol. 37, no. 3, pp. 231-243, May 2008.

[5] IGPS. Wijaya, K. Uchimura, and Z. Hu, "Improving the Pose Invariant Face Recognition Using Double Frequency Analysis and Color Information", International Session of 2008 joint Conference of IEEE Computer Society Fukuoka Chapter, Oita Japan, September 2008 (CD-ROM).

[6] IGPS. Wijaya, K. Uchimura, and Z. Hu, "Why the Alternative PCA Provide Better Performance for Face Recognition", 10th The International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), pp. 169-172, London, UK, May 2009.

[7] M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol.3, No. 1, pp. 71-86, 1991.

[8] J. Yang, D. Zhang, A.F. 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, vol. 26, no.1, pp. 131-137, January 2004.

[9] D. Zhang and Z-H. Zhou, "$(2D)^2PCA$: Two-Directional Two-Dimensional PCA for Efficient Face Representation and Recognition", Neurocomputing, 69, pp. 224-231, 2005.

[10] 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, 205, pp. 601-607, 2008.

[11] S. Noushath, G.H. Kumar, and P. Shivakumara, "$(2D)^2LDA$: An Efficient Approach for Face Recognition", Pattern Recognition, 39, pp. 1396-1400, 2006.

[12] K. Etemad, and R. Chellappa, "Discriminant analysis for Recognition of Human Face Images", J. opt. Soc. Am. A/vol. 14, no. 8, pp. 1724-1733, August 1997.

[13] Z.M. Hafed, and M.D. Levine, "Face Recognition Using the Discrete Cosine Transforms", International Journal of Computer Vision, 43(3), pp. 167-188, 2001.

[14] D-Q. Dai and P.C. Yuen, "Wavelet Based Discriminant Analysis for Face Recognition", Applied Mathematics and Computation, 175, pp. 307-318, 2006.

[15] R.C. Gonzalez, and R.E.. Woods, "Digital Image Processing", Third Edition, Pearson Prentice Hall, USA, pp. 839-842, 2008.

[16] J. Ruiz-del-Solar, and J. Quinteros, "Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of different pre-processing approaches", Pattern Recognition Letter, vol. 29, no. 14, pp. 1966-1979, 2008.

[17] IGPS. Wijaya, K. Uchimura, and Z. Hu, "Face Recognition Based on Dominant Frequency Features and Multiresolution Metric", International Journal of Innovative Computing, Information and Control, vol. 5, no. 3, pp. 641-651, March 2009.

[18] IGPS. Wijaya, K. Uchimura, and Z. Hu, "Improving the PDLDA Based Face Recognition Using Lighting Compensation", The Workshop of Image Electronics and Visual Computing 2010, Nice France, March 2010, (CDROM).

[19] P.J. Philips, H. Moon, S.A. Risvi, P.J. Rauss, P.J., "The FERET evaluation methodology for face recognition algorithms", IEEE Trans. Pattern Anal. Machine Intell., vol. 22, no. 10, pp. 1090-1104, 2000.

[20] F. Samaria and A. Harter, "Parameterization of a stochastic model for human face identification", 2nd IEEE Workshop on Applications of Computer Vision, Sarasota (Florida), pp. 138-142, 1994. (http://www.cl.cam.ac.uk/research/dtg/attarchive/face database.html).

[21] H. Zhao and P.C Yuen, "Incremental Linear Discriminant Analysis for Face Recognition," IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, vol. 38, no. 1, pp. 210-221, February 2008.

[22] http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/

I Gede Pasek Suta Wijaya received the B.Eng. degrees in Electrical Engineering from Gadjah Mada University in 1997, M.Eng. degrees in Computer Informatics System from Gadjah Mada university in 2001, and D.Eng. degrees in Computer Science from Kumamoto university, Japan in 2010. During 1998-1999 he worked in Toyota Astra Motor Company in Indonesia as Planning Production Control, and from 1999-2000, next, he worked as lecturer assistance in Yogyakarta National Technology College in Indonesia, and since 2000 up today, he has been full time lecturer and stays in Informatics Systems Laboratory in Electrical Engineering Department, Mataram University, Indonesia. His research interests are pattern recognition, artificial intelligence, and image processing application on computer vision.

Keiichi Uchimura received the B.Eng. and M.Eng. degrees from Kumamoto University, Kumamoto, Japan, in 1975 and 1977, respectively, and the Ph.D. degree from Tohoku University, Miyagi, Japan, in 1987. He is currently a Professor with the Graduate School of Science and Technology, Kumamoto University. He is engaged in research on intelligent transportation systems, and computer vision. From 1992 to 1993, he was a Visiting Researcher at McMaster University, Hamilton, ON, Canada. His research interests are computer vision and optimization problems in the Intelligence Transport System.

Gou Koutaki received the B.Eng., M.Eng., and Ph.D.degree from Kumamoto University, Kumamoto, Japan, in 2002, 2004, and 2007, respectively. From 2007 to 2010, he was with Production Engineering Research Laboratory, Hitachi Ltd. He is currently an Assistant Professor with the Graduate School of Science and Technology, Kumamoto University. He is engaged in research on image processing and pattern recognition of the Intelligence Transport System.