Relevance-Weighted $(2D)^2$LDA Image Projection Technique for Face Recognition

Waiyawut Sanayha and Yuttapong Rangsanseri

In this paper, a novel image projection technique for face recognition application is proposed which is based on linear discriminant analysis (LDA) combined with the relevance-weighted (RW) method. The projection is performed through 2-directional and 2-dimensional LDA, or $(2D)^2$LDA, which simultaneously works in row and column directions to solve the small sample size problem. Moreover, a weighted discriminant hyperplane is used in the between-class scatter matrix, and an RW method is used in the within-class scatter matrix to weigh the information to resolve confusable data in these classes. This technique is called the relevance-weighted $(2D)^2$LDA, or RW$(2D)^2$LDA, which is used for a more accurate discriminant decision than that produced by the conventional LDA or 2DLDA. The proposed technique has been successfully tested on four face databases. Experimental results indicate that the proposed RW$(2D)^2$LDA algorithm is more computationally efficient than the conventional algorithms because it has fewer features and faster times. It can also improve performance and has a maximum recognition rate of over 97%.

Keywords: Linear discriminant analysis (LDA), 2DLDA, 2-directional 2DLDA, $(2D)^2$LDA, face recognition.

Manuscript received Nov. 17, 2008; revised June 6, 2009; accepted June 24, 2009.

Waiyawut Sanayha (phone: +66 2326 4242, email: waiyawut.s@egat.co.th) and Yuttapong Rangsanseri (email: rangsanseri@yahoo.com) are with the Department of Telecommunications Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.

doi:10.4218/etrij.09.0108.0667

I. Introduction

Linear discriminant analysis (LDA) is one of the most popular linear projection techniques. It is a well-known feature extraction and data representation technique which is widely used in the areas of pattern recognition for feature extraction and dimension reduction. It finds the set of the largest discriminant projection vectors which can map high-dimensional samples onto a low-dimensional space. Principal component analysis (PCA) and LDA research (started in 1991 by Turk and Pentland [1]) presented the Eigenfaces method for the linear projection of face images onto a reduced dimension feature space. Belhumeur and others [2] presented a projection method based on Fisher's linear discriminant (FLD) in 1997. From 2000 to 2004, there was much work on the theory of PCA and LDA [3]-[9] which was motivated by the need to solve the small sample problem.

The objective of LDA is to find the optimal projection so that the ratio of the determinants of the between-class and within-class scatter matrices of the projected samples reaches its maximum. However, concatenating 2D matrices into a 1D vector leads to a very high-dimensional image vector, where it is difficult to evaluate the scatter matrices accurately due to its large size and the relatively small number of training samples. Furthermore, the within-class scatter matrix is always singular, making the direct implementation of the LDA algorithm an intractable task. To overcome these problems, a new technique called 2-dimensional LDA (2DLDA) was recently proposed. This method directly computes the eigenvectors of the scatter matrices without matrix-to-vector conversion. Thus, PCA and LDA were developed into the 2-dimensional space methods which are known as 2DPCA and 2DLDA, respectively [10]-[17].

The scatter matrices in 2DLDA are quite small compared to the scatter matrices in LDA. The size of the 2DLDA matrix is proportional to the width of the image. 2DLDA evaluates the scatter matrices more accurately and computes the corresponding eigenvectors more efficiently than LDA or PCA. However, the main drawback of 2DLDA is that it needs more coefficients for image representation than the conventional PCA and LDA-based schemes.

In addition to the basic 2DLDA method, a two-directional LDA has been proposed. It works by simultaneously combining 2DLDA applied to the row direction of face images with alternative 2DLDA applied to the column direction of face images. This is called $(2D)^2$LDA. Similarly, the 2-dimesional version of PCA is known as $(2D)^2$PCA. These interesting algorithms were first developed in 2005. Zhang and Zhou [18] developed the 2-directional 2DPCA or $(2D)^2$PCA, achieving the same or even higher recognition accuracy than 2DPCA. The main difference between $(2D)^2$PCA and existing 2DPCA is that the latter only works in the row direction of face images, while the former works simultaneously in the row and column directions of face images. The main advantage of $(2D)^2$PCA over 2DPCA is that far fewer coefficients are needed by $(2D)^2$PCA for face representation and recognition than are needed by 2DPCA. Nagabhushan and others [19] introduced the $(2D)^2$FLD method which has the advantage of higher recognition rates, a smaller memory requirement, and better computing performance than the standard PCA/2D-PCA/2D-FLD method. The major advantage of the proposed method is that it requires fewer coefficients for object/face image representation than the standard PCA/2D-FLD/2D-PCA because it works simultaneously in both row and column directions. Noushath and others [20] showed that $(2D)^2$LDA requires fewer coefficients and less computing time for face image representation and recognition than standard PCA, 2DPCA, and 2DLDA methods.

In recent years, LDA research has developed several refinements to improve its performance, such as using the weighting function with the between-class scatter matrix. The relevance-weighted (RW) method is also combined with the within-class scatter matrix. In addition, both weights are used to perform the best recognition process for the LDA and 2DLDA. The research on weighted conventional LDA started by Li and others [21] introduced a weighting factor for each pairwise scatter that enables integration of confusable information into the between-class covariance matrix. There are many possibilities in choosing weighting factors. Li and others considered a few of them depending on the Euclidean and Kullback-Leibler distances between classes when a single Gaussian approximation is used for each class. This is called weighted pairwise scatter linear discriminant analysis (WPS-LDA) transform. Lotlikar and Kothari [22] introduced the concept of fractional dimensionality and developed an incremental procedure called the fractional-step LDA (F-LDA) as the weighting function. Loog and others [23] introduced a weighted variant of the well-known K-class Fisher criterion associated with the LDA. It can be seen that the LDA weighs contributions of individual class pairs according to the Euclidian distance of the respective class means. An interesting subclass of these criteria is the approximate pairwise accuracy criteria (aPAC). Yu and others [24] redefined the between-class scatter by adding a weighted function according to the between-class distance, which helps to separate the classes as much as possible. At the same time, it projects the between-class scatter into the null space of the within-class scatter that contains the most discriminant information. Lu and others [25] combined the strengths of the direct LDA (D-LDA) [4] and F-LDA approaches in the proposed framework, which will hereafter be referred to as DF-LDA. Also, a weighting function is introduced into the proposed variant of D-LDA, so that a subsequent F-LDA step can be applied to carefully reorient the small sample size (SSS)-free subspace, which results in a set of optimal discriminant features for face representation. Price and Gee [26] proposed a new algorithm as direct weighted LDA or DW-LDA. It combines direct LDA (D-LDA) with weighted pairwise Fisher criteria. Wang and others [27] proposed the weighted two-dimensional maximum margin criterion (W2DMMC), which has an additional weighted parameter β that further broadens the margin. Wang and others [28] developed a multi-block 2DLDA (MB2DLDA) to apply to the sub-images instead of the whole image by weighting the 2DLDA feature of a block; thus, the verification performance is improved.

Another approach in the development of relevance-weighted LDA was introduced in 2005 by Tang and others [29] who incorporated the inter-class relationships as relevance weights into the estimation of the overall within-class scatter matrix. Liang and others [30] proposed a generalizing relevance-weighted LDA or GRW-LDA. When compared with the LDA-GSVD and Fisherfaces, the GRW-LDA can extract more powerful discriminatory features, thereby achieving the best performance using the least features. Chougdali and others [31] presented a relevance-weighted LDA and QR decomposition matrix analysis. However, all algorithms presented in this paper are linear methods. Since facial variations are mostly non-linear, LDA, LDA/QR and RW-LDA/QR projections only provide suboptimal solutions. Jarchi and Boostani [32] proposed a variant LDA method for the multi-class problem which redefined the between-class and within-class scatter matrices by incorporating a weight function into each of them.

Our research indicates that LDA has in fact been improved in the $(2D)^2$LDA method. In other words, the RW method can be used to achieve better face recognition performance than that of LDA or 2DLDA. Therefore, it is best to combine the $(2D)^2$LDA method with RW method in order to improve its discrimination performance. This algorithm will be hereafter referred to as the relevance-weighted $(2D)^2$LDA (RW$(2D)^2$LDA).

This paper is organized as follows. Section II presents background of the LDA and its development starting from the conventional LDA, the 2-dimensional LDA (2DLDA), an alternative 2DLDA, the 2-dimensional and 2-directional LDA, and application with the pairwise scatter approaches. Section III describes the proposed algorithm, which uses weighted and RW methods with the $(2D)^2$LDA method. Experimental results and discussion are presented in section IV. The conclusion is given in section V. Finally, future work is proposed in section VI.

II. LDA Background

By using the set of projection vectors determined by the LDA as the projection axes, all projected samples form the maximum between-class scatter and the minimum within-class scatter simultaneously in the projective feature space.

1. Conventional LDA

When using appearance-based methods, the 2D face image matrices must be first transformed into 1D image vectors. The LDA maps high dimensional samples of the projection vectors onto a low-dimensional space and computes eigenvectors in the underlying space that give the best discrimination among classes. More formally, given a number of independent features relative to which the data is described, LDA creates a linear combination of these which yields the largest mean differences between the desired classes. Mathematically speaking, for all the samples of all classes, two measures are defined. One measure is called the within-class scatter matrix, as given by

$$S_{\mathrm{w}} = \sum_{j=1}^{C} \sum_{i=1}^{N_j} (A_i^j - \mu_j)(A_i^j - \mu_j)^T, \tag{1}$$

where A_i^j is the i-th sample of class j, μ_j is the mean of class j, C is the number of classes, and N_j is the number of samples in class j. The other measure is called the between-class scatter matrix, as given by

$$S_{\mathrm{b}} = \sum_{j=1}^{C} (\mu_j - \mu)(\mu_j - \mu)^T, \tag{2}$$

where μ represents the mean of all classes.

Once S_{b} and S_{w} are computed, we compute the optimal projection axes, denoted by X, so that the total scatter of the projected samples of the training images is maximized. To maximize the total scatter of the projected samples we used Fisher's criterion:

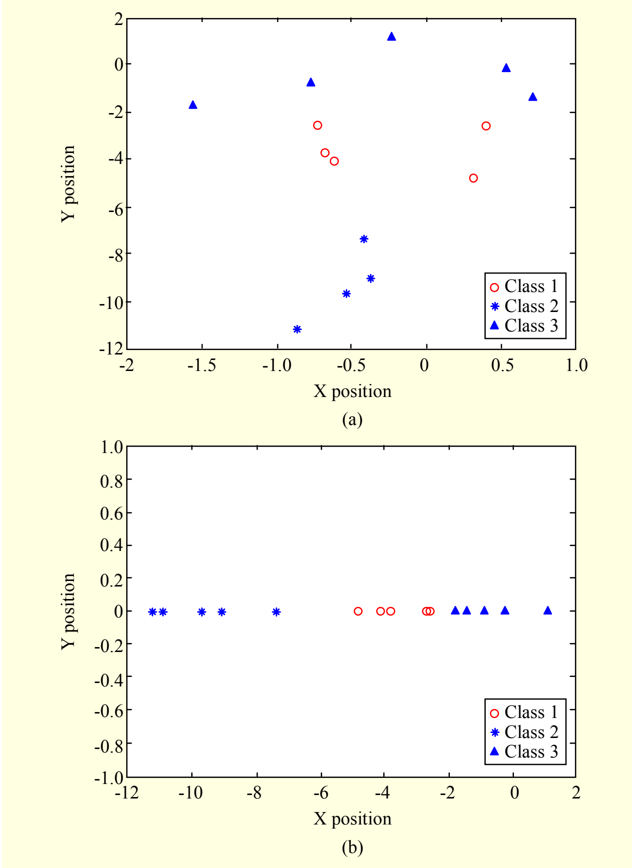

Fig. 1. LDA technique results: (a) data samples and (b) result of 1D projection.

$$J(X) = \frac{X^T S_{\mathrm{b}} X}{X^T S_{\mathrm{w}} X}. \tag{3}$$

The maximized $J(X)$ can be obtained by applying an eigenvector corresponding to the maximum eigenvalues of $S_{\mathrm{w}}^{-1}S_{\mathrm{b}}$. This is the optimal projection axis. It is usually not enough to have only one optimal projection axis, and we compute q projection axes, that is, $X_1, X_2, \cdots, X_q$, which are eigenvectors corresponding to the first q largest eigenvalues of $S_{\mathrm{w}}^{-1}S_{\mathrm{b}}$.

To show the principle of the LDA algorithm, example results of the LDA projection technique are shown in Fig.1. Example data of three classes is shown in Fig. 1(a), and 1D projection results using the LDA method are shown in Fig. 1(b).

2. Two-Dimensional LDA (2DLDA)

The LDA has two drawbacks when directly applied to the original input space. First, some non-face information such as image background data is misclassified when the face of the same subject is presented on different backgrounds. Secondly, when the SSS problem occurs, the within-class scatter matrix is singular. This is the so-called singularity problem. Projecting the high-dimensional input space into low-dimensional subspace via PCA can solve these LDA problems. Nevertheless, the spatial structure information is still lost. To overcome this drawback, the 2-dimensional LDA is based on 2D matrices rather than 1D vectors. This means that the image matrix does not need to be converted into a vector. As a result, the 2DLDA has two advantages: it is easier to evaluate the covariance matrix accurately, and it has lower time-consumption.

Let A_i^j be an image of size $a \times b$ representing the i-th sample in the j-th class. The between-class scatter matrix G_b and within-class scatter matrix G_w are computed as

$$G_b = \sum_{j=1}^{C} (\mu_j - \mu)^T (\mu_j - \mu), \tag{4}$$

$$G_w = \sum_{j=1}^{C} \sum_{i=1}^{N_j} (A_i^j - \mu_j)^T (A_i^j - \mu_j). \tag{5}$$

Once G_b and G_w are computed, we compute the optimal projection axes, denoted by X, so that the total scatter of the projected samples of the training images is maximized. To maximize the total scatter of the projected images we used Fisher's criterion as follows:

$$J(X) = \frac{X^T G_b X}{X^T G_w X}. \tag{6}$$

Thus, the eigenvectors of the final covariance matrix $G_w^{-1} G_b$ are computed, and then q eigenvectors corresponding to the first q largest eigenvalues of $G_w^{-1} G_b$ are chosen. Thus, the dimension of the optimal projection axes X is $b \times q$.

Projection of a training image onto these optimal projection axes results in a feature matrix of the respective training image. That is, we define the feature matrix of A_i^j as

$$Z_i^j = A_i^j X, \tag{7}$$

where Z_i^j is the feature matrix of dimension $a \times q$.

The proposed 2DLDA works in the row-wise direction as the image covariance matrix G_w is obtained by the outer products of the row vectors of the training images.

Training images are subsequently projected onto the optimal projection axes, and their dimension is reduced. During the query phase, query images are also projected onto the optimal projection axes to reduce their dimensionality, and they are subjected to a Euclidean nearest neighbors classifier to contrast them with the projected training images. The class label of the training images which is nearest to the query is retrieved as their class label. In this case, the class of the nearest training face to that of the query face image is identified as its class.

3. Alternative 2DLDA

Equation (5) reveals that the scatter matrix G_w can be obtained from the outer products of row vectors of images, assuming that the training images have a zero mean. For this reason, we claim that the original 2DLDA works in the row direction of images. Apparently, a natural extension is to use the outer product between column vectors of images to construct G_b and G_w. To devise an alternative 2DLDA, we propose that the between-class scatter matrix H_b and the within-class scatter matrix H_w be computed as

$$H_b = \sum_{j=1}^{C} (\mu_j - \mu)(\mu_j - \mu)^T, \tag{8}$$

$$H_w = \sum_{j=1}^{C} \sum_{i=1}^{N_j} (A_i^j - \mu_j)(A_i^j - \mu_j)^T. \tag{9}$$

It can be observed that H_b and H_w in (8) and (9) are obtained in this new formulation as outer products of column vectors, unlike G_b and G_w in (4) and (5) in the case of the original 2DLDA. Using these two scatter matrices, which are similar to the original 2DLDA, in this proposed model we also find the optimal projection axes W $(m \times q)$ so that the total scatter of the projected samples is maximized using the same Fisher's criterion given by

$$J(W) = \frac{W H_b W^T}{W H_w W^T}. \tag{10}$$

Thus, the eigenvectors of $H_w^{-1} H_b$ are computed, and then q eigenvectors corresponding to the first q largest eigenvalues of $H_w^{-1} H_b$ are chosen. Finally, projection of a training image onto these optimal projection axes results in a feature matrix of the respective training image. That is, if Z_i^j represents the feature matrix of A_i^j, then

$$Z_i^j = W^T A_i^j. \tag{11}$$

Equation (9) reveals that the image covariance matrix H_w can be obtained from the outer products of the column vectors of the training images, assuming that they have a zero mean. Therefore, the proposed alternative 2DLDA works in the column direction of images. An illustration of the 2DLDA projection technique is shown in Fig. 2.

4. $(2D)^2$LDA

Let X denote the $n \times d$ optimal projection matrix obtained in the original 2DLDA method as explained in section II.2, and let W denote the $m \times q$ matrix obtained by an alternative

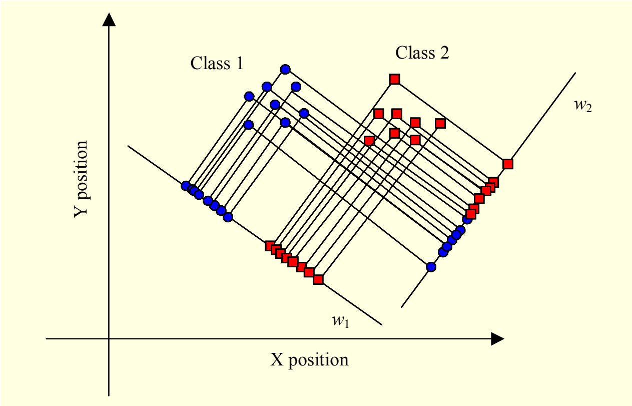

Fig. 2. 2DLDA projection technique.

2DLDA method as explained in section II.3. For $(2D)^2$LDA method, each training image A_i^j is projected onto both X and W simultaneously to obtain the respective feature matrix F_i^j, which is of dimensions $q \times d$ as follows:

$$F_i^j = W^T A_i^j X. \tag{12}$$

The matrix F is also called the coefficient matrix in image representation. When used for face recognition, the matrix F is also called the feature matrix. After projecting each training image A_i^j onto X and W, the feature matrices F_i^j can be obtained.

In a data set, distance or similarity relationships between pairs of classes is the important information for classification. Distance or similarity relationships, usually acquired through statistical approaches such as Euclidean distance, Mahalanobis distance, and so on, reflect how well two classes are separated in the feature space. In multi-class LDA, the relationships between pairs of classes are likely to be different from one pair to another. The classes that are closer to each other are potentially more confusing, and they should be given more attention during the feature extraction stage. The weighted LDA works by using the pairwise scatter matrix approach which only points at the between-class scatter matrix. Generally, the S_b of this pairwise scatter is computed by

$$S_b = \sum_{j=1}^{C-1} \sum_{k=j+1}^{C} (\mu_j - \mu_k)(\mu_j - \mu_k)^T. \tag{13}$$

Therefore, the 2DLDA and an alternative 2DLDA method can rewrite the between-class scatter matrix (4) and (8) as follows:

$$G_b = \frac{1}{N} \sum_{j=1}^{C-1} \sum_{k=j+1}^{C} N_j N_k (\mu_j - \mu_k)^T (\mu_j - \mu_k), \tag{14}$$

$$H_b = \frac{1}{N} \sum_{j=1}^{C-1} \sum_{k=j+1}^{C} N_j N_k (\mu_j - \mu_k)(\mu_j - \mu_k)^T. \tag{15}$$

Equation (14) is the between-class scatter matrix of 2DLDA which essentially works in the row-direction of images, and (15) is the between-class scatter matrix of an alternative 2DLDA which works in the column direction of images.

The within-class scatter matrix is still not changed because the pairwise scatter approach analyzes how the classes are discriminated from each other pairwise. Thus, it does not measure discrimination inside the within-class scatter matrix. However, we can rewrite the within-class scatter matrices (5) and (9) as

$$G_w = \frac{1}{N} \sum_{j=1}^{C} \sum_{i=1}^{N_j} N_j (A_i^j - \mu_j)^T (A_i^j - \mu_j), \tag{16}$$

$$H_w = \frac{1}{N} \sum_{j=1}^{C} \sum_{i=1}^{N_j} N_j (A_i^j - \mu_j)(A_i^j - \mu_j)^T. \tag{17}$$

Similar to (14) and (15), (16) is the within-class scatter matrix of 2DLDA which essentially works in the row-direction of images, and (17) is the within-class scatter matrix of an alternative 2DLDA which works in the column direction of images.

III. Proposed Algorithm

1. Weighted $(2D)^2$LDA

Because (14) and (15) are not directly related to classification accuracy and focus equally on every pair of classes, the outlier classes may negatively influence the estimation of the overall between-class covariance matrices G_b and H_b. Therefore, Loog and others [23] proposed an extended criterion named the *approximate pairwise accuracy criterion* (aPAC) which replaces (14) and (15) with

$$G_b = \frac{1}{N} \sum_{j=1}^{C-1} \sum_{k=j+1}^{C} L_{jk} N_j N_k (\mu_j - \mu_k)^T (\mu_j - \mu_k), \tag{18}$$

$$H_b = \frac{1}{N} \sum_{j=1}^{C-1} \sum_{k=j+1}^{C} L_{jk} N_j N_k (\mu_j - \mu_k)(\mu_j - \mu_k)^T, \tag{19}$$

where the weights L_{jk} are usually estimated based on relationships between classes j and k.

In order to keep enough discriminant information, we need to adjust the weights. A natural candidate is a normalization weight equal to the square of the inverse of the Euclidean distance between class means:

$$L_{jk} = \frac{1}{\left\| \mu_j - \mu_k \right\|^2} = \frac{1}{\left(\mu_j - \mu_k\right)^T \left(\mu_j - \mu_k\right)}. \tag{20}$$

Here, L_{jk} is defined as the dissimilarity between class j and k or how well classes j and k are separated in the original space.

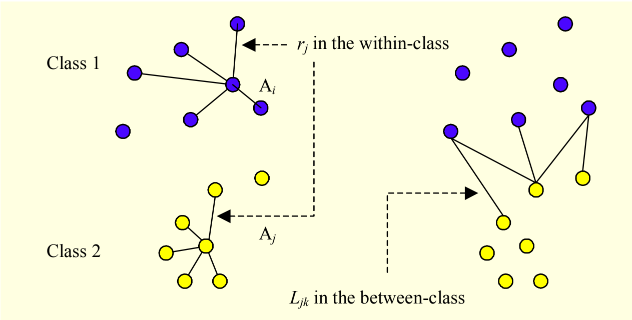

Fig. 3. Adjacency relationships in weighted and relevance-weighted calculation.

2. Relevance-Weighted (2D)^{2}LDA

We can see that in addition to assigning different considerations to classes when estimating the between-class covariance matrix, a weighting scheme should also be employed when estimating the within-class covariance matrix. To reduce the influence of outlier classes, (16) and (17) are modified to give

$$G_{\mathrm{w}} = \frac{1}{N}\sum_{j=1}^{C}\sum_{i=1}^{N_j} r_j N_j \left(A_i^j - \mu_j\right)^T \left(A_i^j - \mu_j\right), \tag{21}$$

$$H_{\mathrm{w}} = \frac{1}{N}\sum_{j=1}^{C}\sum_{i=1}^{N_j} r_j N_j (A_i^j - \mu_j)(A_i^j - \mu_j)^T, \tag{22}$$

where r_j is the relevance-based weight. By integrating r_j in (16) and (17), we intend to ensure that if class i is an outlier class, it only has a slight influence on the estimated G_{w} and H_{w}. This is reasonable because if one class is well separated from the other classes in the data set, the within-class covariance matrix of this class in the new space is compact, and it will not have much influence on the classification.

To calculate a class's separation from other classes, a straightforward weighting function is defined:

$$r_j = \sum_{j \neq k} \frac{1}{L_{jk}}. \tag{23}$$

We normalize r_j so that the largest weight is 1. Although several dissimilarity measures have been proposed in the past, it is impossible to choose one of them as the best measure independent of the data set.

We weight the classes which have their means closer to each other more heavily than those that have means further apart. In this sense, more confusable classes are weighted more heavily, and less confusable classes are weighted more lightly. The weighted between-class relationships and the relevance-weighted within-class relationships are shown in Fig. 3.

3. Classification

Given a test face image A, we first use (12) to get the feature


Fig. 4. Sample images from the FERET database.

matrix F_i^j. Then, we use a nearest neighbor classifier for classification. Here, the distance between $F1_i^j$ and $F2_i^j$ is defined by

$$\left\| F1_i^j - F2_i^j \right\| = \sqrt{\sum_{i=1}^{q}\sum_{j=1}^{d}\left(F1_i^j - F2_i^j\right)^2}. \tag{24}$$

IV. Experimental Results and Discussion

In our work, we have used four face databases, namely, the Facial Recognition Technology Database (FERET) [33]; the CMU Pose, Illumination, and Expression (PIE) Database [34]; the ORL Face Database [35]; and the YALE Face Database [36]. All of our experiments were carried out on a PC with an Intel Core2Duo CPU E6750 at 2.66 GHz with 1.96 GB of RAM memory under a MATLAB R2007b platform.

1. RW(2D)^{2}LDA Performance Test

We first used the FERET face database, which is designed to advance the state of the art in face recognition, with the collected images directly supporting both algorithm development and the FERET evaluation tests. The database is divided into a development set which is provided to researchers and a set of sequestered images for testing. The dataset tested included 2,413 still facial images of 856 individuals. Some sample FERET face images are shown in Fig. 4.

The results of our first accuracy test of the RW(2D)^{2}LDA algorithm are presented in a 3D view. We tested the recognition rate in row and column feature dimensions. The results of this performance test are shown in Fig. 5.

The number of feature dimensions was varied from 0 to 60. The RW(2D)^{2}LDA accuracy is over 70% when tested with 2-3 dimensions and over 97% at 5-8 dimensions. It becomes less steady when there are over 10 feature dimensions.

2. Pose, Illumination, and Expression Condition Test

It is not easy to compare all three algorithms within a 3D view graph because the lines of the graph would be

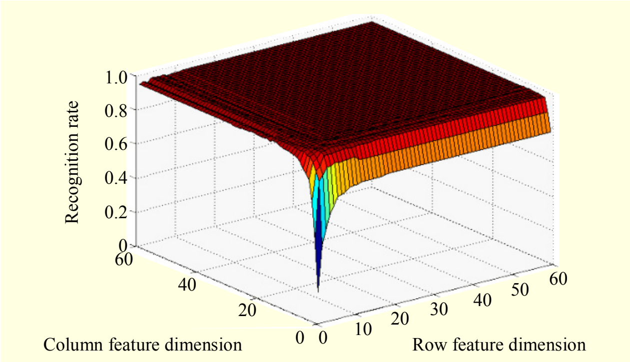

Fig. 5. RW(2D)^{2}LDA accuracy in 3D view.

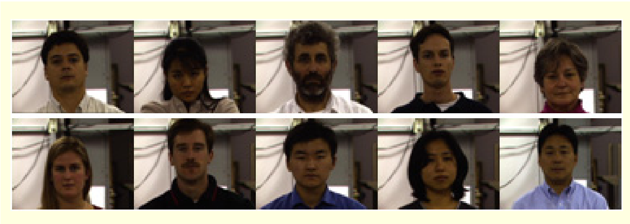

Fig. 6. Sample images from the PIE database.

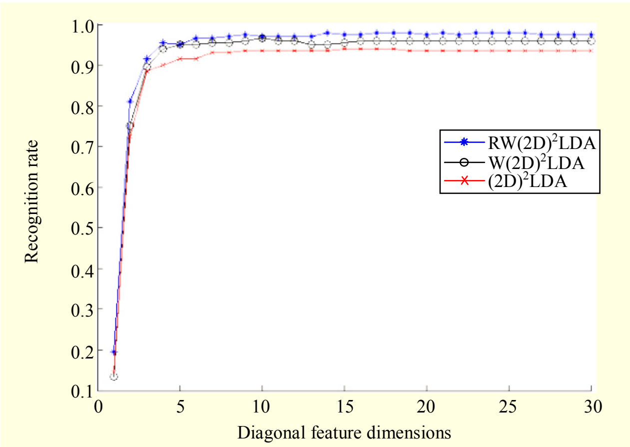

Fig. 7. Result of pose, illumination, and expression test in 2D.

complicated. Therefore, it is more suitable to use a 2D view graph for accurate comparisons between these algorithms. For this reason, we conducted a second experiment to show the performance improvement of the proposed method (RW(2D)^{2}LDA and W(2D)^{2}LDA) compared with the previous method ((2D)^{2}LDA).

We used the PIE face database for pose, illumination, and expression condition tests. This database consists of 41,368 images of 68 individuals. The PIE database was able to image each person under 13 different poses, 43 different illumination conditions, and with 4 different expressions. Some sample PIE face images are shown in Fig. 6.

As seen in Fig. 7, the results of this test demonstrate that, for

Table 1. Result of pose, illumination, and expression test.

Feature dimensions	Methods		
	(2D)^{2}LDA	W(2D)^{2}LDA	RW(2D)^{2}LDA
1	14.50	13.50	19.50
2	72.50	75.00	81.00
3	88.50	89.50	91.50
4	90.00	94.00	95.50
5	91.50	95.00	95.00
6	91.50	95.00	96.50
7	93.00	95.50	96.50
8	93.00	95.50	97.00
9	93.50	96.00	97.50
10	93.50	96.50	97.00

the accuracy of training groups, the RW(2D)^{2}LDA and W(2D)^{2}LDA methods can improve the performance of the previous (2D)^{2}LDA method.

The interesting aspect of the RW(2D)^{2}LDA and W(2D)^{2}LDA methodologies is that they are more computationally efficient than the original formulation of (2D)^{2}LDA because they have fewer features and are more accurate.

Table 1 reveals the recognition accuracy obtained by three methods when the number of diagonal feature dimensions varies from 1 to 10. The W(2D)^{2}LDA and RW(2D)^{2}LDA methods can increase the performance by up to 2.10% and 8.11%, respectively, when compare with the (2D)^{2}LDA.

3. Facial Scaling and Rotation Condition Test

Next, we experimentally evaluated our proposed RW(2D)^{2}LDA, W(2D)^{2}LDA, and (2D)^{2}LDA methods on well-known face databases. The first of these is the ORL database. It is used to test the performance of face recognition algorithms under the condition of minor variation of scaling and rotation.

The ORL database contains 400 images of 40 individuals. These images were captured at different times and have different variations including expression (open or closed eyes, smiling or not smiling) and facial details (glasses or no glasses). The images were taken with a tolerance for some tilting and rotation of the face up to 20 degrees. All images are in grayscale and normalized to the resolution of pixels and histogram equilibrium in the preprocessing step. Some sample images from the ORL database are shown in Fig. 8. In our experiments, we split the whole database into two equal parts. The first five images of each class were used for training, and

Fig. 8. Sample images from the ORL database.

Table 2. Result of scaling and rotation.

Methods	Top recognition	Dimensions	Running time
$(2D)^2$LDA	98.50	8×8	4.14
$W(2D)^2$LDA	98.75	7×8	5.26
$RW(2D)^2$LDA	99.15	6×7	7.85

Fig. 9. Sample images from the YALE database.

the rest of the images were used for testing.

Table 2 compares the three methods in term of their top recognition accuracy and gives the corresponding dimensions of the feature matrices and running times. It can be found that the top recognition accuracy of the original $(2D)^2$LDA method is comparable to that of the other methods. This table also reveals that the top recognition accuracy of the proposed $RW(2D)^2$LDA and $W(2D)^2$LDA methods is significantly higher than that of the existing $(2D)^2$LDA method despite the $RW(2D)^2$LDA having a reduced feature matrix. Conversely, the $RW(2D)^2$LDA and $W(2D)^2$LDA methods consume more running time than the $(2D)^2$LDA.

4. Facial Expression and Lighting Condition Test

Finally, we compared the proposed $RW(2D)^2$LDA and $W(2D)^2$LDA methods with the $(2D)^2$LDA using the YALE face database. We tested the performance under the condition of minor variations of facial expression and lighting conditions. The images are in grayscale with white background and are 320×243-pixels in size. We used a total of 165 images of 15 individuals in GIF format. There are 11 images per subject, and each image shows a different facial expression or lighting configuration. Some sample images from the YALE database are shown in Fig. 9.

Table 3. Result of facial expression and lighting conditions.

Methods	Number of training samples per class			
	2	4	6	8
$(2D)^2$LDA	86.25	93.60	94.67	96.00
$W(2D)^2$LDA	87.36	95.00	96.50	98.45
$RW(2D)^2$LDA	89.57	96.48	97.20	99.40

Table 3 shows the top recognition accuracy under the condition of minor variation of facial expression and lighting conditions obtained by three methods for varying numbers of training samples. The proposed $RW(2D)^2$LDA and $W(2D)^2$LDA methods are comparable to the $(2D)^2$LDA method in terms of recognition accuracy.

V. Conclusion

This paper examined two efficient face recognition methods called the weighted $(2D)^2$LDA or $W(2D)^2$LDA, and the relevance-weighted $(2D)^2$LDA or $RW(2D)^2$LDA incorporating weighted outlier class relationships into the estimation of the overall between-class scatter matrix. In the same way, relevance-weighted inner class relationships are incorporated into the overall within-class scatter matrix to improve the performance of the $(2D)^2$LDA method. The experimental results have shown that our algorithms can be used under various conditions and can achieve better performance accuracy than existing methods.

VI. Future Work

The proposed $RW(2D)^2$LDA method applies the weighting function in the within-class scatter matrix, but the $W(2D)^2$LDA method applies the weighting function in the between-class scatter matrix. It would be good to combine the proposed methods with other methods to improve the performance of the $(2D)^2$LDA. However, the effectiveness of kernel-based methods has been reported recently. It may be interesting to focus on the problem of kernelizing an existing supervised Mahalanobis distance, such as the *kernelizing Mahalanobis distance learning algorithm*. Therefore, we plan to apply neighborhood component analysis, large margin nearest neighbors, and discriminant neighborhood embedding to this algorithm. An alternative kernelization framework will also be used in the between and within-class scatter matrix in our future work.

References

[1] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, 1991, pp. 71-86.

[2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kiregman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, 1997, pp. 711-720.

[3] L. Chen et al., "A New LDA-Based Face Recognition System Which Can Solve the Small Sample Size Problem," *Pattern Recognition,* vol. 33, no. 10, 2000, pp. 1713-1726.

[4] H. Yu and J. Yang, "A Direct LDA Algorithm for High-Dimensional Data—with Application to Face Recognition," *Pattern Recognition,* vol. 34, no. 10, 2001, pp. 2067-2070.

[5] A.M. Martinez and A.C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 2, 2001, pp. 228-233.

[6] H-C. Kim, D. Kim, and S.W. Bang, "Face Recognition Using LDA Mixture Model," *Pattern Recognition Letters,* vol. 24, no. 15, 2003, pp. 2815-2821.

[7] M. Ordowski and G. Meyer, "Geometric Linear Discriminant Analysis for Pattern Recognition," *Pattern Recognition Letters*, vol. 37, no. 3, 2004, pp. 421-428.

[8] X.-Y. Jing, D. Zhang, and Y.-F. Yao, "Improvements on the Linear Discriminant Technique with Application to Face Recognition," *Pattern Recognition Letters,* vol. 24, no. 15, 2003, pp. 2695-2701.

[9] F. Song et al., "A Parameterized Direct LDA and Its Application to Face Recognition," *Neurocomputing,* vol. 24, 2003, pp. 2695-2701.

[10] J. Yang et al., "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, 2004, pp, 131-137.

[11] J. Meng and W. Zhang, "Volume Measure in 2DPCA-Based Face Recognition," *Pattern Recognition Letters,* vol. 28, no. 10, 2007, pp. 1203-1208.

[12] Y. Wen and P. Shi, "Image PCA: A New Approach for Face Recognition," *Proc. IEEE Inter. Conf. on Acoustics, Speech, and Signal Processing (ICASSP),* vol. 1, 2007, pp. I-1241-1244.

[13] H. Xiong, M.N.S. Swamy, and M.O. Ahmad, "Two-Dimensional FLD for Face Recognition," *Pattern Recognition,* vol. 38, no. 7, 2005, pp. 1121-1124.

[14] J. Yang et al., "Two-Dimensional Discriminant Transform for Face Recognition," *Pattern Recognition*, vol. 38, no. 7, 2005, pp. 1125-1129.

[15] M. Li and B. Yuan, "2D-LDA: A Statistical Linear Discriminant Analysis for Image Matrix," *Pattern Recognition Letters*, vol. 26, no. 5, 2005, pp. 527-532.

[16] S. Yan et al., "Multilinear Discriminant Analysis for Face Recognition," *IEEE Trans. Image Processing*, vol. 16, 2007, pp. 212-220.

[17] D.S. Guru and T.N. Vikram, "2D Pairwise FLD: A Robust Methodology for Face Recognition," *IEEE Workshop on Automatic Identification Advanced Technologies*, 2007, pp. 99-102.

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

[19] P. Nagabhushan, D.S. Guru, and B.H. Shekar, "$(2D)^2$ FLD: An Efficient Approach for Appearance Based Object Recognition," *Neurocomputing,* vol. 69, 2006, pp. 934-940.

[20] S. Noushath, G. Hemantha, and P. Shivakumara, "$(2D)^2$LDA: An Efficient Approach for Face Recognition," *Pattern Recognition,* vol. 39, no. 7, 2006, pp. 1396-1400.

[21] Y. Li, Y. Gao, and H. Erdogan, "Weighted Pairwise Scatter to Improve Linear Discriminant Analysis," *Proc. 6th Inter. Conf. Spoken Language Processing* (*ICSLP*), vol. 4, 2000, pp. 608-611.

[22] R. Lotlikar and R. Kothari, "Fractional-Step Dimensionality Reduction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, 2000, pp. 623-627.

[23] M. Loog, R.P.W. Duin, and R. Haeb-Umbach, "Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 7, 2001, pp. 762-766.

[24] B. Yu, L. Jin, and P. Chen, "A New LDA-Based Method for Face Recognition," *Proc. IEEE Int. Conf. Pattern Recognition*, vol. 1, 2002, pp. 168-171.

[25] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face Recognition Using LDA-Based Algorithms," *IEEE Trans. Neural Networks,* vol. 14, no. 1, 2003, pp. 195-200.

[26] J.R. Price and T.F. Gee, "Face Recognition Using Direct, Weighted Linear Discriminant Analysis and Modular Subspaces," *Pattern Recognition,* vol. 38, no. 2, 2005, pp. 209-219.

[27] H. Wang, S. Chen, and Z. Hu, "Image Recognition Using Weighted Two-Dimensional Maximum Margin Criterion," *Proc. 3rd Int. Conf. Natural Computation (ICNC),* vol. 1, 2007, pp. 582-586.

[28] J. Wang, E. Sung, and W. Yau, "Weighting Multi-Block Two-Dimensional Linear Discriminant Features for Face Verification," *Proc. 6th Int. Conf. on Information, Communications and Signal Processing (ICICS)*, 2007, pp. 1-5.

[29] E.K. Tang et al., "Linear Dimensionality Reduction Using Relevance Weighted LDA," *Pattern Recognition*, vol. 38, no. 4, 2005, pp. 485-493.

[30] Y. Liang et al., "Generalizing Relevance Weighted LDA," *Pattern Recognition*, vol. 38, no. 11, 2005, pp. 2217-2219.

[31] K. Chougdali, M. Jedra, and N. Zahid, "Face Recognition Using Relevance Weighted LDA with QR Decomposition," *J.*

Graphics, Vision and Image Processing (GVIP), vol. 6, no. 2, 2006, pp. 27-31.

[32] D. Jarchi and R. Boostani, "A New Weighted LDA Method in Comparison to Some Versions of LDA," *Proc. World Academy of Science, Engineering and Technology (WASET)*, vol. 18, 2006, pp. 233-238.

[33] P.J. Phillips et al., "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, 2000, pp. 1090-1104.

[34] The CMU Technical Report CMU-RI-TR-01-02, "The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces."

[35] F.S. Samaria and A.C. Harter, "Parameterisation of a Stochastic Model for Human Face Identification," *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, 1994, pp. 138-142.

[36] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, 2001, pp. 643-660.

Waiyawut Sanayha received the BInd degree in instrumentation technology from King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1986, the MEng degree in electrical engineering from Prince of Songkla University, Thailand, in 1997, and the MSc degree in information technology from Rangsit University, Thailand, in 2003. He is currently working toward the DEng with the Department of Telecommunications Engineering at KMITL. His research interests include pattern recognition in the areas of character and face recognition.

Yuttapong Rangsanseri received the BEng and the MEng in electrical engineering from King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1985 and 1987, respectively, and the PhD in engineering from the Institute National Polytechnique de Grenoble, France, in 1992. He is currently an associate professor with the Department of Telecommunications Engineering at KMITL. His research interests include image processing, pattern recognition, and multimedia signal processing.