International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 9, No.2 (Mar-2020)
http://dx.doi.org/10.12785/ijcds/090218

Human Face Recognition Using Class-wise Two-dimensional Principal Component Analysis

Falah Alsaqre[1]

[1]*Al-Hikma University College, Department of Computer Engineering Techniques, Baghdad, Iraq*

Received 19 Jun. 2019, Revised 19 Jan. 2020, Accepted 23 Feb. 2020, Published 01 Mar. 2020

Abstract: Two-dimensional principal component analysis (2DPCA) and its variants have been successfully used for the task of face recognition (FR). However, one of the major limitations of 2DPCA-based FR methods is that they only consider the holistic information of a given training dataset, ignoring both class-specific discriminant information and class-separation components, which could further improve recognition performance. To address this limitation, this paper suggests a class-wise 2DPCA (CW2DPCA) framework that seeks to model class-specific subspaces, where each subspace retains the discriminatory information of a particular class, as well as class separability information. In this way, CW2DPCA not only feeds discriminative representations of facial images to the classification model, but also enables a high degree of separation between the different classes present in the training dataset. Experimental evaluation on two face datasets proved the effectiveness of the proposed CW2DPCA in FR.

Keywords: Face recognition, Two-Dimensional Principal Component Analysis (2DPCA), Feature Extraction.

1. INTRODUCTION

Face recognition (FR) is an important and common task within a wide range of pattern recognition and computer vision applications, including human-computer interaction [1], biometrics [2], and visual surveillance [3]. The definitive goal of FR systems is to automatically specify the correct class membership of unknown human faces through extrapolation from training datasets. While there are various approaches to developing FR systems [4][5], subspace methods, particularly those based on principal component analysis (PCA), have attracted intense interest over the last few decades due to their algorithmic simplicity and efficacy [6]–[8].

In the context of FR, PCA is known as an eigenface method used to produce a compact set of principal components or eigenfaces of the original training dataset. In other words, PCA applies a linear transformation to the facial training samples and extracts a new subspace (eigenspace) that holds the key features of these samples. The fundamental assumption in PCA-based methods is that all the facial samples under consideration must be unfolded into column vectors, which, however, induce extensive computational costs and usage of space. Moreover, the use of vectorized facial images results in disregarding the fact that such images very often possess specific spatial structures. In order to obviate these implications, Yang et al. [9] developed the two-dimensional PCA (2DPCA) method, where the feature space is specified through a direct use of 2D face images. In essence, 2DPCA learns a projection matrix reflecting the underlying structural information in the row direction of image-as-matrix training instances. This entails that once an image covariance (scatter) matrix is obtained, the projection matrix can be formed from its orthonormal eigenvectors that correspond to the dominant eigenvalues. Besides being less computationally demanding compared to PCA, 2DPCA allows for the spatial relationships between facial images to be preserved and more structural information to be incorporated into the extracted features, hence endowed with significantly better performance. However, one shortcoming of 2DPCA is that it uses more feature coefficients than the classical PCA for representing image content [10]. Since it was first introduced, 2DPCA has been widely reported as an effective, yet simple, technique for FR [11]–[13].

Inspired by its success in face recognition and representation, many variants of 2DPCA [14]–[23] have been developed in an effort to improve its performance. The three notable examples are perhaps the bilateral-projection-based 2DPCA (B2DPCA) [14], the two-directional 2DPCA ($(2D)^2$PCA) [15], and the diagonal PCA (DiaPCA) [16]. The B2DPCA has been formulated as a more general framework that employs a bilateral projection scheme, instead of unilateral projection scheme as in 2DPCA, in order to cut off redundant information in both directions of 2D images, bringing down the number of coefficients within the feature matrices. Similarly,

E-mail: alsaqre@ieee.org

$(2D)^2$PCA projects each facial image onto two separate (left and right) projection matrices simultaneously; the left one is obtained via alternative 2DPCA (A2DPCA) [15] and the right one via 2DPCA. Note that both B2DPCA and $(2D)^2$PCA achieve very similar recognition results if the dimensionalities of the counterpart projection matrices are equal. In DiaPCA, the image scatter matrix is evaluated from the diagonalized facial images, such that the row and column information can be intermingled to exploit some useful local structures carried by the original images. Later, a group of methods have been developed as extensions of or modifications to $(2D)^2$PCA, such as block-wise two-directional 2DPCA ($B(2D)^2$PCA) [21] and sequential row-column 2DPCA (RC2DPCA) [17]. In parallel, other methods have focused on minimizing the reconstruction error in 2DPCA by replacing its default error measurement metric (L2-norm) with L1-norm [20], Lp-norm [22], or F-norm [23]. Thus, such methods are favorable for image compression.

2DPCA and its variants share the same general structure: a set of feature matrices representing facial images and a distance function evaluating the similarities between the testing and training images in the feature space. Notwithstanding their apparent usefulness in providing a compact representation of a given dataset, 2DPCA methods do not lend themselves to identifying the discriminative features of each class (subject), i.e., those features that capture class-specific invariant characteristics, with which further improvement in recognition accuracy can be brought. For instance, in 2DPCA, while this representation is valuable from an image compression perspective, it means that each class contributes equally to the calculation of the image scatter matrix. Consequently, each column entry of the learned projection matrix is a linear combination of all the training images, making the identification of class-specific discriminant information almost impossible. In addition, as the elements of the projection matrix are evenly involved in the computation of the feature matrices, a test feature matrix may tend to correspond to different classes present in the dataset, especially when the number of eigenvectors that make up the projection matrix is relatively large. This could partly explain why the recognition accuracy of 2DPCA diminishes, to some extent, as additional eigenvectors are included in the projection matrix. Therefore, the features extracted by 2DPCA methods are suboptimal from an image recognition standpoint.

In order to enhance the recognition capability of 2DPCA, this paper proposes a class-wise 2DPCA (CW2DPCA) framework, and demonstrates its effectiveness in FR. The premise behind CW2DPCA is to construct class-wise projection matrices, each holding the invariant characteristics of a particular class and class-separation components. Consistent with this premise, the classification is performed in class-specific subspaces generated by the constructed projection matrices. Under this framework, the classification task is performed based on the rule that the feature representation of a test image belonging to a specific subject is more likely to lie close to that of the training set from the same subject. A nearest neighbor (NN) classifier is deployed to accomplish this task.

It is worthwhile to point out that there are very few attempts to exploit the idea of face-specific/class-wise subspaces representation for developing FR methods. Shan et al. [24] proposed to use face-specific subspaces modeled by applying PCA to each subject independently. Apart from the inherent limitations of the PCA itself, this approach is far from being practical because it requires enlargement of the training set of each subject with additional or virtually derived facial images, so as to be able to infer sufficient and well representative face-specific principal components. Recently, Turhan and Bilge [25] tried to develop FR method based on class-wise 2DPCA operated in Haar-like space, where the projection matrix for each particular class is found by minimizing the within-class scatter. Admittedly, the minimization of within-class scatter does not yield any discriminative or class-separation information, and it only reflects classes distribution [26].

The rest of this paper is structured as follows: Section 2 presents a brief review of the related work; details of the proposed CW2DPCA are offered in Section 3; the experimental results are reported in Section 4, and, finally, Section 5 concludes the paper.

2. BRIEF REVIEW OF RELATED WORK

A. Two-dimensional PCA (2DPCA)

2DPCA [9] attempts to learn a projection matrix whose column vectors maximize the scatter of the projected training instances. Note here that these column vectors are the orthonormal eigenvectors of the image scatter/covariance matrix associated with the dominant eigenvalues. More formally, let $\boldsymbol{F} = \{\boldsymbol{F}_i\}_{i=1}^{c}, \boldsymbol{F}_i \in \mathbb{R}^{m\times n}$, be a training dataset composed of c facial samples divided into exclusive classes. The image covariance matrix ($\boldsymbol{G}_{2DPCA}$) is computed as follows:

$$\boldsymbol{G}_{2DPCA} = \frac{1}{c}\sum_{i=1}^{c}(\boldsymbol{F}_i - \overline{\boldsymbol{F}})^T\,(\boldsymbol{F}_i - \overline{\boldsymbol{F}}), \#(1)$$

where $\overline{\boldsymbol{F}}$ is the total mean of the training dataset and $\boldsymbol{G}_{2DPCA} \in \mathbb{R}^{n\times n}$. It follows, by performing eigenvalue decomposition on the $\boldsymbol{G}_{2DPCA}$ matrix, that the selected d most dominant eigenvectors form the projection matrix $\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_d] \in \mathbb{R}^{n\times d}$. This projection is then used to transform the training images into feature matrices, producing the following set of training feature matrices:

$$\boldsymbol{Y} = \{\boldsymbol{Y}_i\}_{i=1}^{c}, \qquad \boldsymbol{Y}_i = \boldsymbol{F}_i\boldsymbol{U} \in \mathbb{R}^{m\times d}. \#(2)$$

To classify a given test sample $\boldsymbol{R} \in \mathbb{R}^{m\times n}$, its feature matrix $\boldsymbol{X} = \boldsymbol{R}\boldsymbol{U} \in \mathbb{R}^{m\times d}$ is matched against all the training feature matrices and is assigned the identity of the class that has the highest similarity score, most often via an NN technique.

B. Two-directional 2DPCA ((2D)^{2}PCA)

In [15], 2DPCA is extended to operate in both row and column directions of the sample images, and is termed two-directional 2DPCA ((2D)^{2}PCA). The goal of (2D)^{2}PCA is to find two (left and right) projection matrices to project facial images onto a common space. The right projection matrix is actually the $\boldsymbol{U}$ of 2DPCA, while the left, $\boldsymbol{V} = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_q] \in \mathbb{R}^{m\times q}$, is formed by the q most dominant eigenvectors of the alternative image covariance matrix ($\boldsymbol{G}_{A2DPCA}$), which is given by

$$\boldsymbol{G}_{A2DPCA} = \frac{1}{c}\sum_{i=1}^{c}(\boldsymbol{F}_i - \overline{\boldsymbol{F}})(\boldsymbol{F}_i - \overline{\boldsymbol{F}})^T, \#(3)$$

where $\boldsymbol{G}_{A2DPCA} \in \mathbb{R}^{m\times m}$. Like 2DPCA, a set of c training feature matrices can be obtained as follows:

$$\boldsymbol{Y} = \{\boldsymbol{Y}_i\}_{i=1}^{c}, \qquad \boldsymbol{Y}_i = \boldsymbol{V}^T\boldsymbol{F}_i\boldsymbol{U} \in \mathbb{R}^{q\times d}. \#(4)$$

For any test image $\boldsymbol{R}$, the feature matrix $\boldsymbol{X}$ is computed as $\boldsymbol{X} = \boldsymbol{V}^T\boldsymbol{R}\boldsymbol{U} \in \mathbb{R}^{q\times d}$. The feature matrices are then used to represent facial images for classification.

The one advantage of (2D)^{2}PCA over 2DPCA is that the former performs faster during the testing phase, since it requires fewer coefficients to represent an image. But as regards recognition accuracy, (2D)^{2}PCA only achieves a slight improvement over 2DPCA.

Similarly, in [14], 2DPCA is extended to bilateral-projection-based 2DPCA (B2DPCA), in which the projection matrices are learned simultaneously. However, compared to (2D)^{2}PCA, B2DPCA takes more training time because the projection matrices are computed through iterative processes. Interestingly, both (2D)^{2}PCA and B2DPCA deliver very similar recognition performances when their counterpart projection matrices are the same size.

C. Diagonal PCA (DiaPCA)

In [16], diagonal PCA (DiaPCA) is developed to enhance the recognition accuracy of 2DPCA. DiaPCA aims at uncovering more local structural information by learning a projective matrix from the diagonalized version of the original facial images. Conceptually, DiaPCA is quite similar to 2DPCA, with the exception that it replaces the image covariance matrix with a diagonal covariance matrix. In this sense, after computing the diagonal covariance matrix ($\boldsymbol{G}_{DiaPCA}$), the remainder of the procedure is identical to the 2DPCA procedure. As such, $\boldsymbol{G}_{DiaPCA}$ is obtained by means of (1) with the diagonalized version of $\boldsymbol{F}$, and the generation of the feature matrices $\boldsymbol{Y}$ and $\boldsymbol{X}$ follows that in 2DPCA.

The advantage of DiaPCA is that it takes into consideration the global structure of the training dataset as well as certain salient facial structures to model a more expressive feature space than those modeled by 2DPCA and (2D)^{2}PCA. On the other hand, although 2DPCA, (2D)^{2}PCA, and DiaPCA exhibiting relative differences in the required time for training or testing phases, their computational performances are generally comparable with respect to the overall time expense.

Moreover, while the aforementioned methods are guaranteed to produce intrinsic representation of the underlying face dataset, they, however, only address the global structure of the facial samples. Indeed, this is a consequence of the fact that the projection matrices of these methods are solved by maximizing the total scatter of the transformed samples, leading to disregard both the inherent characteristics of each individual class and class separability information, which are essential in satisfying the requirements for modelling discriminant feature spaces.

3. CLASS-WISE 2DPCA (CW2DPCA)

In contrast to 2DPCA, which finds a projection matrix that maximizes the total scatter in the modeled feature space, CW2DPCA constructs class-wise projection matrices, each one containing the projection axes that maximize the scatter of projected sample images of a particular class and the principal components that maximize between-class separability. This ensures that for a given training dataset, CW2DPCA is more likely to model distinct feature subspaces, which in turn facilitating the subsequent classification task. That is, within each subspace, the similarity measurement is performed to evaluate the identity of a test sample, and the classification decision is taken from the subspace that exhibits the highest similarity measure.

The proposed CW2DPCA-based FR method is implemented through two major stages. The first is the feature extraction stage, at which the class-specific feature subspaces are derived from the training dataset. The second stage assigns a test image to one of the training classes.

A. Feature Extraction

In view of the fact that sample images of the same class are naturally highly correlated, the application of 2DPCA to each class separately allows to select the most discriminative features within individual classes. In line with this, the suggested method initially applies 2DPCA to each class present in the training dataset. To do so, suppose that there are p classes $\boldsymbol{F}^k$, $k = 1, \cdots, p$, in the $\boldsymbol{F}$ training dataset, where each class contains v images, and the kth class is expressed as $\boldsymbol{F}^k = \{\boldsymbol{F}_i^k\}_{i=1}^{v}$. Based on the 2DPCA concept, CW2DPCA defines the image scatter matrix ($\boldsymbol{G}_{CW2DPCA}^k$) for the kth class as:

$$\boldsymbol{G}_{CW2DPCA}^k = \frac{1}{v}\sum_{i=1}^{v}(\boldsymbol{F}_i^k - \overline{\boldsymbol{F}}^k)^T\left(\boldsymbol{F}_i^k - \overline{\boldsymbol{F}}^k\right), \#(5)$$

where $\bar{\boldsymbol{F}}^k$ refers to the mean image of the kth class and $\boldsymbol{G}_{CW2DPCA}^k \in \mathbb{R}^{n\times n}$. Also, the projection matrix for the kth class, $\boldsymbol{U}^k = [\boldsymbol{u}_1^k, \cdots, \boldsymbol{u}_d^k] \in \mathbb{R}^{n\times d}$, is obtained by applying eigenvalue decomposition to the $\boldsymbol{G}_{CW2DPCA}^k$ and taking the first d dominant eigenvectors.

Although each subject possesses specific invariant characteristics, human faces do share common configural properties across various subjects. Without taking this observation into account, the class-wise projection matrices may retain similar global geometrical information, in particular those conveyed by the class means, thereby limiting the degree of class-separation. A possible way to account for this and at the same time to ensure wider separation between the different classes is to augment $\boldsymbol{U}^k$ with the orthonormal eigenvectors that maximize the between-class scatter ($\boldsymbol{S}$) [27], which is given, in the case of image-as-matrix, by the following expression [28]:

$$\boldsymbol{S} = \sum_{k=1}^{p} v\,(\bar{\boldsymbol{F}}^k - \bar{\boldsymbol{F}})^T(\bar{\boldsymbol{F}}^k - \bar{\boldsymbol{F}}), \#(6)$$

where $\boldsymbol{S} \in \mathbb{R}^{n\times n}$. Accordingly, let the columns of $\boldsymbol{U}^s = [\boldsymbol{u}_1^s, \cdots, \boldsymbol{u}_r^s]$, $\boldsymbol{U}^s \in \mathbb{R}^{n\times r}$, be the first r orthonormal eigenvectors of $\boldsymbol{S}$ corresponding to the leading eigenvalues. By concatenating $\boldsymbol{U}^k$ and $\boldsymbol{U}^s$ into a single matrix, followed by orthonormalizing the columns of the concatenated matrix using the well-known Gram–Schmidt (QR) orthonormalization procedure, the final projection matrix for the kth class is obtained and denoted by $\boldsymbol{W}^k = [\boldsymbol{U}^k\boldsymbol{U}^s] \in \mathbb{R}^{n\times(d+r)}$. Recall that the significance of $\boldsymbol{W}^k$ is that its first d columns preserve the invariant characteristics of the kth class, whereas the remaining r columns reflect the class separability. With this matrix, the feature matrices of training images in the kth class can be derived as follows:

$$\boldsymbol{Y}^k = \{\boldsymbol{Y}_i^k\}_{i=1}^{v}, \qquad \boldsymbol{Y}_i^k = \boldsymbol{F}_i^k\boldsymbol{W}^k \in \mathbb{R}^{m\times(d+r)}. \#(7)$$

From the above formulation, it is obvious that CW2DPCA learns p projection matrices $\boldsymbol{W}^1, \cdots, \boldsymbol{W}^p$ to respectively transform the images from p classes $\boldsymbol{F}^1, \cdots, \boldsymbol{F}^p$ onto p distinct subspaces $\boldsymbol{Y}^1, \cdots, \boldsymbol{Y}^p$, such that each subspace encodes both the most expressive information of its corresponding class and the class-separation information. This is in contrast to other related works, which endeavor to derive a common feature space, where neither class-specific discriminant information nor class separability are considered. Therefore, it is natural to expect that feature extraction through CW2DPCA would induce a more discriminative classification model.

B. Classification Based on Class-wise Representation

As mentioned earlier, the key property of CW2DPCA is that its ability to project the original training dataset into class-specific subspaces, each uniquely representing its respective class. Thus, each facial class is represented by a number of feature matrices corresponding to its training samples. Based on this representation, the classification of a given test sample is conducted by first projecting it onto each subspace, then employing an NN classifier to assign the class identity of the nearest subspace to the test sample.

More specifically, assume that $\boldsymbol{R} \in \mathbb{R}^{m\times n}$ is a test image. The evidence for $\boldsymbol{R}$ belonging to either of the p classes is the minimum distance, within each subspace, between the feature matrix of $\boldsymbol{R}$ and the training feature matrices. Without loss of generality, let $\boldsymbol{X}^k$ be the feature matrix of $\boldsymbol{R}$ generated with the subspace basis of the kth class as $\boldsymbol{X}^k = \boldsymbol{R}\boldsymbol{W}^k \in \mathbb{R}^{m\times(d+r)}$. The minimum distance between $\boldsymbol{X}^k$ and $\boldsymbol{Y}^k$ can be easily determined as follows:

$$\mathcal{D}^k = \min_i(\|\boldsymbol{X}^k - \boldsymbol{Y}_i^k\|_F), \qquad i = 1, \cdots, v\#(8)$$

where $\|\cdot\|_F$ stands for Frobenius norm. By returning $\mathcal{D} = \min(\mathcal{D}^1, \cdots, \mathcal{D}^p)$, the class membership of $\boldsymbol{R}$ is simply the class identity of the returned minimum distance to $\mathcal{D}$.

4. EXPERIMENTAL RESULTS

This section demonstrates the effectiveness of the introduced CW2DPCA in FR and compares it with the 2DPCA [9], $(2D)^2$PCA [15], and DiaPCA [16] on two publicly available face datasets: the ORL (AT&T Laboratories Cambridge) [29] and the Yale [30]. In all the experiments, the classification is carried out with NN technique based on the Frobenius norm. The performance of each method is evaluated regarding its recognition ability.

A. Experiments on the ORL Dataset

The ORL [29] face dataset comprises 400 images from 40 subjects. For each subject, there are 10 sample images with various face details, expressions, scales, and poses. All the images are gray-level with size of 112×92 pixels. Fig. 1 displays the sample images of one subject from the ORL dataset.

For evaluation purposes, four experimental scenarios are created through the partitioning of the ORL dataset into non-overlapping training and testing sets, such that the first f ($f = 2$, 3, 4, and 5) samples of each subject are assigned as training set and the remainder as testing set. The training samples are used to compute the projection matrices of the 2DPCA, $(2D)^2$PCA, DiaPCA, and CW2DPCA methods.

Figure 1. Sample images of one subject in the ORL dataset.

TABLE I. COMPARISON OF CW2DPCA WITH 2DPCA, (2D)^{2}PCA, AND DIAPCA ON THE ORL DATASET IN TERMS OF ARR (%) AND HRR (%).

Method	$f = 2$		$f = 3$		$f = 4$		$f = 5$	
	ARR	HRR	ARR	HRR	ARR	HRR	ARR	HRR
2DPCA	81.68	86.87	84.84	88.57	88.38	91.66	89.51	93.00
(2D)^{2}PCA	81.71	86.68	84.90	88.07	88.43	91.58	89.76	93.00
DiaPCA	83.17	85.87	86.26	89.57	89.83	92.08	91.18	93.50
CW2DPCA	85.19	87.61	88.29	91.12	91.89	94.00	93.20	95.50

Since, in general, the recognition rates vary with respect to the dimensionality of the produced feature matrices, the number of selected eigenvectors d in each projection matrix is varied from 1 to 30 in increments of one. Note that for (2D)^{2}PCA, the number of eigenvectors q in the left projection matrix $\boldsymbol{V}$ is set to 27, as it yielded the best recognition rates on this dataset [15]. For the same reason, the number of eigenvectors r used to form $\boldsymbol{U}^s$ in the class-wise projection matrices of CW2DPCA is fixed at 15.

Fig. 2 exhibits the recognition performances of the 2DPCA, (2D)^{2}PCA, DiaPCA, and CW2DPCA on the ORL dataset. As can be seen, in all the stated scenarios, CW2DPCA comprehensively outperformed the other three methods and provided a more consistent recognition performance, despite the increased dimensionality of class-wise projection matrices.

Table 1 presents a detailed comparison of the four methods in terms of the average recognition rate (ARR) and the highest recognition rate (HRR). In this comparison, CW2DPCA improved the ARR by about 3.5% compared to the 2DPCA and (2D)^{2}PCA. It also resulted in an up to 2% better ARR than DiaPCA. Furthermore, as shown in Table 1, CW2DPCA persistently attained the best HRR in all the experimental scenarios. For instance, in the case of $f = 5$, the achieved HRR by CW2DPCA is 95.5%, outperforming 2DPCA/(2D)^{2}PCA and DiaPCA by a margin of 2.5% and 2%, respectively.

B. Experiments on the Yale Dataset

The Yale dataset [30] contains 15 distinct subjects, with 11 frontal gray-level images per subject, characterizing various variations, including facial details, expressions, and lighting conditions. In the experiments on this dataset, the original images are cropped to 100 × 80 pixels, according to the face's location. Fig. 3 shows cropped images of a typical subject in the Yale dataset.

In a same manner as with the ORL dataset, four experimental scenarios are adopted, in which the first f ($f = 2$, 3, 4, and 5) samples from each subject are designated as training set, while the rest are employed for testing. In all the experiments, the number of eigenvectors d is varied in the range of 1 to 30, with an interval of one. Also, the numbers of eigenvectors q and r, used for building $\boldsymbol{V}$ and $\boldsymbol{U}^s$, are set to 20 and 8, respectively.

Fig. 4 displays the recognition rates of the 2DPCA, (2D)^{2}PCA, DiaPCA, and CW2DPCA methods from each experimental scenario. Again, it can clearly be seen that the recognition performance achieved with CW2DPCA is better than those achieved with the other three methods in this challenging dataset. Along with this, CW2DPCA exhibited the smallest performance degradation among all the competing methods.

The ARRs and HRRs obtained by the four methods are reported in Table 2. In an overall sense, CW2DPCA increased the ARR by 3.3-4.25% compared to both 2DPCA and (2D)^{2}PCA. It also increased the ARR by 2-2.75% compared to DiaPCA. Additionally, CW2DPCA consistently outperformed 2DPCA, (2D)^{2}PCA, and DiaPCA in terms of HRR. For example, in the case of

$f = 2$, CW2DPCA reached an HRR of 68.19%, while 2DPCA/(2D)^{2}PCA and DiaPCA reached HRRs of 65.92% and 66.22%, respectively.

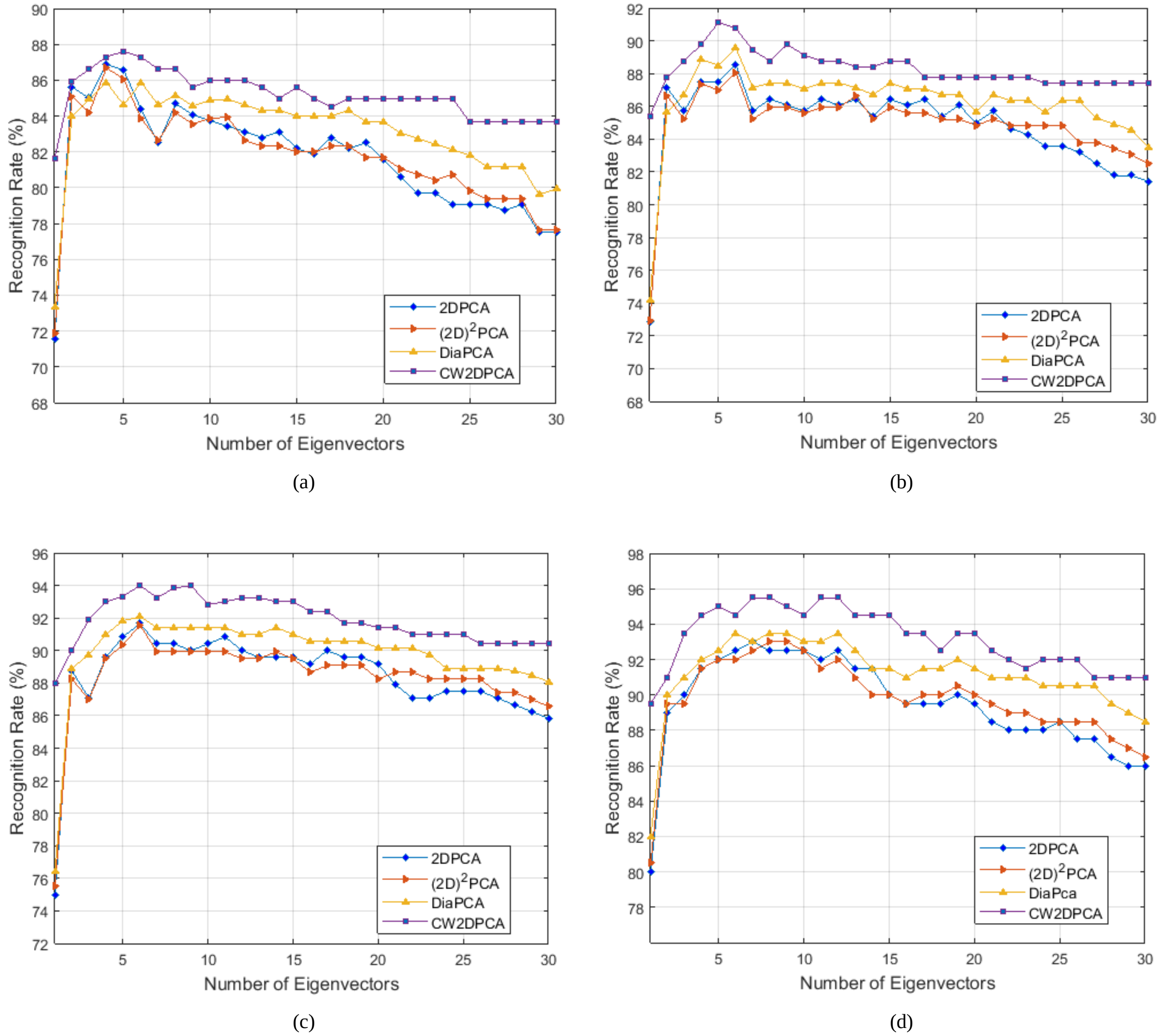

Figure 2. Recognition rates of the 2DPCA, (2D)^{2}PCA, DiaPCA, and CW2DPCA on the ORL dataset: (a) $f = 2$, (b) $f = 3$, (c) $f = 4$, and (d) $f = 5$.

Figure 3. Cropped images of one subject from the Yale dataset.

TABLE II. COMPARISON OF CW2DPCA WITH 2DPCA, $(2D)^2$PCA, AND DIAPCA ON THE YALE DATASET IN TERMS OF ARR (%) AND HRR (%).

Method	$f = 2$		$f = 3$		$f = 4$		$f = 5$	
	ARR	HRR	ARR	HRR	ARR	HRR	ARR	HRR
2DPCA	60.96	65.92	78.00	81.33	84.15	88.57	86.85	91.11
$(2D)^2$PCA	60.99	65.92	78.13	82.50	84.30	87.62	86.92	90.00
DiaPCA	62.29	66.22	79.50	83.33	85.54	89.74	88.40	92.11
CW2DPCA	64.77	68.19	82.24	84.83	87.58	90.32	90.89	93.52

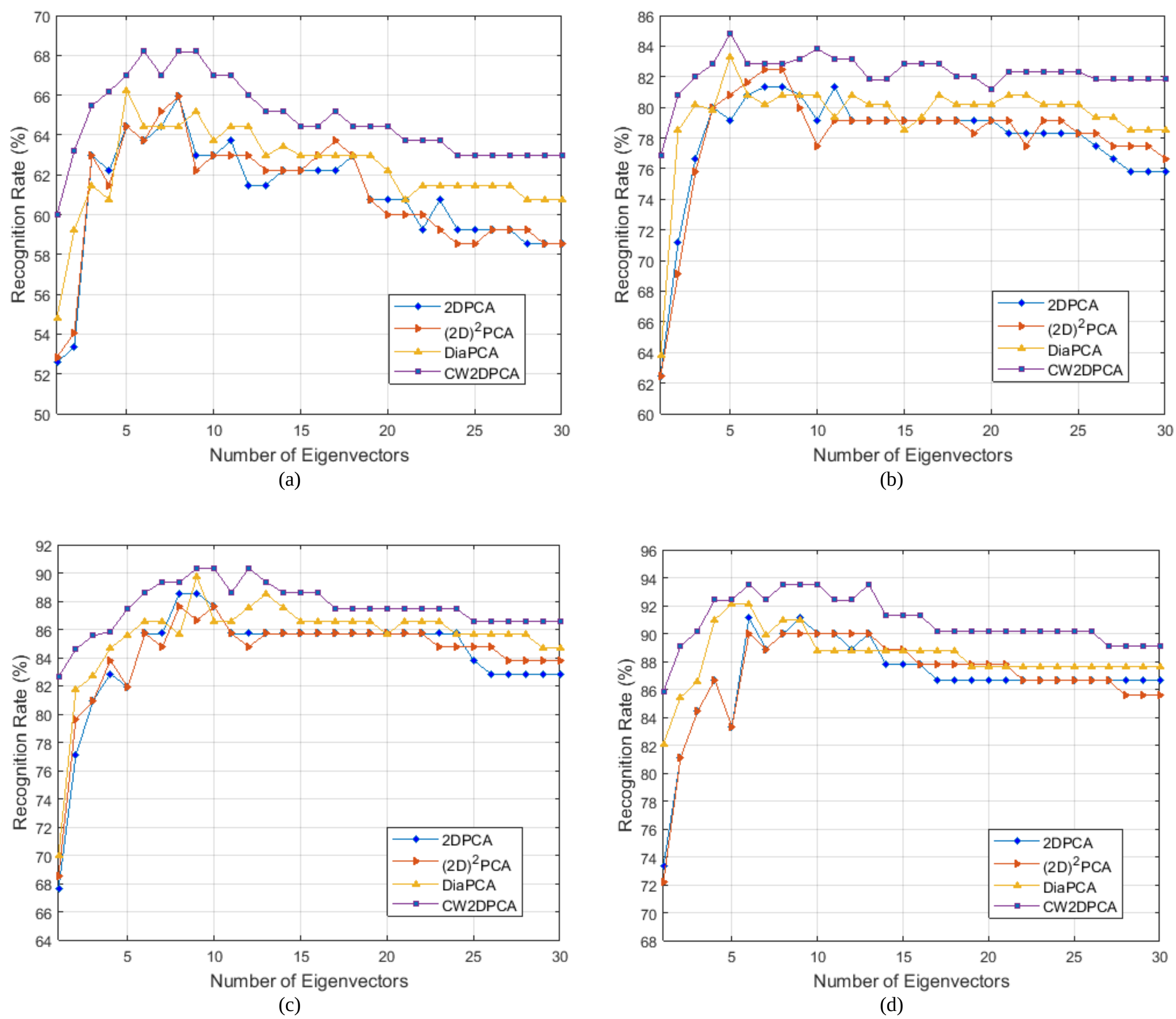

Figure 4. Recognition rates of the 2DPCA, $(2D)^2$PCA, DiaPCA, and CW2DPCA on the Yale dataset: (a) $f = 2$, (b) $f = 3$, (c) $f = 4$, and (d) $f = 5$.

C. *Computational Performance*

The CW2DPCA and the competitive methods are implemented in MATLAB running on an Intel Core i3 at 2.26 GHz CPU speed with 4GB RAM. It may be interesting to mention that, due to the formulation of 2DPCA-based FR methods, including the proposed CW2DPCA, is made up based on image-as-matrix calculation, their overall computational performances are expected to be relatively short and comparable. Nevertheless, as the CW2DPCA method requires explicit computation of p sets of training feature matrices in the training phase and p test feature matrices in the testing phase, it is not surprising that the total computation time (TCT) of this method is slightly longer among the compared methods. Table 3 presents the TCT of each method when $d = 1$ in the preceding experiments on the ORL and Yale datasets. It is apparent in this table that the TCT of CW2DPCA still fairly comparable to those of the 2DPCA, $(2D)^2$PCA, and DiaPCA.

TABLE III. THE REQUIRED TCT (S) FOR THE 2DPCA, $(2D)^2$PCA, DIAPCA, AND CW2DPCA.

Method	**ORL**				**Yale**			
	$f = 2$	$f = 3$	$f = 4$	$f = 5$	$f = 2$	$f = 3$	$f = 4$	$f = 5$
2DPCA	0.175	0.187	0.204	0.225	0.482	0.538	0.570	0.598
$(2D)^2$PCA	0.173	0.181	0.197	0.218	0.470	0.522	0.553	0.587
DiaPCA	0.209	0.238	0.250	0.277	0.529	0.613	0.669	0.727
CW2DPCA	0.293	0.321	0.361	0.387	0.629	0.709	0.795	0.838

D. Discussion

According to the results of the conducted experiments on the ORL and Yale datasets, the CW2DPCA framework presented in this paper continuously produced promising recognition performance, in terms of ARR and HRR, across all the experimental scenarios. These results suggest that by capturing sufficient class-wise features and class-separation information, it is possible to model expressive and separable face-specific subspaces, which are much useful for classifying human faces.

As expected, the experimental results revealed that CW2DPCA method always delivered better and more consistent recognition performances over the compared methods, namely 2DPCA, $(2D)^2$PCA, and DiaPCA. However, the one drawback of this method is that it requires slightly more total computational time, which, fortunately, can be remarkably shortened by the pre-computation of the class-specific projection matrices. Further, the results also showed that the four methods share similar behavior regarding the ARR and TCT, that is, both of them tended to increase as the size of training sets increased.

5. CONCLUSIONS

This paper introduced class-wise 2DPCA (CW2DPCA) as a variant of the classical 2DPCA technique. Unlike other 2DPCA-based FR methods that only consider the holistic information of a given training dataset, CW2DPCA emphasized the importance of class-specific discriminant information to contribute to improving FR accuracy. Furthermore, it leveraged the strength of 2DPCA to prune redundant information within individual classes and the advantage of inter-class relationships to account for class-separation. As a result, the CW2DPCA method modeled discriminative subspaces and maximized the separation between the classes in the training dataset.

The recognition performance of CW2DPCA was evaluated using the ORL and Yale face datasets. The reported experimental results verified the effectiveness of this method for FR. In addition, CW2DPCA outperformed 2DPCA, $(2D)^2$PCA, and DiaPCA methods, and yielded a more consistent recognition performance for the benchmark datasets. Overall, in terms of average recognition rate, the ordering was CW2DPCA > DiaPCA > $(2D)^2$PCA ≥ 2DPCA. Finally, since CW2DPCA is a general framework, it can be utilized to solve other computer vision problems, such as multiple objects classification.

REFERENCES

[1] C. L. Lisetti and D. J. Schiano, "Automatic Facial Expression Interpretation : Where Human-Computer Interaction , Artificial Intelligence and Cognitive Science Intersect," Pragmat. Cogn., vol. 8, no. 1, pp. 185–235, 2000.

[2] P. Tome, R. Vera-rodriguez, J. Fierrez, and J. Ortega-garcia, "Facial soft biometric features for forensic face recognition," Forensic Sci. Int., vol. 257, pp. 271–284, 2015.

[3] X. Yuan, Z. Sun, Y. Varol, and G. Bebis, "A Distributed Visual Surveillance System," in Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, 2003, pp. 199–204.

[4] W. Zhao and A. Rosenfeld, "Face Recognition : A Literature Survey," ACM Comput. Surv., vol. 35, no. 4, pp. 399–458, 2003.

[5] A. S. Tolba, "Face Recognition : A Literature Review," Int. J. Comput. Inf. Eng., vol. 2, no. 7, pp. 2556–2571, 2008.

[6] M. Turk and A. Pentland, "Eigenfaces for Recognition," J. Cogn. Neurosci., vol. 3, no. 1, pp. 71–86, Jan. 1991.

[7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 711–720, Jul. 1997.

[8] A. Rao and S. Noushath, "Subspace methods for face recognition," Comput. Sci. Rev., vol. 4, no. 1, pp. 1–17, 2010.

[9] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 1, pp. 131–137, 2004.

[10] W. Yang, C. Sun, L. Zhang, and K. Ricanek, "Laplacian bidirectional PCA for face recognition," Neurocomputing, vol. 74, no. 1–3, pp. 487–493, 2010.

[11] C. Wang, B. Yin, X. Bai, and Y. Sun, "Color face recognition based on 2DPCA," 2008 19th Int. Conf. Pattern Recognit., pp. 1–4, 2008.

[12] L. Oliveira, M. Mansano, A. Koerich, and A. de S. Britto, "2D Principal Component Analysis for Face and Facial-Expression Recognition," Comput. Sci. Eng., vol. 13, no. 3, pp. 9–13, 2011.

[13] D. I. Islam, S. R. N. Anal, and A. Datta, "Facial Expression Recognition Using 2DPCA on Segmented Images," Adv. Comput. Commun. Paradig., pp. 289–297, 2018.

[14] H. Kong, L. Wang, E. K. Teoh, X. Li, J.-G. Wang, and R. Venkateswarlu, "Generalized 2D principal component analysis for face image representation and recognition," Neural Networks, vol. 18, no. 5, pp. 585–594, 2005.

[15] D. Zhang and Z.-H. Zhou, "(2D)2PCA: Two-directional two-dimensional PCA for efficient face representation and recognition," Neurocomputing, vol. 69, no. 1, pp. 224–231, 2005.

[16] D. Zhang, Z.-H. Zhou, and S. Chen, "Diagonal principal component analysis for face recognition," Pattern Recognit., vol. 39, no. 1, pp. 140–142, 2006.

[17] W. Yang, C. Sun, and K. Ricanek, "Sequential Row--Column 2DPCA for face recognition," Neural Comput. Appl., vol. 21, no. 7, pp. 1729–1735, Oct. 2012.

[18] Q. Wang and Q. Gao, "Robust 2DPCA and Its Application," in 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016, pp. 1152–1158.

[19] Q. Gao, L. Ma, Y. Liu, X. Gao, and F. Nie, "Angle 2DPCA : A New Formulation for 2DPCA," IEEE Trans. Cybern., vol. 48, no. 5, pp. 1672–1678, 2018.

[20] X. Li, Y. Pang, and Y. Yuan, "L1-Norm-Based 2DPCA," IEEE Trans. Syst. Man, Cybern. Part B, vol. 40, no. 4, pp. 1170–1175, 2009.

[21] A. Mashhoori and M. Z. Jahromi, "Block-wise two-directional 2DPCA with ensemble learning for face recognition," Neurocomputing, vol. 108, pp. 111–117, 2013.

[22] J. Wang, "Generalized 2-D Principal Component Analysis by Lp-Norm for Image Analysis," IEEE Trans. Cybern., vol. 46, no. 3, pp. 792–803, 2016.

[23] Y. Wang and Q. Li, "Robust 2DPCA with F -Norm Minimization," IEEE Access, vol. 7, pp. 68083–68090, 2019.

[24] S. Shan, W. Gao, and D. Zhao, "Face recognition based on face-specific subspace," Int. J. Imaging Syst. Technol., vol. 13, no. 1, pp. 23–32, 2003.

[25] C. Guzel Turhan and H. S. Bilge, "Class-wise two-dimensional PCA method for face recognition," IET Comput. Vis., vol. 11, no. 4, pp. 286–300, 2016.

[26] E. K. Tang, P. N. Suganthan, X. Yao, and A. K. Qin, "Linear dimensionality reduction using relevance weighted LDA," Pattern Recognit., vol. 38, no. 4, pp. 485–493, 2005.

[27] K. Das and Z. Nenadic, "An efficient discriminant-based solution for small sample size problem," Pattern Recognit., vol. 42, pp. 857--866, 2009.

[28] M. Li and B. Yuan, "2D-LDA : A statistical linear discriminant analysis for image matrix," Pattern Recognit. Lett., vol. 26, pp. 527–532, 2005.

[29] "The ORL Database of Faces," AT&T Laboratories Cambridge. [Online]. Available: https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[30] "Yale Face Database." [Online]. Available: http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html.

Falah Alsaqre received the BS degree in control and systems engineering from the University of Technology, Baghdad, Iraq, in 1988, the MS degree in control and instrumentation engineering from the same university, in 1995, and the PhD (Eng.) degree in signals and information processing from the Beijing Jiaotong University, Beijing, China, in 2004. Following the appointment as a postdoctoral fellow at Beijing Jiaotong University, he joined the University Malaysia Perils as a senior lecturer. He also appointed as an assistant professor at the Gulf University, Bahrain, and Sohar University, Oman. Most recently, he joined Al-Hikma Univerity College, Bahgdad, Iraq. His research interests include information processing, computer vision, pattern recognition, and machine learning. He is a member of the IEEE.