Response to the comments on "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition"

Jian Yang [a]   David Zhang [b] *   Jing-yu Yang [c]

[a] Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA
[b] Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong
[c] Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, P. R. China

E-mail: csjyang@comp.polyu.edu.hk (Jian Yang)
csdzhang@comp.polyu.edu.hk (David Zhang)
yangjy@mail.njust.edu.cn (Jing-yu Yang)

First of all, we would like to thank Xiaoqing Ding for her intrests in our paper [2] and providing her comments [1] on it. We also thank the Associate Editor of PAMI give us a chance to respond to Ding's comments and to clarify our idea.

X. Ding [1]'s comments can be sumarized as follows:

Comment 1: 2DPCA is actually a column vector based sub-image PCA, which is an approximation of image PCA that only considers between-column correlations.

Comment 2: The comparison between image PCA and the so-called 2DPCA were carried out under wrong conditions which lead to invalid results. Under the correct conditions, image PCA in fact substantially outperforms the so-called 2DPCA in most of the experiments.

Comment 3: The only advantage Yang's 2D-PCA has is that it requires fewer total training images. This advantage diminishes once the number of classes is close to n×m (where n and m are the image dimensions). In other words, if the experiment were carried out in a dataset containing larger number of smaller images, the results would be much different.

Comment 4: (a) For 2DPCA, there is no theoretical backing to its proclaimed superior performance over image PCA. (b) The choice of using column vector based sub-image PCA over using row vector based sub-image PCA in their scheme appears entirely arbitrary and is not discussed at all in Yang's paper.

The following are the detailed responses to these comments.

Response to Comment 1

On Comment 1, we understand the author's intention and partially agree with it. It is true that 2DPCA only considers and eliminates correlations between column vectors (by eliminating correlations between any two elements of a row vector). But, we don't think the author's proposition "2DPCA is actually a column vector based sub-image PCA" is proper. This is because Formula (5) in the comment [1] is not a standard *column vector based covariance matrix*. If we rewrite an $m \times n$ image $\mathbf{A}$ as a vector of a set of column vectors $\mathbf{A} = (\mathbf{A}^{[1]}, \mathbf{A}^{[2]}, \cdots, \mathbf{A}^{[n]})$, the *column vector based covariance matrix* should be

$$\Sigma = E\{(\sum_{j=1}^{n} \mathbf{A}^{[j]} (\mathbf{A}^{[j]})^T\} = \sum_{j=1}^{n} E\{\mathbf{A}^{[j]} (\mathbf{A}^{[j]})^T\}, \tag{1}$$

which is an $m \times m$ matrix, while the size of image covariance matrix of 2DPCA is $n \times n$. A proper proposition should be "2DPCA is equivalent to an ***image row vector*** (rather than column vector) based PCA" [3, 4, 13].

In our original idea of 2DPCA [2], we view an image as a whole matrix. We construct image covariance matrix directly based on given image matrices and then project image matrix onto a set of principal eigenvectors. It is interesting that 2DPCA, in implementation, is indeed equivalent to an *image row vector* based PCA. This is fully due to the operation rule of matrix multiplication: the product of two matrices is determined by the inner product of row vectors and column vectors. Specifically, given M training image samples $\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_M$, each of which is an $m \times n$ matrix and can be rewritten as

$$\mathbf{A}_j = \begin{pmatrix} \mathbf{A}_j^{(1)} \\ \mathbf{A}_j^{(2)} \\ \vdots \\ \mathbf{A}_j^{(m)} \end{pmatrix}, \text{ where } \mathbf{A}_j^{(1)}, \mathbf{A}_j^{(2)}, \cdots, \mathbf{A}_j^{(m)} \text{ are row vectors of } \mathbf{A}_j. \tag{2}$$

Denoting the mean image as $\overline{\mathbf{A}} = \frac{1}{M}\sum_{j=1}^{M}\mathbf{A}_j = \begin{pmatrix} \overline{\mathbf{A}}^{(1)} \\ \overline{\mathbf{A}}^{(2)} \\ \vdots \\ \overline{\mathbf{A}}^{(m)} \end{pmatrix}$, the image covariance matrix of 2DPCA is

$$\mathbf{G}_t = \frac{1}{M}\sum_{j=1}^{M}(\mathbf{A}_j - \overline{\mathbf{A}})^{\mathrm{T}}(\mathbf{A}_j - \overline{\mathbf{A}}) = \frac{1}{M}\sum_{j=1}^{M}\begin{pmatrix} \mathbf{A}_j^{(1)} - \overline{\mathbf{A}}^{(1)} \\ \mathbf{A}_j^{(2)} - \overline{\mathbf{A}}^{(2)} \\ \vdots \\ \mathbf{A}_j^{(m)} - \overline{\mathbf{A}}^{(m)} \end{pmatrix}^T \begin{pmatrix} \mathbf{A}_j^{(1)} - \overline{\mathbf{A}}^{(1)} \\ \mathbf{A}_j^{(2)} - \overline{\mathbf{A}}^{(2)} \\ \vdots \\ \mathbf{A}_j^{(m)} - \overline{\mathbf{A}}^{(m)} \end{pmatrix}$$

$$= \frac{1}{M}\sum_{j=1}^{M}\sum_{i=1}^{m}(\mathbf{A}_j^{(i)} - \overline{\mathbf{A}}^{(i)})^T(\mathbf{A}_j^{(i)} - \overline{\mathbf{A}}^{(i)}) \tag{3}$$

Thus, $\mathbf{G}_t$ can be viewed as a ***special***[1] covariance matrix constructed by all Mm row vectors $\mathbf{A}_j^{(i)}$ ($i = 1,2,\cdots,m$, $j = 1,2,\cdots,M$).

In addition, after the projection of an image $\mathbf{A}$ onto $\mathbf{G}_t$'s principal eigenvectors $\mathbf{X}_1,\cdots,\mathbf{X}_d$, we obtain its *principal component vectors*

$$\mathbf{Y}_k = \mathbf{A}\mathbf{X}_k = \begin{pmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \\ \vdots \\ \mathbf{A}^{(m)} \end{pmatrix}\mathbf{X}_k = \begin{pmatrix} \mathbf{A}^{(1)}\mathbf{X}_k \\ \mathbf{A}^{(2)}\mathbf{X}_k \\ \vdots \\ \mathbf{A}^{(m)}\mathbf{X}_k \end{pmatrix}, \tag{4}$$

which indicates that the projection of an image matrix is essentially the projection of its row vectors.

---

[1] "Special" means Formula (3) is not a standard covariance matrix constructed by all Mm row vectors $\mathbf{A}_j^{(i)}$. A standard covariance matrix constructed by all row vectors should be

$$\Sigma = \frac{1}{Mm}\sum_{j=1}^{M}\sum_{i=1}^{m}(\mathbf{A}_j^{(i)} - \overline{\mathbf{A}}^{(0)})^T(\mathbf{A}_j^{(i)} - \overline{\mathbf{A}}^{(0)}),$$

where $\overline{\mathbf{A}}^{(0)}$ is the mean of all row vectors, that is, $\overline{\mathbf{A}}^{(0)} = \frac{1}{Mm}\sum_{j=1}^{M}\sum_{i=1}^{m}\mathbf{A}_j^{(i)}$.

In Formula (3), $\overline{\mathbf{A}}^{(i)} = \frac{1}{M}\sum_{j=1}^{M}\mathbf{A}_j^{(i)}$. That means the *i*-th row vector of an image is centered by the mean of the *i*-th row vectors of all images, rather than by the total mean vector $\overline{\mathbf{A}}^{(0)}$.

Thereby, we can draw a conclusion that 2DPCA is equivalent to an "image row vector based PCA". Wang LW et al [3, 4] and H. Kong et al [13] first found this equivalence and presented it in their papers. So, we don't think Comment 1 is new and meaningful since this idea has been published and becomes well-known.

It should be mentioned that to the best of our knowledge, no one presented and applied "image row vector based PCA" to face recognition before the publication of our 2DPCA paper. The finding of "2DPCA is equivalent to an image row vector based PCA" provides an alternative way to understand 2DPCA, but, we don't think it can obliterate the contribution of 2DPCA to face recognition.

In addition, some improved 2DPCA methods [11, 12, 13, 14] have been developed recently. These methods consider *between-column* correlations and *between-row* correlations at the same time. The drawback of 2DPCA "only considering correlations between column vectors" has been overcome.

Response to Comment 2

Comment 2 is related to our experiments. We are afraid we cannot agree with it.

The author claimed that in our paper [2], the comparison between PCA and 2DPCA were carried out under "wrong conditions" just because 2DPCA generally uses more features than PCA. The "correct conditions" that the author claimed is "2DPCA must use the same number of features when comparing with PCA". We don't think this opinion is right and reasonable.

As we know, every model (algorithm) has its parameters. The number of features (dimension) chosen is a parameter of PCA or 2DPCA. When we try to compare the performance of two models, Model_1 (t_1) and Model_1 (t_2), where t_1 and t_2 are parameters. A reasonable way is to choose a t_1^* making the performance of Model_1 best and to choose a t_2^* making the performance of Model_2 best. Generally, $t_1^* \neq t_2^*$, and, we cannot require the two model must use the same parameter. For instance, when PCA and LDA (Fisher linear discriminant analysis) are applied to face recognition, PCA generally use more features than LDA. It is unreasonable to say "LDA can only use c-1 features (c is number of classes), so PCA cannot use more features when comparing with LDA".

In our paper [2], we compare PCA and 2DPCA as they both achieve their best performance when the parameter, the number of features, is well chosen. 2DPCA needs more features to achieve its best recognition rate while PCA needs less. ***The disadvantage of 2DPCA for image representation has been pointed out in our paper [2] (We did not intend to hide this weakness; actually, we mentioned this weakness twice. The first time is in Section 4.1*** *(see Page 133, the last paragraph in the right column)* ***and the second time is in the last paragraph of Section 5 "Conclusion and future work")***. However, 2DPCA, when using more features, can achieve better (recognition or reconstruction) performance than PCA, while PCA cannot even if it uses more features. In addition, we also suggested a way on how to avoid this disadvantage of 2DPCA in the paper, i.e., 2DPCA plus PCA (that is, PCA is followed to further reduce the dimension of 2DPCA).

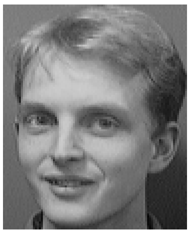
In this response, we will specify our idea from the following two aspects:

(i) On image construction (compression)

Here, we have to clarify that we never claim that "2DPCA is better than PCA for image compression" in our paper. Since 2DPCA needs more features for image representation, its compression rate is much lower than PCA.

In our paper, we just try to show that in small sample size case (as we know, face recognition is typically a small sample size problem), the quality of 2DPCA-based reconstruction image can be better than PCA-based one, especially for the image out of training sample set. The author claimed the comparison of PCA and 2DPCA based reconstruction in Fig. 4 of our paper [2] is invalid. He thinks 2DPCA should use the same number of features to compare with PCA.

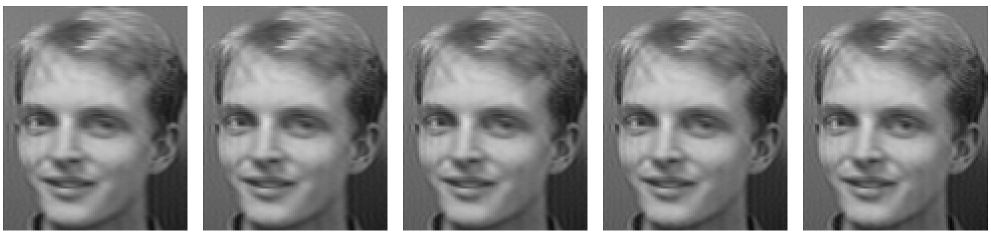
We know that the number of PCA components (features) depends on the number of training samples. In Fig. 4 case, since there are 200 images training, the total number of features should be 199. In contrast, the number of 2DPCA features has no this constraint. In Fig. 4, we try to show 2DPCA, using more features, can get better reconstruction of an image than PCA. Due to page limit of a short correspondence paper, we only listed a small number of reconstruction images (for 2DPCA, d =2, 4, 6, 8, 10, and for PCA, d=5, 10, 20, 30, 40). This may cause some misunderstanding, for example, some one may think if PCA use all 199 components (features), it can get better reconstruction than 2DPCA. Here, we list the remaining reconstruction images (for 2DPCA, d =12, 14, ---, 40, and for PCA, d=60, 70, ---, 190, 199) in Figure 1. Figure 1 shows PCA cannot achieve better reconstruction than 2DPCA, even if all 199 features are used.



Original image

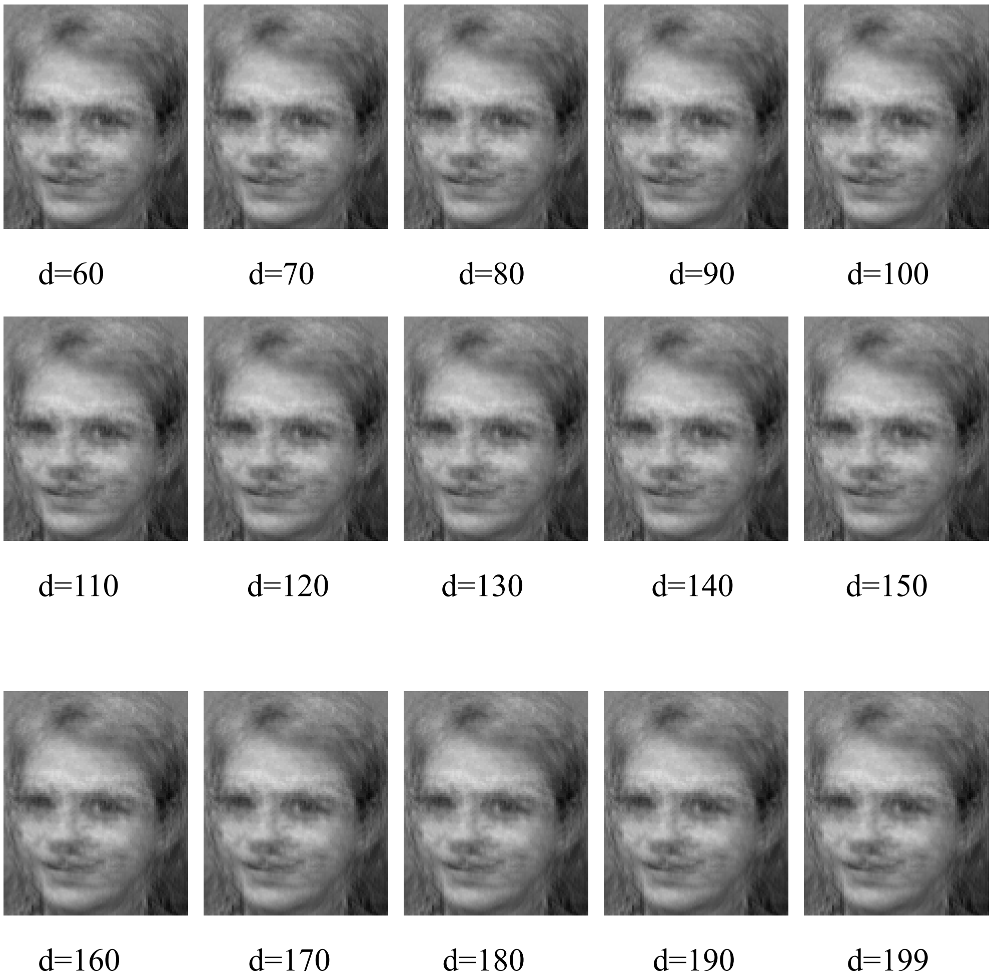
d=12 d=14 d=16 d=18 d=20

d=22 d=24 d=26 d=28 d=30



d=32 d=34 d=36 d=38 d=40

(a) 2DPCA-based reconstruction



(b) PCA-based reconstruction

Figure 1 The remaining reconstruction images of Fig. 4 in our paper [2]

In Ding's comment [1], 360 images from all 400 images of ORL database are used for training. In this case, it is true PCA based reconstruction results are much better than before. But, we are still sure 2DPCA, using more features (for example, 40x112), can achieve better reconstruction performance than PCA using all 359 features.

Finally, it should be mentioned that 2DPCA has been improved and its weakness of for image compression (representation) has been overcome. H. Kong [13] proposed a generalized 2DPCA (G2DPCA) method and J. Ye [14] presented a method called Generalized Low Rank Approximations (GLRA). Both methods are similar in their basic idea. G2DPCA (or GLRA) can achieve better reconstruction performance than PCA, even if they use the same number of features.

(ii) On recognition performance

The author tried to show us that PCA can outperform 2DPCA in the so-called "correct conditions" (where both methods use the same number of features) in Tables 1 and 2. From Table 1, we can see that the recognition rate of PCA does not improve with the increase of features. However, the recognition rate of 2DPCA improves rapidly with the increase of features. It can be predicted that 2DPCA could achieve better (at least comparable) performance if more features (for example, 4x112) were used. Table 2 seems to tell us the same thing.

It should be mentioned that a number of researchers [3-13] have evaluated our 2DPCA and demonstrated its performance advantage over PCA in their experiments. From Ding's opinion, their experiments were all performed under wrong conditions and their results are all invalid?

In addition, some improved 2DPCA versions [11, 12, 13, 14] have been proposed to overcome the weakness of 2DPCA-based image representation. These improved 2DPCA versions can perform better than (or comparable with) PCA using the same number of features [11-14].

Response to Comment 3

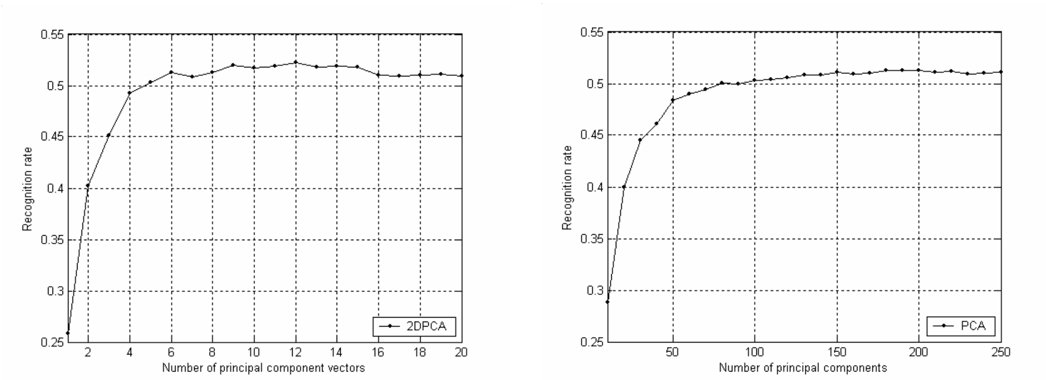
It seems that the advantage of 2DPCA over PCA in small sample size cases has been accepted by the author. When the training sample size becomes large, how about the performance of 2DPCA? Is Ding's opinion right? (Ding [1], however, didn't give any justification to support her opinion.)

To address these questions, we performed an experiment on FERET database [15]. The basic gallery of the FERET 1996 standard subset contains 1,196 face images of 1,196 subjects (one image per subject). There are four sets of probe images used for testing, where the *fafb* probe set contains 1,195 images taken at the same time as the gallery images but with different facial expressions. The *fafc* probe set contains 194 images taken under different lighting conditions. The *duplicate I* probe set contains

722 images taken anywhere between one minute and 1,031 days after their respective gallery matches. Finally, the *duplicate II* probe set is a subset of the *duplicate I* set, containing 234 images taken at least 18 months after their gallery entries. In our experiment, the face portion of each original image is automatically cropped based on the location of eyes and mouth (i.e., the known coordinates of two eyes and the mouth in the images) and resized to an image of $\mathbf{30 \times 30}$ pixels.

In our experiment, we use 1196 gallery images to form the training set, and unit four probe sets as a testing set. 2DPCA and PCA are used for feature extraction and a nearest neighbor classifier with Euclidean distance is used for classification. The experimental results are shown in Figure 2 and Table 1. It can be seen that 2DPCA still outperforms PCA.

This experiment shows us, although the training sample size (or number of classes, 1196) is larger than the size of image (30x30=900), the performance advantage of 2DPCA still exists. It seems that the performance comparison between PCA and 2DPCA is independent of the image size and training sample size. So, we cannot agree with Ding's opinion.



2DPCA PCA

Figure 2 Performance comparisons of 2DPCA and PCA on FERET database

Table 1 Maximal recognition rates of 2DPCA and PCA and the corresponding dimensions

Method	Maximal recognition rate	Number of features
PCA	51.30 %	180
2DPCA	52.27 %	12x30=360

Response to Comment 4

The involved problems of this comment are very interesting, *but they are beyond the scope of our correspondence paper. After all, no paper can contain everything. Every paper leaves some questions open.*

(a) In our paper [2], we just demonstrated 2DPCA is more effective than PCA by experiments. With respect to the problem "Why does 2DPCA outperform PCA in face recognition", we only gave our explanation in conclusion part. On this problem, H. Kong tried to give a more reasonable explanation [13].

(b) Our paper just presented the basic idea and formulation of 2DPCA model. It is impossible for us to address the problem that X. Ding [1] mentioned in our paper since at that time, the equivalence relationship between 2DPCA and *image row vector* based PCA has not been found yet.

As discussed before, we know that 2DPCA is equivalent to an *image row vector* based PCA. So, the original 2DPCA can be called row-based 2DPCA, which can be denoted by 2DPCA (row). If we use the transpose of image **A**, that is, $\mathbf{A}^T$, instead of **A**, as an input to 2DPCA model, we will get an alternative 2DPCA version, which is column-based 2DPCA and denoted by 2DPCA (column). For these two 2DPCA versions, which one is better for face recognition? It is a question difficult to answer. We have done some research on this last year. Based on our experimental results on a modified FERET 2000 subset which contains 200 subjects, as shown in Table 2, we can get some interesting results.

Table 2 The maximal recognition rates (%) of PCA and two versions of 2DPCA on four Probe sets of the modified FERET 2000 subset

Method	**P1** (pose ±15º) (400 images)	**P2** (Pose ±25º) (400 images)	**P3** (Expression) (200 images)	**P4** (Illumination) (200 images)
PCA	56.8	28.2	75.5	42.5
2DPCA (row)	*73.0*	*56.8*	82.0	56.5
2DPCA (column)	55.2	27.8	*85.5*	*58.0*

From Table 2, we can see that On Probe set 1 (pose 15º) and Probe set 2 (pose 25º), the row-based 2DPCA performs much better than the column-based one. On Probe set 3 (expression) and Probe set 4 (illumination), the column-based 2DPCA performs better than the row-based one. It seems that row-based 2DPCA is more insensitive to the pose variations, while the column-based one is more robust under the variations of facial expression and lighting conditions.

Our Recent Work

Considering two 2DPCA versions, row-based and column-based, might be complementary in discrimination, we proposed a combined 2DPCA+PCA framework recently, as illustrated in Figure 2. We would like to test this framework and compare it with the common image PCA using the largest face image database available in the world－FRGC version 2.0 database [16].

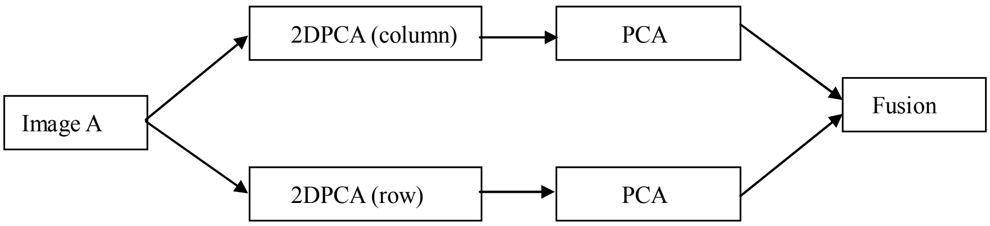


Figure 3. Illustration of the proposed combined 2DPCA+PCA framework

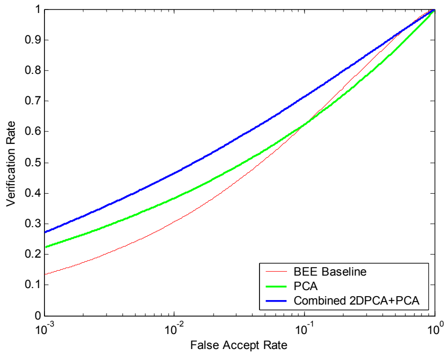
FRGC version 2.0 consists of six experiments. The experiments measure performance on still images taken with controlled lighting and background, uncontrolled lighting and background, 3D imagery, multi-still imagery, and between 3D and still images. Experiment 4 is designed to measure progress on recognition from uncontrolled frontal still images. In Experiment 4, the training set contains 12,776 still images, which are either controlled or uncontrolled. The target set consists of 16,028 controlled still images, and the query set consists of 8014 uncontrolled still images. The FRGC baseline algorithm, a PCA with Mahalinobis cosine distance measure, reveals that Experiment 4 is the most challenging FRGC experiment. We therefore choose FRGC Experiment 4 to evaluate our algorithm.

In our experiment, the face region of each image is first cropped from the original high-resolution still images and resized to 32x32. Figure 4 show some example images used in our experiment. Some previous research shows that the R component image carries more discriminating information than the grayscale images converted from the RGB image [18]. We therefore use only the R component images (by discarding the G and B component images) in our experiment.

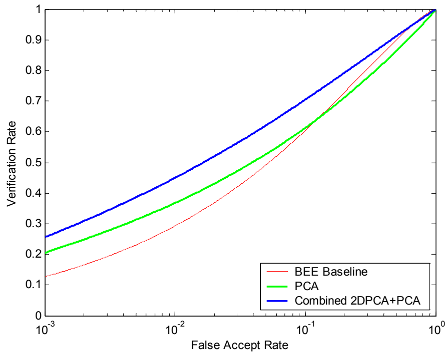
Figure 4 Example cropped images in FRGC version 2

According to the experimental protocol of FRGC, the face recognition performance is reported using the Receiver Operating Characteristic (ROC) curves, which plot the Face Verification Rate (FVR) versus the False Accept Rate (FAR). The ROC curves are automatically generated by the available Biometric Experimentation Environment (BEE) when a similarity matrix (distance matrix) is input to the system. The generated three ROC curves, ROC I, ROC II, and ROC III, correspond to the evaluation results based on images collected within semesters, within a year and between semesters, respectively [17]. The similarity matrix stores the similarity score of every query image versus target image pair. So, the size of the similarity matrix is T x Q, where T is the number of target images and Q is the number of query images.

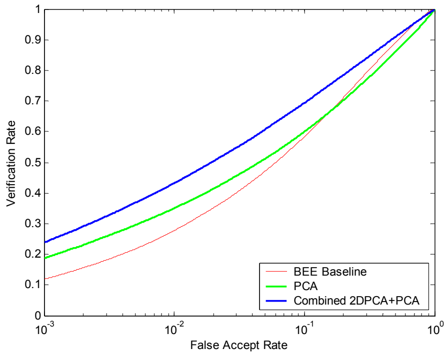
PCA and the proposed Combined 2DPCA+PCA are, respectively, employed for image representation. For PCA, 800 principal components (PCs) are finally extracted to represent a face (we choose the number of PCs as 800 because in this case PCA achieves its best performance). For Combined 2DPCA+PCA, we choose d=14 in column-based and row-based 2DPCA transform, and the 440 principal components in the following PCA. Mahalinobis cosine distance is used to calculate the similarity matrix. The generated ROC curves by BEE are shown in Figure 5. The verification rates (%) corresponding to three methods when the False Accept Rate is 0.1% are listed in Table 3. These results show that Combined 2DPCA can significantly improve the performance of PCA on FRGC version 2.



ROC I



ROC II



ROC III

Figure 5 Three ROC curves corresponding to PCA, Combined 2DPCA+PCA, and BEE Baseline

Table 3 Verification rate (%) comparison when the False Accept Rate is 0.1%

Method	ROC I	ROC II	ROC III
BEE-Baseline	13.36	12.67	11.86
PCA	22.11	20.45	18.70
Combined 2DPCA+PCA	**27.04**	**25.51**	**23.77**

Appendix: A small discussion

On Page 7, in the first two lines, "where as the maximum number of features that can be used in 2DPCA is 39x112". We don't think this presentation is right. Since the image covariance matrix $\mathbf{G}_t$ is generally nonsingular even if there are only 40 images for training, the maximum number of features extracted by 2DPCA should be 92 x112, not 39x112.

References

[1] Xiaoqing Ding, Comments on “Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition”, submitted to PAMI.

[2] Jian Yang, David Zhang, A. F. Frangi, Jing-yu Yang, “Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition”, IEEE Trans. Pattern Analysis and Machine Intelligence, 2004, Vol.26, No.1, pp. 131-137.

[3] Wang LW, Wang X, Zhang XR, et al, The equivalence of two-dimensional PCA to line-based PCA, PATTERN RECOGNITION LETTERS, 2005, 26 (1): 57-60

[4] Wang LW, Wang X, Feng JF, On image matrix based feature extraction algorithms, IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS PART B-CYBERNETICS, 2006, 36 (1): 194-197

[5] Visani M, Garcia C, Laurent C, Comparing robustness of two-dimensional PCA and eigenfaces for face recognition, LECTURE NOTES IN COMPUTER SCIENCE, 3212: 717-724 Part 2 2004

[6] Chen SC, Zhu YL, Zhang DQ, et al., Feature extraction approaches based on matrix pattern: MatPCA and MatFLDA, PATTERN RECOGNITION LETTERS, 26 (8): 1157-1167 JUN 2005

[7] Rodriguez-Aragon LJ, Conde C, Serrano A, et al., Face verification advances using spatial dimension reduction methods: 2DPCA & SVM, LECTURE NOTES IN COMPUTER SCIENCE, 3617: 978-985 2005

[8] Rodriguez-Aragon LJ, Serrano A, Conde C, et al., Spatial approach to pose variations in face verification, LECTURE NOTES IN COMPUTER SCIENCE, 3646: 351-361 2005

[9] Zuo WM, Zhang D, Wang KQ, An assembled matrix distance metric for 2DPCA-based image recognition, PATTERN RECOGNITION LETTERS, 27 (3): 210-216 FEB 2006

[10] Nagabhushan P, Guru DS, Shekar BH, Visual learning and recognition of 3D objects using two-dimensional principal component analysis: A robust and an efficient approach, PATTERN RECOGNITION, 39 (4): 721-725 APR 2006

[11] Zhang DQ, Chen SC, Liu J, Representing image matrices: Eigenimages versus eigenvectors, LECTURE NOTES IN COMPUTER SCIENCE, 3497: 659-664 2005

[12] Zhang DQ, Zhou ZH, (2D)(2)PCA: Two-directional two-dimensional PCA for efficient face representation and recognition, NEUROCOMPUTING 69 (1-3): 224-231 DEC 2005

[13] H. Kong, L Wang, EK Teoh, X Li, JG Wang et al, Generalized 2D principal component analysis for face image representation and recognition, Neural Networks, 2005, Volume 18, Issue 5-6, Pages: 585 - 594

[14] Jieping Ye, Generalized Low Rank Approximations of Matrices, Machine Learning, 2005, Volume 61, Numbers 1-3, Pages: 167 – 191.

[15] P.J. Phillips, H. Wechsler, J. Huang, and P. Rauss, “The FERET Database and Evaluation Procedure for Face-Recognition Algorithms,” Image and Vision Computing, vol. 16, pp. 295-306,

[16] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, “Overview of the Face Recognition Grand Challenge,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2005.

[17] P.J. Phillips, FRGC Third Workshop Presentation, FRGC Workshop, February 2005.
[18] P. Shih and C. Liu: "Comparative Assessment of Content-Based Face Image Retrieval in Different Color Spaces", International Journal of Pattern Recognition and Artificial Intelligence, vol. 19, no. 7, pp. 873-893, 2005.