

A Unified View of Two-Dimensional Principal Component Analyses

Kohei Inoue*, Kenji Hara, and Kiichi Urahama

Department of Communication Design Science, Kyushu University,
4-9-1, Shiobaru, Minami-ku, Fukuoka, 815-8540 Japan
{k-inoue,hara,urahama}@design.kyushu-u.ac.jp

Abstract. Recently, two-dimensional principal component analysis (2D-PCA) and its variants have been proposed by several researchers. In this paper, we summarize their 2DPCA variants, show some equivalence among them, and present a unified view in which the non-iterative 2DPCA variants are interpreted as the non-iterative approximate algorithms for the iterative 2DPCA variants, i.e., the non-iterative 2DPCA variants are derived as the first iterations of the iterative algorithm started from different initial settings. Then we classify the non-iterative 2DPCA variants on the basis of their algorithmic patterns and propose a new non-iterative 2DPCA algorithm based on the classification. The effectiveness of the proposed algorithm is experimentally demonstrated on three publicly accessible face image databases.

Keywords: Dimensionality reduction, Principal component analysis, Two-dimensional principal component analysis.

1 Introduction

Principal component analysis (PCA) and linear discriminant analysis (LDA) are well-known techniques for dimensionality reduction. Since they are based on vectors, matrices such as 2D face images must be transformed into 1D image vectors in advance. However, the resultant vectors usually lead to a high-dimensional vector space, where it is difficult to solve the (generalized) eigenvalue problems for PCA and LDA.

Recently, Yang and Yang [1] and Yang et al. [2] have proposed two-dimensional PCA (2DPCA) which can handle matrices directly without vectorizing them. However, 2DPCA is approximately equivalent to the conventional PCA operated only on the row vectors of matrices [3–5], and needs many more coefficients for image representation than PCA. To overcome this problem, several variants of 2DPCA have been proposed recently. Since they have been proposed almost independently and simultaneously, the relationship among them is not clear. Therefore, the systematization of them is desired for a deeper understanding of 2DPCA variants.

* This work was partially supported by the Japan Society for the Promotion of Science under the Grant-in-Aid for Scientific Research (23700212).

In this paper, we summarize the variants of 2DPCA [2] and show some equivalence among them. Moreover, we present a unified view of 2DPCA variants, where the non-iterative 2DPCA variants are interpreted as the non-iterative approximate algorithms for the iterative ones, i.e., we show that the non-iterative ones are derived as the first iterations of the iterative algorithm started from different initial settings. Then we classify the non-iterative ones on the basis of their algorithmic patterns and present a new non-iterative 2DPCA algorithm based on the classification result.

The rest of this paper is organized as follows: Section 2 briefly surveys the related work. Section 3 summarizes the original 2DPCA and its variants and shows some equivalence among them. Section 4 presents a unified view of 2DPCA variants and classifies the non-iterative 2DPCA variants on the basis of their algorithmic patterns. From the classification result, a new non-iterative algorithm is derived. Section 5 shows experimental results which demonstrate the effectiveness of the derived algorithm compared with the conventional non-iterative 2DPCA variants. Section 6 concludes this paper.

2 Related Work

Yang and Yang [1] and Yang et al. [2] proposed two-dimensional principal component analysis (2DPCA) which is based on 2D image matrices rather than 1D vectors so the image matrix does not need to be transformed into a vector prior to feature extraction. Ye et al. [6] proposed generalized PCA (GPCA) which is formulated as an optimization problem and derived an iterative procedure for GPCA. Kong et al. [7] proposed a framework of generalized 2DPCA to extend the original 2DPCA in two perspectives: a bilateral-projection-based 2DPCA (B2DPCA) and a kernel-based 2DPCA. Zhang and Zhou [8] proposed two-directional 2DPCA, i.e., $(2D)^2PCA$ which combines 2DPCA and alternative 2DPCA. Zhang et al. [9] proposed a method for representing 2D image matrices using eigenimages, which are 2D matrices with the same size as the original images and can be directly computed from the original image matrices. Benito and Peña [10] proposed a method for dimensionality reduction based on the projection of images as matrices. Xu et al. [11] proposed complete 2DPCA (C2DPCA) in which two image covariance matrices are constructed and their eigenvectors are derived for image feature extraction. Xu et al. [12] proposed a two-stage strategy, parallel image matrix compression (PIMC), to compress the image matrix redundancy among both row vectors and column ones. Zuo et al. [13] proposed bi-directional PCA (BDPCA) and an assembled matrix distance metric to calculate the distance between two feature matrices. Wen and Shi [14] proposed image PCA (IPCA) in which a family of projective feature vectors, which is called the projective feature image, is obtained by 2DPCA and then the transpose of the projective feature image is processed by 2DPCA again. Lu et al. [15] proposed doubleside 2DPCA (D2DPCA) and the constructive method for incrementally adding observation to the existing eigen-space model, called incremental D2DPCA. Xi and Ramadge [16] proposed separable PCA (SPCA)

and unified 2DPCA [2], BDPCA [13] and generalized low rank approximations of matrices (GLRAM) [17]. Yang et al. [18] proposed Bi-2DPCA which performs 2DPCA [2] twice: the first one is in horizontal direction and the second is in vertical direction.

The above 2DPCA variants proposed almost independently and simultaneously. Therefore, the theoretical relationship among them is not clear. In the following, we will discuss the relationship theoretically.

3 2DPCA and Its Variants

In this section, we summarize the original 2DPCA [2] and its variants which are roughly divided into two categories: iterative and non-iterative algorithms, and present some equivalence among them.

3.1 2DPCA

Suppose that there are M training image samples, the k th training image is denoted by an $m \times n$ matrix $A_k \in \mathfrak{R}^{m \times n}$ where \mathfrak{R} denotes the set of real numbers, and the average image of all training samples is denoted by $\bar{A} = \frac{1}{M} \sum_{k=1}^M A_k$. Then the image covariance (scatter) matrix [2] is defined by

$$G = \frac{1}{M} \sum_{k=1}^M (A_k - \bar{A})^T (A_k - \bar{A}) \in \mathfrak{R}^{n \times n}, \quad (1)$$

where T denotes the transpose of a matrix, and the generalized total scatter criterion [2] is expressed by

$$J(X) = \text{tr}(X^T G X), \quad (2)$$

where tr denotes the matrix trace and $X \in \mathfrak{R}^{n \times \tilde{n}}$ for $\tilde{n} < n$ is subject to $X^T X = I_{\tilde{n}}$ where $I_{\tilde{n}}$ is the $\tilde{n} \times \tilde{n}$ identity matrix. The optimal X that maximize $J(X)$ is obtained by $X = [x_1, \dots, x_{\tilde{n}}]$ where x_j ($j = 1, \dots, \tilde{n}$) denotes the eigenvector of G corresponding to the j th largest eigenvalue. Finally, each A_k is transformed into

$$B_k = A_k X \in \mathfrak{R}^{m \times \tilde{n}}. \quad (3)$$

Although the above 2DPCA can reduce the number of columns from n to \tilde{n} in (3), the number of rows, m , is unchanged. Therefore, 2DPCA needs many more coefficients for image representation than PCA [2, 9]. To overcome this problem, several variants of 2DPCA have been proposed recently. They can be classified into two categories: iterative and non-iterative algorithms.

3.2 Non-iterative 2DPCA Variants

In this subsection, we summarize the non-iterative 2DPCA variants, which are further divided into two sub-categories: parallel and serial methods.

Parallel Method. Instead of the image covariance (scatter) matrix in (1), another one can be defined as follows:

$$\tilde{G} = \frac{1}{M} \sum_{k=1}^M (A_k - \bar{A}) (A_k - \bar{A})^T \in \mathfrak{R}^{m \times m}, \quad (4)$$

from which another criterion is obtained by

$$\tilde{J}(Y) = \text{tr} \left(Y^T \tilde{G} Y \right), \quad (5)$$

where $Y \in \mathfrak{R}^{m \times \tilde{m}}$ for $\tilde{m} < m$ is subject to $Y^T Y = I_{\tilde{m}}$. The optimal Y that maximize $\tilde{J}(Y)$ is obtained by $Y = [y_1, \dots, y_{\tilde{m}}]$ where y_i ($i = 1, \dots, \tilde{m}$) denotes the eigenvector of \tilde{G} corresponding to the i th largest eigenvalue. Zhang and Zhou [8] called this method the alternative 2DPCA. Finally, each A_k is transformed into

$$\tilde{B}_k = Y^T A_k X \in \mathfrak{R}^{\tilde{m} \times \tilde{n}}. \quad (6)$$

We call this type of 2DPCA variant the parallel method because X and Y are calculated in a parallel manner. Essentially, (2D)²PCA [8], eigenimages [9], C2DPCA [11], BDPCA [13], and D2DPCA [15] are equivalent to this parallel method.

Serial Method. As opposed to the above parallel method, we can consider the serial method as follows. First, 2DPCA [2] described in Subsection 3.1 is conducted to obtain $\{B_k\}_{k=1}^M$. Next, the image covariance (scatter) matrix is constructed for the set of the transposed matrices $\{B_k^T\}_{k=1}^M$ as follows:

$$\hat{G} = \frac{1}{M} \sum_{k=1}^M (B_k^T - \bar{B}^T)^T (B_k^T - \bar{B}^T) \in \mathfrak{R}^{m \times m}, \quad (7)$$

where $\bar{B} = \frac{1}{M} \sum_{k=1}^M B_k$. Then the total scatter criterion

$$\hat{J}(\hat{Y}) = \text{tr} \left(\hat{Y}^T \hat{G} \hat{Y} \right) \quad (8)$$

for $\hat{Y} \in \mathfrak{R}^{m \times \tilde{m}}$ which is subject to $\hat{Y}^T \hat{Y} = I_{\tilde{m}}$ is maximized by $\hat{Y} = [\hat{y}_1, \dots, \hat{y}_{\tilde{m}}]$ where \hat{y}_i ($i = 1, \dots, \tilde{m}$) denotes the eigenvector of \hat{G} corresponding to the i th largest eigenvalue. Each A_k is transformed into

$$\hat{B}_k = \hat{Y}^T B_k = \hat{Y}^T A_k X \in \mathfrak{R}^{\tilde{m} \times \tilde{n}}. \quad (9)$$

Essentially, PIMC [12], IPCA [14], and Bi-2DPCA [18] are equivalent to this serial method.

3.3 Iterative 2DPCA Variants

In the above non-iterative 2DPCA variants, two matrices X and Y (or \hat{Y}) are derived from different criteria. On the other hand, the iterative 2DPCA variants are formulated as an optimization of a single criterion as follows [19]:

$$\max_{X,Y} \frac{1}{M} \sum_{k=1}^M \left\| Y^T \tilde{A}_k X \right\|_F^2 \quad (10)$$

$$\text{subj.to } X^T X = I_{\tilde{n}}, Y^T Y = I_{\tilde{m}}, \quad (11)$$

where $\tilde{A}_k = A_k - \bar{A}$. Let $F(X, Y)$ be the objective function in (10). Then the Lagrange function for (10)-(11) is given by

$$\mathcal{L} = F(X, Y) - \text{tr} [\Lambda_X (X^T X - I_{\tilde{n}})] - \text{tr} [\Lambda_Y (Y^T Y - I_{\tilde{m}})], \quad (12)$$

where Λ_X and Λ_Y are symmetric matrices of which the elements are the Lagrange multipliers. From $\partial \mathcal{L} / \partial X = 0$ and $\partial \mathcal{L} / \partial Y = 0$, we have

$$G_Y X = X \Lambda_X, \quad (13)$$

$$G_X Y = Y \Lambda_Y, \quad (14)$$

respectively, where $G_Y = \frac{1}{M} \sum_{k=1}^M \tilde{A}_k^T Y Y^T \tilde{A}_k$ and $G_X = \frac{1}{M} \sum_{k=1}^M \tilde{A}_k X X^T \tilde{A}_k^T$. Hence, for a fixed Y , the optimal X is obtained by $X = [x_1, \dots, x_{\tilde{n}}]$ where x_j ($j = 1, \dots, \tilde{n}$) denotes the eigenvector of G_Y corresponding to the j th largest eigenvalue, and similarly, for the obtained X , the optimal Y is obtained by $Y = [y_1, \dots, y_{\tilde{m}}]$ where y_i ($i = 1, \dots, \tilde{m}$) denotes the eigenvector of G_X corresponding to the i th largest eigenvalue. This procedure is repeated until the convergence. Essentially, GPCA [6], B2DPCA [7], Benito's method [10], and SPCA [16] are equivalent to this method.

4 A Unified View of 2DPCA Variants

In this section, we present a unified view of the 2DPCA variants described in the above section. That is, we show that the non-iterative 2DPCA variants including the original 2DPCA [2] can be interpreted as the non-iterative approximate algorithms for the iterative algorithm in Subsection 3.3.

First, the original 2DPCA [2] is derived from the iterative algorithm in Subsection 3.3 as follows: if we initialize $Y = I_m$, then $F(X, Y)$ becomes

$$F(X, Y) = \text{tr} \left[X^T \left(\frac{1}{M} \sum_{k=1}^M \tilde{A}_k^T Y Y^T \tilde{A}_k \right) X \right] = \text{tr} (X^T G_X X) = J(X). \quad (15)$$

Therefore, X obtained by the first iteration coincides with that of 2DPCA [2].

On the other hand, if we initialize $X = I_n$, then we have $F(X, Y) = \tilde{J}(Y)$. Therefore, Y obtained by the first iteration in this setting coincides with that

of the parallel method in Subsection 3.2. That is, the parallel method uses X and Y obtained by the first iterations of the iterative algorithm started from different initial settings: $Y = I_m$ and $X = I_n$, respectively.

The serial method in Subsection 3.2 can be derived by initializing $Y = I_m$. Since X obtained by the serial method coincides with that of 2DPCA [2], it can be obtained by the first iteration. Next, the obtained X is used for computing Y , i.e., we have

$$F(X, Y) = \text{tr} \left[Y^T \left(\frac{1}{M} \sum_{k=1}^M \tilde{A}_k X X^T \tilde{A}_k^T \right) Y \right] = \text{tr} (Y^T \hat{G} Y). \quad (16)$$

Therefore, Y obtained by the first iteration coincides with that of the serial method; $Y = \hat{Y}$.

Thus, the conventional non-iterative 2DPCA variants can be derived as the first iterations of the iterative algorithm started from different initial settings. Furthermore, this viewpoint suggests the existence of the other non-iterative variant, i.e., we can consider another (alternative) serial method which is initialized as $X = I_n$. Then we obtain a pair of Y and \hat{X} which is the solution to $\max_X F(X, Y)$. Finally, we can combine the two serial methods to obtain the selective method as follows:

$$(X^*, Y^*) = \arg \max_{(X, Y) \in \{(X, \hat{Y}), (\hat{X}, Y), (\hat{X}, \hat{Y})\}} F(X, Y). \quad (17)$$

Each A_k is transformed into

$$B_k^* = (Y^*)^T A_k X^* \in \mathbb{R}^{\tilde{m} \times \tilde{n}}. \quad (18)$$

This method will achieve better performance than the conventional non-iterative 2DPCA variants because it is guaranteed that the objective function value obtained by the selective method is greater than or equal to that of the serial methods. The superiority of the proposed method to the other methods will be experimentally demonstrated in the next section. Table 1 shows the classification of the non-iterative 2DPCA variants. The proposed selective method in (17) fills up the blank in Table 1.

Table 1. Classification of non-iterative 2DPCA variants

	Renew X or Y	Renew X and Y
Initialize X or Y	(alternative) 2DPCA	(alternative) serial
Initialize X and Y	parallel	selective (proposed)

5 Experimental Results

In this section, we show experimental results on the ORL face image database [20], the Caltech Faces [21] and the UMIST face database [22]. The ORL database [20]

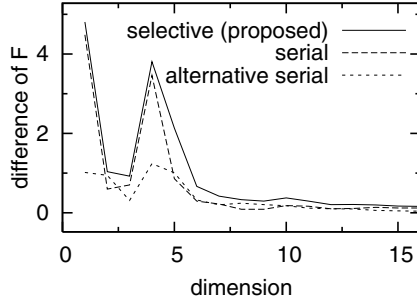


Fig. 1. Difference of F for the ORL face image database [20]

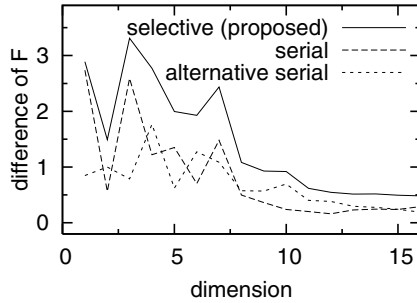


Fig. 2. Difference of F for the Caltech Faces [21]

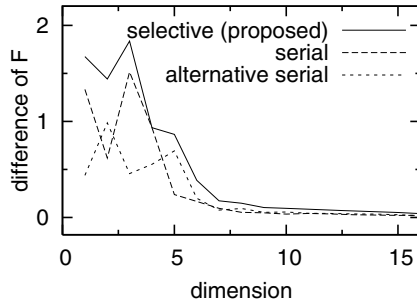


Fig. 3. Difference of F for the UMIST face database [22]

contains face images of 40 persons. For each person, there are 10 different face images. In our experiment, we used the first 5 images per person, i.e., $M = 5 \times 40 = 200$. The height and width of each image are $m = 112$ and $n = 92$, respectively. Fig. 1 shows the differences of the objective function values: $F(\xi, \eta) - F(X, Y)$, where $\xi = X$, $\eta = \hat{Y}$ for broken line (the serial method), $\xi = \hat{X}$, $\eta = Y$ for dotted line (the alternative serial method) and $\xi = \hat{X}$, $\eta = \hat{Y}$

for solid line (the proposed selective method), and $F(X, Y)$ denotes the objective function value for the parallel method, i.e., X and Y are obtained by maximizing (2) and (5), respectively. The horizontal axis denotes \tilde{m} ($= \tilde{n}$). Since all lines lie in the positive region, it is clear that the serial, alternative serial and selective methods achieve higher objective function values than the parallel method. Furthermore, among the three methods, the proposed selective method achieves the highest objective function value.

Figs. 2 and 3 show the results for the Caltech Faces [21] and the UMIST face database, respectively. In the Caltech Faces [21], we used 445 cropped face images. The height and width of each image are $m = 165$ and $n = 122$, respectively. In the UMIST face database, we used 380 face images. The height and width of each image are $m = 112$ and $n = 92$, respectively. Figs. 2 and 3 also demonstrate the superiority of the proposed selective method to the other methods.

6 Conclusion

In this paper, we summarized the 2DPCA variants which have been proposed by several researchers recently, and presented a unified view of the 2DPCA variants. We discussed some equivalence of the 2DPCA variants and classified them on the basis of their algorithmic patterns. Then we proposed a new non-iterative 2DPCA algorithm based on the classification result. The proposed method achieved higher objective function value than the other non-iterative 2DPCA variants. Future work will include the summarization of the variants of two-dimensional linear discriminant analysis.

References

1. Yang, J., Yang, J.-Y.: From Image Vector to Matrix: A Straightforward Image Projection Technique - IMPCA vs. PCA. *Pattern Recognition* 35(9), 1997–1999 (2002)
2. Yang, J., Zhang, D., Frangi, A.F., Yang, J.-Y.: Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Trans. PAMI* 26(1), 131–137 (2004)
3. Wang, L., Wang, X., Zhang, X., Feng, J.: The equivalence of two-dimensional PCA to line-based PCA. *Pattern Recogn. Lett.* 26(1), 57–60 (2005)
4. Zhang, D., Chen, S., Liu, J.: Representing Image Matrices: Eigenimages Versus Eigenvectors. In: Wang, J., Liao, X.-F., Yi, Z. (eds.) *ISNN 2005. LNCS*, vol. 3497, pp. 659–664. Springer, Heidelberg (2005)
5. Gao, Q.: Is two-dimensional PCA equivalent to a special case of modular PCA? *Pattern Recogn. Lett.* 28(10), 1250–1251 (2007)
6. Ye, J., Janardan, R., Li, Q.: GPCA: An Efficient Dimension Reduction Scheme for Image Compression and Retrieval. In: *KDD*, pp. 354–363. ACM, New York (2004)
7. Kong, H., Li, X., Wang, X., Teoh, E.K., Wang, J.-G., Venkateswarlu, R.: Generalized 2D Principal Component Analysis. In: *Proc. IJCNN*, pp. 108–113 (2005)
8. Zhang, D., Zhou, Z.-H. (2D)²PCA: Two-Directional Two-Dimensional PCA for Efficient Face Representation and Recognition. *Neurocomputing* 69(1-3), 224–231 (2005)

9. Zhang, D., Chen, S., Liu, J.: Representing Image Matrices: Eigenimages Versus Eigenvectors. In: Wang, J., Liao, X.-F., Yi, Z. (eds.) ISSN 2005. LNCS, vol. 3497, pp. 659–664. Springer, Heidelberg (2005)
10. Benito, M., Peña, D.: A Fast Approach for Dimensionality Reduction with Image Data. *Pattern Recognition* 38(12), 2400–2408 (2005)
11. Xu, A., Jin, X., Jiang, Y., Guo, P.: Complete Two-Dimensional PCA for Face Recognition. In: *Proc. ICPR*, vol. (3), pp. 481–484 (2006)
12. Xu, D., Yan, S., Zhang, L., Li, M., Ma, W., Liu, Z., Zhang, H.: Parallel Image Matrix Compression for Face Recognition. In: *MMM 2005*, pp. 232–238. IEEE Computer Society, Washington (2005)
13. Zuo, W., Zhang, D., Wang, K.: Bidirectional PCA with Assembled Matrix Distance Metric for Image Recognition. *IEEE trans. SMC-B* 36(4), 863–872 (2006)
14. Wen, Y., Shi, P.: Image PCA: A New Approach for Face Recognition. In: *ICASSP 2007*, vol. (1), pp. 1241–1244. IEEE (2007)
15. Lu, C., Liu, W., Liu, X., An, S.: Double Sides 2DPCA for Face Recognition. In: Huang, D.-S., Wunsch II, D.C., Levine, D.S., Jo, K.-H. (eds.) *ICIC 2008*. LNCS, vol. 5226, pp. 446–459. Springer, Heidelberg (2008)
16. Yongxin Taylor Xi, Y.T., Ramadge, P.J.: Separable PCA for Image Classification. In: *ICASSP*, pp. 1805–1808. IEEE (2009)
17. Ye, J.: Generalized Low Rank Approximations of Matrices. *Mach. Learn.* 61(1-3), 167–191 (2005)
18. Yang, J., Xu, Y., Yang, J.-Y.: Bi-2DPCA: A Fast Face Coding Method for Recognition. In: *Pattern Recognition Recent Advances*, pp. 313–340. InTech (2010)
19. Inoue, K., Urahama, K.: Equivalence of Non-Iterative Algorithms for Simultaneous Low Rank Approximations of Matrices. In: *Proc. CVPR*, pp. 154–159. IEEE (2006)
20. Samaria, F.S., Harter, A.C.: Parameterisation of a Stochastic Model for Human Face Identification. In: *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, pp. 138–142 (1994)
21. Caltech Faces, <http://www.vision.caltech.edu/html-files/archive.html>
22. Graham, D.B., Allinson, N.M.: Characterizing Virtual Eigen Signatures for General Purpose Face Recognition. *Face Recognition: From Theory to Applications* 163, 446–456 (1998)