THE 2D FACTOR ANALYSIS AND ITS APPLICATION TO FACE RECOGNITION WITH A SINGLE SAMPLE PER PERSON

Alexei M. C. Machado

Pontifical Catholic University of Minas Gerais - Department of Computer Science
R. Dom Jose Gaspar,500, Belo Horizonte MG, 30535901, Brazil

ABSTRACT

In this paper, a novel theoretical model of data reduction and multivariate analysis is proposed. The Two-dimensional Factor Analysis is an extension of classical factor analysis in which the images are treated as matrices instead of being converted to unidimensional vectors. By maximally representing the correlation among the pixels, it is able to capture meaningful information about the spatial relationships of the elements in a two-dimensional signal. The method is illustrated in the problem of face recognition with superior results when compared to other approaches based on principal component analysis. Experiments using public databases under different pose and illumination conditions show that the proposed method is significantly more effective than the two-dimensional principal component analysis while dealing with samples composed by a single image per person.

***Index Terms*—** Face recognition, factor analysis, principal component analysis, data reduction.

1. INTRODUCTION

Despite the increasing interest of both academia and industry, unconstrained face recognition remains an open problem, as the challenges imposed by this application are far from being trivial. Face recognition becomes a difficult task when factors such as illumination, occlusion and pose variation should be jointly accounted for. As well as other computer vision problems, face recognition suffers from the *curse of dimensionality* that occurs when the signals being analyzed are represented in high-dimensional spaces. A common approach for dealing with this problem is to use a method of data reduction in which the images are projected onto a new lower-dimensional space with possible loss of information.

One of the most relevant mathematical frameworks used for data reduction is the Principal Component Analysis (PCA). It is a simple and computationally efficient method that relies on the assumption of gaussianity and is suitable to the analysis of samples with a single image per person. Other approaches based on the Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA) are much more restrictive while dealing with small samples or Normal distributions. In spite of it, important methods based on these techniques can also be found in the literature with the purpose of data reduction, whether used separately or associated with PCA.

In 2004, Yang *et al.* [1] proposed the Two-dimensional Principal Component Analysis (2DPCA), an extension to PCA that does not require the image to be converted into a one-dimensional vector before eigendecomposition. In this sense, much of the topological relationships between the pixels of the image are preserved. Following this work, many variations of the technique have been proposed, with incremental improvements in its accuracy.

Another linear model suitable to data reduction that is similar to PCA is the Factor Analysis (FA). While the major objective of PCA is to represent data in a new basis whose axes correspond to the principal modes of the sample *variance*, FA aims at maximally representing the sample *correlation*. Since the pixels in images depicting faces are highly correlated to each other, the nature of FA can be an advantageous feature to data reduction performance. Nevertheless, the use of FA in face recognition methods has been restricted to much fewer works, perhaps due to its resemblance to PCA.

In this paper, we propose the Two-dimensional Factor Analysis (2DFA), an extension of FA that is applied to the image as a two-dimensional matrix. By maximally representing the correlation among the pixels, it is able to capture meaningful information about spatial relationships of the elements in a two-dimensional signal. The method is illustrated in the problem of face recognition with superior results when compared to 2DPCA.

1.1. BACKGROUND

Since the groundbreaking work of Kirby and Sirovich [2] and the publication of the EigenFaces method by Turk and Pentland [3], PCA has been the most widely used framework for data reduction in the problem of face recognition. In a variation of the method, the Projection-Combined Principal Component Analysis ($(\mathrm{PC})^2\mathrm{A}$) combines the original face image with its first-order projection with superior re-

This work is partially supported by Pontifcia Universidade Catlica de Minas Gerais - FIP-1S2015.

sults while considering a single sample per person [4]. By using higher-order projections, Chen *et al.* [5] proposed the Enhanced Projection-Combined Principal Component Analysis (E(PC)^{2}A) in which additional features of the images are jointly accounted for in the process of face recognition. More recent works derived from PCA include the methods of Jiang *et al.* [6] on eigenfeatures, Wang *et al.* [7] and Al-Arashi *et al.* [8] that combine PCA with support vector machines and genetic algorithms, respectively.

Other multivariate analysis methods have been also applied to the problem of face recognition, such as the Singular Value Decomposition [9], ICA [10], LDA in one or two dimensions [11–13] and the LaplacianFaces [14], together with different classifiers. Although claiming to be superior to PCA-based methods, these approaches suffer from important drawbacks, mainly when the training sample is composed by a single image per person.

Differently from PCA, the 2DPCA technique and its variations are based on the analysis of the images as matrices instead of converting them to unidimensional vectors. After the publication of the original method [1], several derived method have been proposed in the past decade that follows the same approach. For instance, Kong *et al.* used two sets of projection directions in order to simultaneously account for the rows and columns of the images. Zuo *et al.* [15] proposed an assembled matrix distance metric to measure the distance between two feature matrices obtained through 2DPCA. The extension of (PC)^{2}A to two dimensions was presented by Jun-Bao *et al.* [16]. More recently, Mohammed *et al.* [17] tackled the problem of face recognition based on the bidirectional two-dimensional principal component analysis (B2DPCA) and neural networks. Hou *et al.* [18] proposed the local two-dimensional principal component analysis (L2DPCA) which more effectively characterizes the modes of facial variability and improves dimensionality reduction.

The large number of works derived from 2DPCA motivates the investigation of factor analysis as an alternative to principal components in the main core of the method. Previous works have already shown that FA is superior to PCA in many imaging applications [19–21]. In the next section we present the two-dimensional factor analysis, a natural extension of FA that is more effective to face recognition when compared to 2DPCA.

2. METHODS

Factor analysis is a multivariate method that aims at representing a set of p variables as linear combinations of m hypothetical constructs called factors[1]. A factor can be seen as a non-observable variable associated with a cluster of correlated observed variables.

[1]Compare with the PCA model in which a component is a linear combination of the original variables

2.1. Proposed model

In the proposed two-dimensional factor analysis model, the original variables $\mathbf{Y}$ of a sample of K subjects are the $r \times c$ images themselves, with r rows and c columns:

$$\mathbf{Y} = \mathbf{F}\mathbf{A}' + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_m)$ is a set of m *common factors*, each one being a r-dimensional column vector, $\mathbf{A}$ is a $c \times m$ matrix of *loadings* and $\boldsymbol{\epsilon}$ is the *unique factor* matrix or residual terms that account for the portion of $\mathbf{Y}$ that is not common to other variables. For applications in which data reduction is required, $\boldsymbol{\epsilon}$ is neglected and it is assumed that the common factors are sufficient to account for most of the correlation embedded in the model. Variables $\mathbf{Y}$ are standardized, so that their expected values are 0 and each column of $\mathbf{Y}$ in the sample is an unitary vector, i.e.

$$\sum_{k=1}^{K} Y_{ij}^{(k)} = 0, \forall i, j \quad and \quad \sum_{i=1}^{r} (Y_{ij}^{(k)})^2 = 1, \forall j, k. \tag{2}$$

The expected value of $\mathbf{F}$ is also the null matrix and it is assumed that the factors can be chosen to be uncorrelated.

2.2. Computation of loadings

Let $\mathbf{G}$ and $\mathbf{S}$ denote, respectively, the image correlation (scatter) matrix and the factor correlation matrix. Since the factors are uncorrelated, $\mathbf{S} = \mathcal{E}(\mathbf{F}'\mathbf{F}) = \mathbf{I}$ and $\mathbf{G}$ can be defined as:

$$\mathbf{G} = \mathcal{E}(\mathbf{Y}'\mathbf{Y}) = \mathcal{E}(\mathbf{A}\mathbf{F}'\mathbf{F}\mathbf{A}') = \mathbf{A}\mathcal{E}(\mathbf{F}'\mathbf{F})\mathbf{A}' = \mathbf{A}\mathbf{A}', \tag{3}$$

where $\mathcal{E}(\dot{})$ denotes the expected value. From (3), the loadings can be determined using spectral decomposition on $\mathbf{G}$:

$$\mathbf{G} = \mathbf{A}\mathbf{A}' = \mathbf{Q}\mathbf{L}\mathbf{Q}', \tag{4}$$

from which $\mathbf{A}$ is computed as

$$\mathbf{A} = \mathbf{Q}\mathbf{L}^{1/2}, \tag{5}$$

where $\mathbf{L}^{1/2} = diag(L_1^{1/2}, \ldots, L_p^{1/2})$ is the diagonal matrix with the square root of the eigenvalues of $\mathbf{G}$ and $\mathbf{Q}$ is the matrix of corresponding eigenvectors.

2.3. Rotation of loadings

A remarkable property of factor analysis is that the loading matrix can be rotated and still be able to represent the covariance of the original dataset [22]. The rotation of loadings plays an important role in factor interpretation, as it is possible to obtain a matrix that assigns few high loadings for each variable, keeping the other loadings small. If such matrix is obtained, each variable will be related to a single factor or at least to few ones. Since the variables are related to pixels in the image, the resulting factors can be visually identified as regions associated with facial features. The main algorithms used for rotation are the *varimax* and *quartimax*, in the case of orthogonal rotation, and the *promax* algorithm for oblique rotation [23].

2.4. Factor scores

Once the loadings are determined and rotated, the original data should be projected onto the factor space, so that a set of *scores* can be computed and used to represent the images in a lower-dimensional space. From (1) and (5) it follows that

$$\mathbf{F} = \mathbf{YAL}^{-1}. \qquad (6)$$

Factor scores can be used for classification, just as any other feature vector. In this paper we applied a simple nearest neighbor classifier with the purpose of comparing the results of 2DFA to the ones obtained by 2DPCA. Let $\mathbf{F}^{(x)}$ be the factor scores computed for image $\mathbf{Y}^{(x)}$ that should be assigned to one of the K classes ω_k of individuals in the sample $\mathbf{Y}^{(1)}, \ldots \mathbf{Y}^{(K)}$, with corresponding scores $\mathbf{F}^{(1)}, \ldots \mathbf{F}^{(K)}$. Thus, the decision function can be defined as

$$\omega_k = \arg\min_k \sum_{i=1}^{r} \sum_{j=1}^{m} (\mathbf{F}_{ij}^{(x)} - \mathbf{F}_{ij}^{(k)})^2. \qquad (7)$$

2.5. Complexity analysis

The most expensive procedure in the method is the computation of the eigenvectors of the sample correlation matrix, which is $O(c^3)$, i.e., it will present cubic complexity with respect to the number of columns of the images. The rotation of loadings is an iterative procedure with fast convergence, each iteration being $O(m^2)$ with respect to the number of factors m, a usually small number. The computation of the scores requires the inversion of $\mathbf{L}$, which is $O(m)$ since it is a diagonal matrix. Therefore, the overhead of computational time required by 2DFA when compared to 2DPCA is the rotation of loadings, a procedure that is executed only once during the training step.

3. MATERIALS

The experiments were conducted over 2 public databases. The AT&T (formerly ORL) Database [24] is composed of 400 images of 40 distinct subjects, taken under controlled illumination conditions and background. For some subjects, the images were taken at different times, varying the illumination, facial expressions and the use of glasses. All images are in grayscale, with a resolution of 92×112 pixels. The extended Yale Face Database B [25] contains 16,128 single light source images of 28 subjects taken under 576 combinations of pose and directed illumination. For each subject in a particular pose, an image with ambient (diffuse) illumination was also captured. All images are in grayscale, with a resolution of 112×92 pixels.

Table 1. Recognition rates obtained with the AT&T database for different numbers of components (m).

m	2DPCA	2DFA		
		Varimax	Quartimax	Promax
2	**74.86**	75.81	**76.92**	**75.31**
5	73.14	72.14	74.25	58.53
8	70.14	73.69	75.86	49.89
11	69.19	75.22	76.00	46.22
14	68.08	76.53	74.92	39.50
17	68.00	**76.86**	75.31	34.00
20	67.36	76.17	74.61	29.58

4. EXPERIMENTAL PROCEDURE AND RESULTS

Two sets of experiments were conducted in this study, in order to investigate the behavior of the proposed method when compared to 2DPCA, under different pose and illumination conditions. The 2DPCA and 2DFA methods were implemented using IDL language (Research Systems) and run in an Intel Core i3-2367M processor, 1.4 GHz, 4GB of RAM under 64-bit Windows 7 operating system. The varimax, quartimax and promax rotation algorithms were implemented as specified by Reyment and Joreskog [23].

4.1. Experiments using the AT&T Database

In the first set of experiments, the AT&T Database was used in a leave-p-out cross-validation design. Since the sample is composed of only 10 images per subject, one image from each person was used for training and the remaining images used to validate the method. The experiment was repeated 10 times, each one using a different set of 40 images as the training set, with a single sample per person. The choice of the images for the training set followed the numbering of the files in the database, so that the k-th experiment used the k-th image from each subject. The test sample was therefore composed by 360 images. For each training sample, the 2DPCA and 2DFA based on 3 different rotation methods were tested varying the number of components/factors from 2 to 20.

The recognition rates obtained in the set of experiments were averaged and the most relevant ones are shown in Table 1. As can be seen, the 2DPCA was always inferior to 2DFA based on orthogonal rotation.The best recognition rate was obtained for the 2DFA method using quartimax rotation and 2 factors (76.92%), followed by the 2DFA method with varimax (76.86%, 17 factors) and with promax (75.31%, 2 factors). The 2DPCA was the least effective method, yielding a recognition rate of 74.86% for 2 components. Paired t-tests show that 2DFA with quartimax is significantly better than 2DPCA at the level of p=0.005. An interesting behavior of the promax oblique rotation algorithm for this dataset is that the recognition rate drops considerably as the number of factors used to represent the subjects increases. Differently

Table 2. Recognition rates obtained with the Yale database for different numbers of components (m).

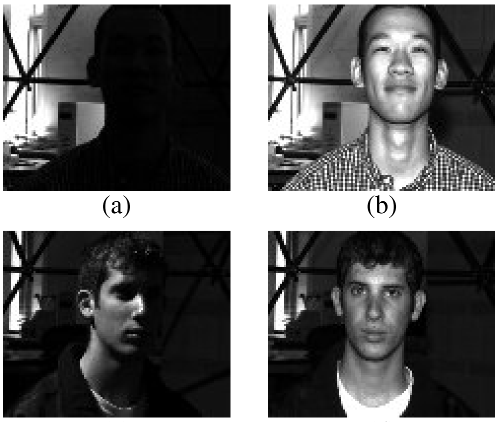
m	2DPCA	2DFA		
		Varimax	Quartimax	Promax
2	44.32	30.05	30.19	35.42
5	51.58	42.55	44.05	57.76
8	60.56	42.63	45.61	57.40
11	61.86	**51.52**	53.59	**71.91**
15	65.01	50.30	**54.54**	66.75
17	**65.65**	49.86	53.43	64.77
20	65.39	50.87	53.71	62.23

from PCA in which the variance tends to be better represented by larger number of components, factor analysis is more sensitive to the choice of the number of factors that should be considered, since correlation is better represented when factors indeed cluster sets of correlated variables. Increasing the number of factors may cause the resulting clusters to be partitioned in a unnatural way, causing the recognition rate to drop. For 2DPCA and 2DFA, there is a tendency for the recognition rates to drop as more components/factors are used. It should be noticed that in the two-dimensional approaches, a component/factor is actually a vector, so that the amount of information they represent are much larger when compared to the classical PCA and FA models.

4.2. Experiments using the Yale Database

A second set of experiments used the Extended Yale B Database, with the training set being composed of 28 frontal images, with uniform illumination, and the remaining 16,100 images used to validate the methods. A cross-validation design was not implemented since the number of samples in the dataset was sufficiently large to yield statistically significant results. Furthermore, the choice of the frontal images to compose the training set is justified, as in most applications the face datasets used as reference are acquired with the users in neutral pose and uniform illumination. The 2DPCA and 2DFA methods were tested varying the number of components/factors from 2 to 20.

The recognition rates obtained in the experiments are shown in Table 2. The best recognition rate was obtained for the 2DFA method using promax rotation and 11 factors (76.92%), followed by the 2DPCA method (65.65%, 17 components). The difference in the recognition rate between the 2DFA and the 2DPCA methods in this set of experiments was much larger than the one observed with the AT&T dataset. The orthogonal rotation did not yield good results in this case. Also, differently from what was observed in the previous experiments, an increase in the number of components/factors was associated with an increase in the recognition rate. This behavior can be partially explained by the nature of the Yale database. Whereas the AT&T dataset presents roughly uni-



(a) (b) (c) (d)

Fig. 1. Examples of images correctly classified by 2DFA but misclassified by 2DPCA ((a) and (c)). The samples used for training are respectively given in (b) and (d).

form illumination and less variation in pose, the Yale sample is characterized by large variations in both aspects. In this sense, the model may require more factors or components to represent this level of variability. Furthermore, the promax algorithm may be more adequate to capture the nonlinear relationships between the cluster of factors, which may occur when severe illumination and pose variability are under analysis. Fig. 1 shows 2 examples of successful classification by the 2DFA method for which the 2DPCA failed due to large illumination and pose differences. Since the 2DFA method accounts for the correlation among parts of the image, illumination differences are attenuated.

With respect to computational time, the 2DPCA with 11 components took on average 0.31 s to compute the eigenvectors and 0.02 s to classify a single image. The 2DFA methods using 11 factors took on average 0.50 s to compute the rotated loadings and 0.02 s to classify an image. Since the computation of the loadings is done only once in the training step, the 2 methods can be considered equally efficient procedures for face recognition.

5. CONCLUSION

A novel method for data reduction, the Two-dimensional Factor Analysis, was proposed. The method is an extension of the factorial analytic model, in which the correlations between variables are represented by a new variable set of lower cardinality. Applied to classification problems in two-dimensional signals, such as face recognition, 2DFA was able to provide a concise description of the data and additionally presented superior results when compared to principal component analysis. Since PCA is still the basis for many state-of-the-art methods, the investigation of 2DFA as an alternative model is a vast field for future work.

The current results are preliminary in nature and much work remains, including studies relating the effects of registration to the effectiveness of face recognition, the application of the method to medical imaging and visual information retrieval, as well as the association of the method with different distance measurements and classifiers. The results obtained with face recognition are nevertheless encouraging. It showed the ability of 2DFA to capture the relationship among facial characteristics under different illumination and pose conditions, based on the analysis of small samples composed of a single image per person.

REFERENCES

[1] J. Yang, D. Zhang, A. Frangi, and J. Yang, "Two-dimensional pca: a new approach to appearance-based face representation and recognition," *IEEE T Pattern Anal*, vol. 26, no. 1, pp. 131–7, 2004.

[2] M. Kirby and L. Sirovich, "Application of the karhunen-love procedure for the characterization of human faces," *IEEE T Pattern Anal*, vol. 12, no. 1, pp. 103–8, 1990.

[3] M. Turk and A. Pentland, "Eigenfaces for recognition," *J Cognitive Neurosci*, vol. 3, no. 1, pp. 71–86, 1991.

[4] J. Wu and Z. Zhou, "Face recognition with one training image per person," *Pattern Recogn Lett*, vol. 23, no. 14, pp. 1711–19, 2002.

[5] S. Chen, D. Zhang, and Z. Zhou, "Enhanced (pc) 2 a for face recognition with one training image per person," *Pattern Recogn Lett*, vol. 25, pp. 1173–81, 2004.

[6] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE T Pattern Anal*, vol. 30, no. 3, pp. 383–94, March 2008.

[7] C. Wang, L. Lan, Y. Zhang, and M. Gu, "Face recognition based on principle component analysis and support vector machine," in *International Workshop on Intelligent Systems and Applications*, 2011, pp. 1–4.

[8] H. Al-Arashi, H. Ibrahim, and S. Azmin, "Optimizing principal component analysis performance for face recognition using genetic algorithm," *Neurocomput*, vol. 128, pp. 415–20, 2014.

[9] D. Zhang, S. Chen, and Z. Zhou, "A new face recognition method based on svd perturbation for single example image per person," *Appl Math Comput*, vol. 163, no. 2, pp. 895 – 907, 2005.

[10] J. Bartlett, M. Movellan and T. Sejnowski, "Face recognition by independent component analysis," *IEEE T Neural Networ*, vol. 13, no. 6, pp. 1450–64, 2002.

[11] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE T Pattern Anal*, vol. 19, no. 7, pp. 711–20, 1997.

[12] M. Li and B. Yuan, "2d-lda: A statistical linear discriminant analysis for image matrix," *Pattern Recogn Lett*, vol. 26, no. 5, pp. 527–32, 2005.

[13] Q. Gao, L. Zhang, and D. Zhang, "Face recognition using flda with single training image per person," *Appl Math Comput*, vol. 205, no. 2, pp. 726 – 34, 2008.

[14] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using laplacianfaces," *IEEE T Pattern Anal*, vol. 27, no. 3, pp. 328–40, 2005.

[15] W. Zuo, D. Zhang, and K. Wang, "An assembled matrix distance metric for 2DPCA-based image recognition," *Pattern Recogn Lett*, vol. 27, pp. 210–6, 2006.

[16] J. Li, S. Chu, and J. Pan, "2d(pc)2 a for face recognition with one training image per person," in *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2007, vol. 1, pp. 131–4.

[17] A. Mohammed, R. Minhas, Q. Wu, and M. Sid-Ahmed, "Human face recognition based on multidimensional {PCA} and extreme learning machine," *Pattern Recogn*, vol. 44, pp. 2588 – 97, 2011.

[18] J. Hou, Y. Wang, and J. Zuo, "A novel matrix-based method for face recognition," *Neural Comput Appl*, vol. 23, no. 7-8, pp. 2261–5, 2013.

[19] A. Machado, J. Gee, and M. Campos, "Visual data mining for modeling prior distributions in morphometry," *IEEE Signal Proc Mag*, vol. 21, no. 3, pp. 20–7, 2004.

[20] J. Simon, D. Prince, J. Elder, J. Warrell, and F. Felisberti, "Tied factor analysis for face recognition across large pose changes," *BMVC*, pp. 889–98, 2006.

[21] B. Tunc, V. Dagli, and M. Gokmen, "Class dependent factor analysis and its application to face recognition," *Pattern Recogn*, vol. 45, no. 12, pp. 4092 – 102, 2012.

[22] A. Rencher, *Methods of Multivariate Analysis*, John Wiley & Sons, New York, 2002.

[23] R. Reyment and K. Joreskog, *Applied Factor Analysis in the Natural Sciences*, Cambridge University Press, New York, 1993.

[24] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in *IEEE Workshop on Applications of Computer Vision*, 1994, pp. 138–42.

[25] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE T Pattern Anal*, vol. 23, no. 6, pp. 643–60, 2001.