FRACTIONAL EIGENFACES

T.B.A. de Carvalho[1,2], *M.A.A. Sibaldo*[1,2], *I.R. Tsang*[2,3], *G.D.C. Cavalcanti*[2], *I.J. Tsang*[4] *and J. Sijbers*[3]

[1]Unidade Acadêmica de Garanhuns, Universidade Federal Rural de Pernambuco, Brazil
[2]Centro de Informática, Universidade Federal de Pernambuco, Recife, Brazil
[3] iMinds-Vision Lab, University of Antwerp, Belgium
[4]Alcatel-Lucent, Bell Labs, Antwerp, Belgium
{tbac,maas2,tir,gdcc}@cin.ufpe.br, ing-jyh.tsang@alcatel-lucent.com, jan.sijbers@uantwerpen.be

ABSTRACT

The proposed Fractional Eigenfaces method is a feature extraction technique for high dimensional data. It is related to Fractional PCA (FPCA), which is based on the theory of fractional covariance matrix, and it is an extension of the classical Eigenfaces. Like FPCA, it computes projections for a low dimensional space from the fractional covariance matrix and similar to the Eigenfaces, it is suited for high dimensional data. Moreover, the proposed technique extends the fractional transformation of the data for more stages of the feature extractions than FPCA. The Fractional Eigenfaces is evaluated in three different face databases. Results show that it achieves a higher accuracy rate than FPCA and Eigenfaces according to the Wilcoxon hypothesis test.

Index Terms— Face recognition, Principal component analysis, Fractional covariance matrix, Dimensionality reduction.

1. INTRODUCTION

Face recognition is typically a high-dimensional problem. It is common to project a face image to a low-dimensional space that is able to preserve or even increase the recognition rate. Several methods exist for dimensionality reduction for face recognition that are extensions of the classical Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). Structured Sparse LDA (SSLDA) [1], which is a fusion of the Supervised LDA and Structured Sparse PCA (SSPCA); Incremental SDA (ISDA) [2] is an extension of the Subclass Discriminant Analysis (SDA) and Incremental LDA (ILDA), both direct extensions of LDA, ISDA takes advantage of the high recognition accuracy of SDA and low time complexity of ILDA; and Average Invariant Factor [3], which is a generalization of Intrinsic Discriminate Analysis (IDA).

This work was partially supported by Brazilian agencies CNPq, CAPES, FACEPE and by the Flemish Government Agency for Innovation by Science and Technology, Belgium through the SBO TOMFOOD project.

A common characteristic for PCA-based and LDA-based techniques is that both methods depend on the covariance matrix. One of the most recently proposed PCA-based methods for dimensionality reduction is the Fractional Principal Component Analysis (FPCA) [4]. In contrast to most of related methods, the FPCA computes data projections from a $fractional$ covariance matrix. Here, we propose an extension of this method for high dimensional data called Fractional Eigenfaces, since it is inspired by both Eigenfaces and Fractional PCA. We also extend the fractional transformation beyond the covariance matrix calculation.

In Section 2, a short overview of the feature extraction method is presented. The proposed Fractional Eigenfaces is described in Section 3. Experiments using three different face datasets are described and analyzed in Section 4, which is followed by the conclusion.

2. RELATED METHODS

In order to compute the discriminant projections of the data, the PCA method calculates the directions of maximal variance within the original data space. These directions are defined as eigenvectors of the data covariance matrix; the inner product of a data vector with an eigenvector is a new extracted feature. This feature is supposed to be more discriminant. Then, k features of a pattern are extracted using PCA such that they are projections from the original pattern into the space of the k eigenvectors with the largest eigenvalues.

Let

$$X = [\mathbf{x}_1 \ldots \mathbf{x}_n] \tag{1}$$

be a data matrix for n samples in which each column $\mathbf{x}_i$ is a pattern, $i = 1, \ldots, n$; in the case of images, all columns of the image are stacked in order to form a single column vector. The image contains m samples, hence, the vector x is a $m \times 1$ column vector. Then, the $m \times m$ data covariance matrix is computed as:

$$C_{m\times m} = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}_i - \bar{\mathbf{x}}\right)\left(\mathbf{x}_i - \bar{\mathbf{x}}\right)^T, \tag{2}$$

and

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \tag{3}$$

is the data mean vector. Let E be the $m \times k$ projection matrix such that each element of its columns, $\mathbf{e}_i$, consisting of eigenvectors of C corresponding to the k largest eigenvalues:

$$E_{m\times k} = [\mathbf{e}_1 \ldots \mathbf{e}_k]. \tag{4}$$

The dimensionality reduction (or feature extraction) takes place by projecting an input pattern $\mathbf{x}_i$ from a m-dimensional space to a k-dimensional space, $k < m$, such that $\mathbf{x}'_i$ is the project pattern:

$$\mathbf{x}'_i = E^T(\mathbf{x}_i - \bar{\mathbf{x}}). \tag{5}$$

2.1. Fractional PCA (FPCA)

The FPCA method is based on the theory of the fractional covariance matrix [4]. It computes the projection directions as the eigenvectors of the fractional (r-order) covariance matrix, which is defined as:

$$C^r_{m\times m} = \frac{1}{n}\sum_{i=1}^{n}\left((\mathbf{x}_i)^r - (\bar{\mathbf{x}})^r\right)\left((\mathbf{x}_i)^r - (\bar{\mathbf{x}})^r\right)^T, \tag{6}$$

with

$$(\mathbf{x}_i)^r = [(x_{i1})^r \ldots (x_{im})^r]^T, \tag{7}$$

C^r is an $m \times m$ matrix, like PCA, that have m eigenvectors from which k of highest eigenvalues are chosen in order to define the projection matrix E (Equation 4), and project the input data according Equation 5.

In [4], FPCA was evaluated on a face recognition task using the ORL database. The value $r = 0.01$ was defined experimentally to provide the best accuracy rate. The experiments showed that the FPCA results are improved compared to PCA, but both methods are not able to obtain projections for high dimensional data, since they are computationally intractable [5]. In order to perform the experiments every image was resized from 92×112 to 23×28. The solution to extend PCA for high dimensional data is to use Eigenfaces.

2.2. Eigenfaces

Eigenfaces [5, 6], previously proposed in [7, 8], is an extension of PCA for high-dimensional data [9], i.e., for problems for which the number of feature m is too high [10]. In high dimensional problems it is common that $n << m$, i.e., the number of samples are much smaller than the number of features. For the ORL face database, there are $n = 400$ samples and $m = 10304$ features. It is stated in [5] that it is intractable to determining the m eigenvectors and eigenvalues of an $m \times m$ matrix for typical image sizes. Indeed the complexity of such algorithms is $O(m^3)$ according to [9, 11]. The solution for this problem, presented in [5, 9], is to not compute eigenvectors directly from the covariance matrix C (Equation 2) but from a smaller matrix D of dimension $n \times n$, then calculate the eigenvector of C from the eigenvectors of D. The D matrix is defined as:

$$D_{n\times n} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})^T(\mathbf{x}_i - \bar{\mathbf{x}}). \tag{8}$$

Let $E'_{m\times k} = [\mathbf{e}'_1 \ldots \mathbf{e}'_k]$, $k = 1, \ldots, n$, be the k eigenvectors of D with highest eigenvalues, the eigenvectors $\mathbf{e}_i$ of E (Equation 4) of the covariance matrix C, are computed from E' as follows:

$$\mathbf{e}_i = \frac{1}{(n\lambda_i)^{1/2}}[(\mathbf{x}_1 - \bar{\mathbf{x}}), \ldots, (\mathbf{x}_n - \bar{\mathbf{x}})]\mathbf{e}'_i, \tag{9}$$

with λ_i as the eigenvalue corresponding to $\mathbf{e}'_i$. Once the projection matrix E is computed, data is projected to a lower dimensional space according to Eq. 5.

Since D has only n eigenvectors, then only up to n of the m eigenvectors of C can be calculated through Equation 9. This is not a limiting factor since C has $n - 1$ meaningful eigenvectors [5], and the other $m - d + 1$ eigenvalues are zero. Once E is computed, the dimensionality reduction is performed by projecting the input data according to Equation 5. In the next section, we propose the Fractional Eigenfaces method that takes advantage of the theory of fractional covariance matrix applied to FPCA and the capabilities of Eigenfaces to compute the eigenvectors of the covariance matrix of high-dimensional data.

3. FRACTIONAL EIGENFACES

The proposed method tries to define a better projection vector that can improve the recognition rate compared to Eigenfaces. Since FPCA showed improvement over PCA [4], it is reasonable to assume that the eigenvectors of the fractional covariance matrix (Eq. 6) for high-dimensional data can improve the Eigenfaces recognition rate. Since it is intractable to compute the eigenvectors of $C^r_{m\times m}$ (Eq. 6) and there are only $n - 1$ non-zero eigenvalues, using a process similar to Eigenfaces, the $n - 1$ eigenvectors corresponding to non-zero eigenvalues are calculated from the $n \times n$ matrix D^r:

$$D^r_{n\times n} = \frac{1}{n}\sum_{i=1}^{n}\left((\mathbf{x}_i)^r - (\bar{\mathbf{x}})^r\right)^T\left((\mathbf{x}_i)^r - (\bar{\mathbf{x}})^r\right), \tag{10}$$

using the same definition as in Eq. 7 for $(\mathbf{x}_i)^r$. Let $E'_{m\times k} = [\mathbf{e}'_1 \ldots \mathbf{e}'_k]$, $k = 1, \ldots, n$, be the k eigenvectors of D^r with highest eigenvalues, the eigenvectors $\mathbf{e}_i$ of E (Equation 4) of the covariance matrix C^r, are computed from E' as shown in Equation 9.

In contrast to all previous reviewed methods for which the projection of the new patterns is performed through Equation

Fig. 1. Two images from each dataset, from left to right: Yale, AT&T, and Sheffield.

5, the proposed Fractional Eigenfaces obtain the new vector of features $\mathbf{x}'_i$ for the input pattern $\mathbf{x}_i$ using:

$$\mathbf{x}'_i = E^T \left((\mathbf{x}_i)^r - (\bar{\mathbf{x}})^r \right). \tag{11}$$

Since the eigenvectors E are calculated for the fractional covariance matrix C^r, they do not represent directions of maximal variance for the input dataset X (Eq. 1) but they represent the direction of the maximal variance for the fractional transformed data X^r:

$$X^r = [((\mathbf{x}_1)^r - (\bar{\mathbf{x}})^r) \ldots ((\mathbf{x}_n)^r - (\bar{\mathbf{x}})^r)], \tag{12}$$

therefore the patterns must be transformed using Equation 11 before being projected to the new space of features. However, this transformation was not considered by the authors of the FPCA method [4].

Fractional Eigenfaces improves FPCA in two ways: first, the proposed method is able to compute the eigenvectors of the fractional covariance matrix for high dimensional data; second, we have proposed a new equation to extract features that considers the fractional transformation of the data. From another point of view, it may be considered that Fractional Eigenfaces extends Eigenfaces to the field of the theory of fractional covariance matrix. Indeed, the proposed method not only combines FPCA and Eigenfaces, but it also extends the fractional transformation to the final feature extraction equation. In Section 4, the Fractional Eigenfaces is evaluated and compared experimently against FPCA and Eigenfaces.

4. EXPERIMENTS AND RESULTS

The experiments were performed on three well-known face dataset: Yale, AT&T (formerly ORL), both used in [3, 4], and Sheffield (formerly UMIST). The Yale face images were resized to 92 $\times$112, which is the size of the images in the other datasets. The Yale dataset has 15 subjects and 11 images per subject, which vary mainly in the illumination conditions. The AT&T dataset has 40 subjects and 10 images per subject, which slightly vary on pose. The Sheffield images vary mainly on the head rotation and have 20 subjects, the number of images per subject vary from 19 to 48, the total number of face images is 574. Fig. 1 shows some examples of images from each dataset.

For every holdout experiment the accuracy of the Nearest Neighbor classifier (1-NN) was calculated and averaged

Table 1. Columns: n = number of dimensions, mean accuracy $\pm$ standard deviation for Eigenfaces, FPCA, and Fractional Eigenfaces.

n	Eigenfaces	FPCA	Frac. Eigen.
	Yale		
1	0.204 ± 0.043	0.256 ± 0.043	0.349 ± 0.053
5	0.671 ± 0.033	0.707 ± 0.029	0.699 ± 0.031
10	0.720 ± 0.035	0.756 ± 0.033	0.741 ± 0.031
15	0.732 ± 0.028	0.763 ± 0.026	0.750 ± 0.024
20	0.749 ± 0.027	0.747 ± 0.033	0.756 ± 0.021
25	0.760 ± 0.029	0.741 ± 0.025	0.766 ± 0.021
30	0.761 ± 0.021	0.751 ± 0.028	0.780 ± 0.014
35	0.759 ± 0.021	0.756 ± 0.031	0.788 ± 0.018
40	0.766 ± 0.022	0.762 ± 0.024	0.790 ± 0.011
45	0.767 ± 0.025	0.757 ± 0.024	0.788 ± 0.014
50	0.769 ± 0.022	0.756 ± 0.030	0.789 ± 0.014
55	0.770 ± 0.022	0.758 ± 0.029	0.789 ± 0.014
60	0.769 ± 0.023	0.758 ± 0.026	0.790 ± 0.014
65	0.771 ± 0.024	0.757 ± 0.029	0.790 ± 0.014
70	0.773 ± 0.024	0.760 ± 0.028	0.792 ± 0.014
	AT&T		
1	0.120 ± 0.017	0.142 ± 0.014	0.148 ± 0.021
5	0.783 ± 0.032	0.822 ± 0.026	0.847 ± 0.025
10	0.899 ± 0.023	0.907 ± 0.022	0.922 ± 0.024
15	0.915 ± 0.026	0.923 ± 0.020	0.936 ± 0.018
20	0.917 ± 0.024	0.931 ± 0.020	0.940 ± 0.019
25	0.923 ± 0.024	0.935 ± 0.025	0.942 ± 0.020
30	0.926 ± 0.025	0.936 ± 0.024	0.944 ± 0.016
35	0.929 ± 0.019	0.936 ± 0.020	0.947 ± 0.022
40	0.928 ± 0.014	0.941 ± 0.017	0.951 ± 0.020
45	0.928 ± 0.016	0.939 ± 0.016	0.952 ± 0.017
50	0.929 ± 0.020	0.939 ± 0.016	0.951 ± 0.017
55	0.929 ± 0.020	0.938 ± 0.017	0.949 ± 0.021
60	0.930 ± 0.018	0.940 ± 0.016	0.947 ± 0.021
65	0.927 ± 0.018	0.939 ± 0.016	0.947 ± 0.019
70	0.927 ± 0.017	0.937 ± 0.016	0.948 ± 0.022
	Sheffield		
1	0.194 ± 0.022	0.203 ± 0.025	0.206 ± 0.018
5	0.903 ± 0.031	0.924 ± 0.017	0.935 ± 0.013
10	0.940 ± 0.021	0.944 ± 0.014	0.951 ± 0.017
15	0.958 ± 0.017	0.962 ± 0.014	0.965 ± 0.016
20	0.969 ± 0.015	0.969 ± 0.013	0.973 ± 0.015
25	0.970 ± 0.013	0.973 ± 0.012	0.975 ± 0.012
30	0.971 ± 0.013	0.974 ± 0.013	0.976 ± 0.010
35	0.972 ± 0.014	0.974 ± 0.012	0.973 ± 0.011
40	0.972 ± 0.014	0.973 ± 0.012	0.975 ± 0.013
45	0.973 ± 0.013	0.974 ± 0.012	0.976 ± 0.012
50	0.972 ± 0.013	0.974 ± 0.011	0.975 ± 0.011
55	0.972 ± 0.012	0.974 ± 0.012	0.976 ± 0.012
60	0.972 ± 0.012	0.974 ± 0.011	0.976 ± 0.011
65	0.973 ± 0.013	0.975 ± 0.012	0.976 ± 0.011
70	0.974 ± 0.012	0.975 ± 0.012	0.977 ± 0.011

for 10 runs; for each run half of the images from each subject were randomly chosen for the test set and the remaining images for the training set. In the case of an odd number of images, the extra image was added to the test set. The same training and test sets were used for every method into a hold-out run. The value of $r = 0.01$ was used for the FPCA and Fractional Eigenfaces for every experiment, since this is the value used in [4] that presented the best accuracy. The results of the experiments are presented in Table 1 and summarized in the plots shown in Fig. 2.

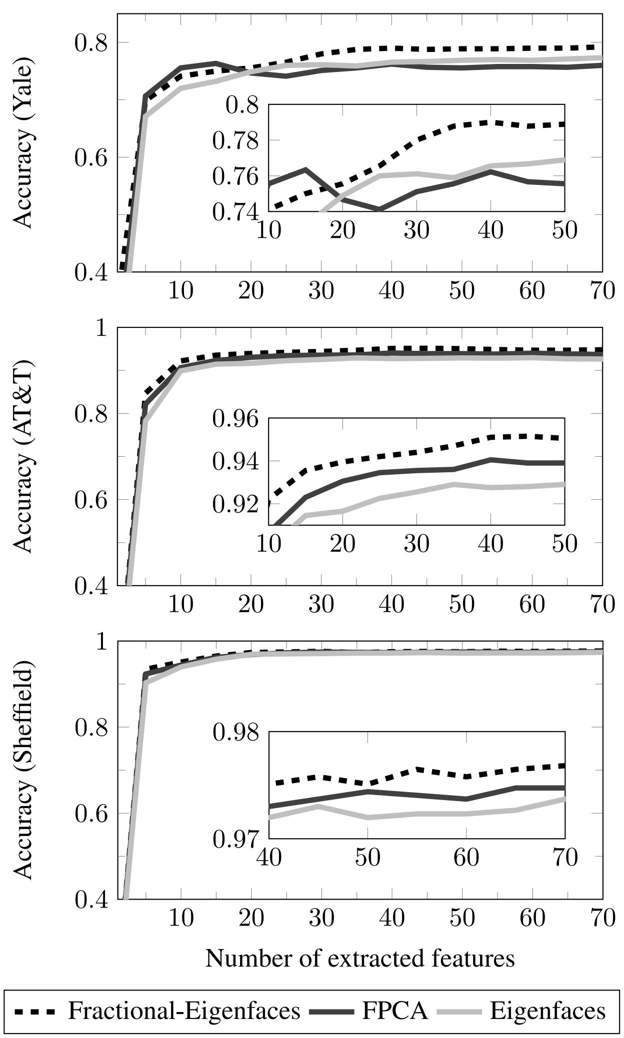

Fig. 2. Accuracy rate per number of features for the three dataset: Yale (top), AT&T (middle), Sheffield (bottom).

The classification accuracy was measured for 1, 5, 10, 15, ..., 70 features, as shown in Table 1. For the Sheffield dataset the proposed method performed slightly better than FPCA and Eigenfaces for every number of features. In most of the cases, this difference is smaller than 1%, for 5 features this difference is 3% for Eigenfaces and 1% for FPCA. However, the recognition rate of this database of around 97% and it can hardly be improved. For the AT&T dataset the proposed method presented an improvement of recognition rate of about 2 to 6% compared to Eigenfaces and 1 to 3% when compared to FPCA, the maximum difference between accuracy rates also occurs when only 5 features are selected. For this database the proposed Fractional Eigenfaces presents a better accuracy in every number of feature. For the Yale dataset, when the number of selected features is at least 30, the proposed method has 2% better accuracy rate than Eigenfaces and 3% than FPCA.

Finally a Wilcoxon signed-rank nonparametric hypothesis test [12] was performed, by comparing 45 (15 for each dataset) pairwise accuracy rates for each feature extraction method. The obtained p-values are: 9.45×10^{-7} for FPCA and Fractional Eigenfaces; 5.06×10^{-9} for Eigenfaces and Fractional Eigenfaces; and 0.01 for FPCA and Eigenfaces. For all cases the p-values are less than the significance level of 0.05, then the null hypothesis can be rejected. The null hypothesis states that the feature extraction methods presents the same accuracy rate for the 1-NN classifier. Therefore with this results, it can be concluded that the proposed Fractional Eigenfaces method presents a higher accuracy rate than previous methods from which it is derived.

5. CONCLUSIONS

This paper proposed the Fractional Eigenfaces technique which is a new method for dimensionality reduction suited for high-dimensional data, such as image databases. The method uses the theory of fractional covariance matrix similary to the Fractional PCA (FPCA) method. However, Fractional Eigenfaces improves FPCA in two ways: the proposed method is able to perform feature extraction in high dimensional data; and a new way to extract features that considers the fractional transformation over the data is presented. From another point of view, it may be considered that Fractional Eigenfaces extends Eigenfaces using the theory of fractional covariance matrix. The proposed method combines ideas from FPCA and Eigenfaces, since it extends the fractional transformation from the training part to the entire process of feature extraction. For future work, this step can be also extended to FPCA. Experimental results on three different face image databases shows that the proposed method presents better results than the two previous methods for the face recognition task. The pairwise Wilcoxon hypothesis test was used to show this improvement.

6. REFERENCES

[1] Zhen Cui, Shiguang Shan, Haihong Zhang, Shihong Lao, and Xilin Chen, "Structured sparse linear discriminant analysis," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, Sept 2012, pp. 1161–1164.

[2] H. Lamba, T.I. Dhamecha, M. Vatsa, and R. Singh, "Incremental subclass discriminant analysis: A case study in face recognition," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, Sept 2012, pp. 593–596.

[3] Zhongxuan Luo, Hao Sun, Xin Fan, and Jielin Zhang, "Face recognition using average invariant factor," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, Sept 2012, pp. 1429–1432.

[4] Chaobang Gao, Jiliu Zhou, and Qiang Pu, "Theory of fractional covariance matrix and its applications in PCA and 2D-PCA," *Expert Systems with Applications*, vol. 40, no. 13, pp. 5395 – 5401, 2013.

[5] Matthew Turk and Alex Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991.

[6] P.N. Belhumeur, J.P. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 711–720, Jul 1997.

[7] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Am. A*, vol. 4, no. 3, pp. 519–524, 1987.

[8] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 1, pp. 103–108, Jan 1990.

[9] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[10] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[11] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, *Numerical Recipes in C, 2nd ed.: The Art of Scientific Computing*, Cambridge University Press, New York, NY, USA, 1992.

[12] Mario F. Triola, *Elementary Statistics, 9th ed.*, Springer Series in Statistics. Addison Wesley, 2003.