SUPERVISED FRACTIONAL EIGENFACES

T.B.A. de Carvalho[1,2], *A.M.Costa*[1], *M.A.A. Sibaldo*[1,2], *I.R. Tsang*[2], *G.D.C. Cavalcanti*[2]

[1]Unidade Acadêmica de Garanhuns, Universidade Federal Rural de Pernambuco, Brazil
[2]Centro de Informática, Universidade Federal de Pernambuco, Recife, Brazil
{tbac,maas2,tir,gdcc}@cin.ufpe.br

ABSTRACT

Supervised Fractional Eigenfaces (SFE) is an extension of Principal Component Analysis (PCA), which uses the fractional covariance matrix, class label information, and nonlinear data transformation to extract discriminant features. The proposed method combines techniques of two state-of-the-art feature extractors: Fractional Eigenfaces and Dual Supervised PCA. Supervised Fractional Eigenfaces was evaluated in three known face datasets and it achieved significant smaller recognition error.

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

1. INTRODUCTION

Eigenfaces is a well-known technique for face recognition that is often used as benchmark. It is a extension of Principal Component Analysis (PCA), which deals with high dimensional data, *i.e.*, dataset that have much more dimensions than samples [1]. PCA finds linear projections that maximizes data variance, dimensionality reduction takes place using only a few projections. There are still many state-of-the-art approaches to extend PCA, some of them focus on face images. MPCA is PCA for mixed size image datasets, it is able to minimize reconstruction error for face images even if images have different sizes [2]. Face hallucination based on PCA [3] is able to reconstruct a high resolution face image from the corresponding low resolution input, providing lower reconstruction error and better visual quality. Woo et al. proposed a fast GPU parallel implementation of PCA for face recognition [4]. Yuan et al. proposed a color facial image denoising based on PCA [5].

Supervised PCA (SPCA) [6] is an extension of PCA that aims to find projections that are more discriminant for the classification task. Also a version for high dimensional data has been proposed, the Dual SPCA (DSPCA). DSPCA have never been used for face recognition, if so it could also be called Supervised Eigenfaces by analogy. Fractional PCA (FPCA) was proposed by Gao et al. [7]. It is based on the theory of the fractional covariance matrix and it finds linear projections that improve the face recognition accuracy. FPCA was restrict to classify small face images due the limitation of PCA to deal with high dimensional data. Fractional Eigenfaces [8] not only solve this limitation of FPCA, but also generate non-linear projections that are more suitable for face recognition. In this paper, we proposed the Supervised Fractional Eigenfaces (SFE), which combined properties of DSPCA and Fractional Eigenfaces to produce projections with smaller classification error for the face recognition task.

The rest of this paper is organized as follows. Section 2 describes the major methods that are related with the proposed method. In Section 3, we described the Supervised Fractional Eigenfaces. Experiments using three different face datasets are described and analyzed in Section 4. Finally, Section 5 presents the conclusion.

2. RELATED METHODS

For high dimensional data, we have $n << m$, *i.e.*, the number of samples n is much smaller than the number of dimensions m. Face recognition is typically in high dimensional. For this reason Eigenfaces was proposed to be able to handle the amount of data. PCA depends on computing eigenvector of the data covariance matrix with size $m \times m$. But Eigenfaces computes the same eigenvectors from a $n \times n$ matrix [1]. The computational complexity of Eigenfaces is $O(n^3)$ that is much smaller than the complexity of PCA, which is $O(m^3)$. This cubic complexity is due to the eigenvector decomposition. There is no information loss in Eigenfaces, when compared with the standard PCA, because the number of non-zero eigenvalues for a covariance matrix is up to $n - 1$ if $n < m$, [9].

The two methods described in this section also depend on the computation of the eigenvectors. The first is very similar to Eigenfaces but performs non-linear dimensionality reduction. The second is a version of Supervised PCA for high dimensional data.

This work was partially supported by Brazilian agencies CNPq, CAPES, FACEPE.

2.1. Fractional Eigenfaces (FE)

Fractional PCA (FPCA) is a version of PCA that computes its projections as eigenvectors of the fractional covariance matrix, it was proposed by Gao et al. [7]. Fractional Eigenfaces (FE), proposed by de Carvalho et al. [8], is an extension of FPCA in two ways: (1) it indirectly computes the eigenvectors of the fractional covariance matrix, reducing the computational complexity similar to Eigenfaces; and (2) performs a non-linear transformation of the data before the final projection, that is, FE performs non-linear dimensionality reduction.

Let $\mathbf{x}_i = [x_{i1} \dots x_{im}]^T$ be the i-th sample $i = 1, \dots, n$:

$$X_{m \times n} = [\mathbf{x}_1 \dots \mathbf{x}_n] \tag{1}$$

is the data matrix, where each column of X is a sample. A matrix, that is usually represented by a matrix, can be converted to a vector associating each matrix position to a vector position with the same number of elements.

Let $(\mathbf{x}_i)^r$ be a fractional transformed sample:

$$(\mathbf{x}_i)^r = [(x_{i1})^r - (\bar{x}_1)^r \dots (x_{im})^r - (\bar{x}_m)^r]^T, \tag{2}$$

with $\bar{x}_j$ as the mean of the j-th feature, $j = 1, \dots, m$; and having r as the fractional coefficient, usually $r \in [0, 1]$. The fractional transformation not only raise each feature to the power of r but also translate the data to a fractional mean. This data preprocessing step is necessary to compute the fractional covariance matrix, which is defined by $(X^r)(X^r)^T$.

$$X^r_{m \times n} = [(\mathbf{x}_1)^r \dots (\mathbf{x}_n)^r] \tag{3}$$

is the fractional transformed data matrix, each column of X^r is a fractional transformed sample.

The $n - 1$ eigenvectors corresponding to non-zero eigenvalues from the fractional covariance matrix are calculated from the $n \times n$ matrix D^r:

$$D^r_{n \times n} = \frac{1}{n}(X^r)^T(X^r). \tag{4}$$

Let $E'_{n \times k} = [\mathbf{e}'_1 \dots \mathbf{e}'_k]$, $k = 1, \dots, n$, be the k eigenvectors of D^r with highest eigenvalues. The eigenvectors of the fractional covariance matrix are computed as:

$$\mathbf{e}_p = \frac{1}{(n\lambda_p)^{1/2}}(X^r)\mathbf{e}'_p, \tag{5}$$

having λ_p as the corresponding eigenvalue of $\mathbf{e}'_p$, $p = 1, \dots, k$. Fractional Eigenfaces obtains the new vector of features $\mathbf{x}'_i$ for the input pattern $\mathbf{x}_i$ using:

$$\mathbf{x}'_i = E^T(\mathbf{x}_i)^r, \tag{6}$$

with $E_{m \times k} = [\mathbf{e}_1 \dots \mathbf{e}_k]$ as the matrix of projection vectors. It must be highlighted that transformed sample $\mathbf{x}'_i$ is not only a linear projection of $\mathbf{x}_i$, former the projection the input sample suffer a nonlinear transformation as described by Equation 2.

2.2. Dual Supervised PCA (DSPCA)

DSPCA is a version of Supervised PCA for high dimensional data [6]. It considers class information in order to find linear projections that are more discriminant for classification. This class information is inserted in the matrix S through matrix Δ, the projection vectors are dependent from the eigenvectors of S. The major restriction of DSPCA is that it can extract up to c features, where c is the number of classes of the problem. DSPCA computes the eigenvectors of S following:

$$S_{c \times c} = \Delta H[X^T X]H\Delta^T, \tag{7}$$

with $\Delta_{c \times n} = [\delta_{li}]$, where $\delta_{li} = 1$ if the i-th sample belongs to the l-th class and $\delta_{li} = 0$ otherwise, $l = 1, \dots, c$ and $i = 1, \dots, n$; $H_{n \times n} = I - n^{-1}\mathbf{h}\mathbf{h}^T$, I is the $n \times n$ identity matrix, $\mathbf{h}$ is the $n \times 1$ vector of ones. Let $U'_{c \times k} = [\mathbf{u}'_1 \dots \mathbf{u}'_k]$, $k = 1, \dots, n$, be the k eigenvectors of S with highest eigenvalues, the eigenvectors for data projections are computed as:

$$\mathbf{u}_p = \frac{1}{(\lambda_p)^{1/2}} XH\Delta^T\mathbf{u}'_p, \tag{8}$$

with λ_p, the corresponding eigenvalue of $\mathbf{u}'_p$, $p = 1, \dots, k$. DSPCA obtains the new vector of features $\mathbf{x}'_i$ for the input pattern $\mathbf{x}_i$, defined as:

$$\mathbf{x}'_i = U^T\mathbf{x}_i, \tag{9}$$

having $U_{m \times k} = [\mathbf{u}_1 \dots \mathbf{u}_k]$ as the matrix of projection vectors.

The effect of multiplying X by H is the centralization of the data around the mean. The effect of multiplying XH by Δ is to sum the samples of each class, reducing the dimensions of S to $c \times c$. Then, U is the eigenvectors of the covariance matrix for the mean vectors for each class. DSPCA implicitly assumes that samples of each class are distributed around the mean of the class. In this way, directions of maximal variance are also more discriminant. A direction that the mean vectors of each class are more spread is also more suitable for classification.

3. SUPERVISED FRACTIONAL EIGENFACES

Supervised Fractional Eigenfaces (SFE) is inspired by the capability of DSPCA to find discriminant projections, fractional covariance matrix and non-linear projections of FE. This combination is possible because FE and DSPCA improve Eigenfaces in different ways. Since SFE extends DSPCA it inherits its limitation of have maximal number of features to extract equals to the number of classes.

SFE computes the eigenvectors of S^r

$$S^r_{c \times c} = \Delta[(X^r)^T(X^r)]\Delta^T, \tag{10}$$

having the same formulation for X^r as defined in Equation 3 and Δ defined in Equation 7.

Let $V'_{c\times k} = [\mathbf{v}'_1 \dots \mathbf{v}'_k]$, $k = 1, \dots, n$, be the k eigenvectors of S^r with highest eigenvalues, the eigenvectors for the data projections are computed as:

$$\mathbf{v}_p = \frac{1}{(\lambda_p)^{1/2}}(X^r)\Delta^T \mathbf{v}'_p, \tag{11}$$

with λ_p as the corresponding eigenvalue of $\mathbf{v}'_p$, $p = 1, \dots, k$. Supervised Fractional Eigenfaces obtains the new vector of features $\mathbf{x}'_i$ for the input pattern $\mathbf{x}_i$ using:

$$\mathbf{x}'_i = V^T(\mathbf{x}_i)^r, \tag{12}$$

where $V_{m\times k} = [\mathbf{v}_1 \dots \mathbf{v}_k]$ is the matrix of projection vectors.

The XH multiplication of DSPCA is substituted by the fractional transformed data matrix (X^r), which is already centered around the fractional mean. The multiplication by Δ sum up fractional samples of each class. The effective number of data points to compute S^r is the number of classes. SFE tries to find the direction that better separates the mean vector of the classes for the fractional transformed data. Since the fractional data transformation improves face recognition, projections that better separates the mean vector of each class should improve even more the results.

4. EXPERIMENTS AND RESULTS

Experiments were performed using three well known face databases (from www.face-rec.org): Georgia Tech, Yale and ORL. The ORL dataset has 40 classes with 10 samples per class. The grayscale images have size equal to 92×112 pixels. All three datasets are composed of frontal face image with variations on face rotation and illumination. The Yale database have 15 classes and 11 samples per class, its grayscale images were cropped and resized to 92×112 pixels. The set of cropped images of the Georgia dataset (50 classes and 15 samples per class) were transformed to gray tone and resized to fit into a frame of size 91×98, the proportion of the cropped image were maintained.

We used the Nearest Neighbor classifier (1-NN) with Euclidian distance and evaluated the recognition error as the number of miss classified samples divided by the number of classified images. The plots shown in Figure 1 are the mean recognition error for 50 stratified holdout, each holdout trained with half of the images. The remaining images were used for testing. The dimensionality reduction was performed using four different methods: Eigenfaces, DSPA, FE, and the proposed Supervised Fractional Eigenfaces (SFE). The value for the fractional parameter is $r = 0.1$ for FE and SFE. The projection methods are evaluated for 1 to 40 extracted features. Note that SFE and DSPCA can extract up to 15 features for the Yale dataset, that is the number of classes of this set. The results for more than 15 feature are just copies of this result.

Figure 1 shows that the error values are smaller for the proposed method. The best results were obtained by SFE,

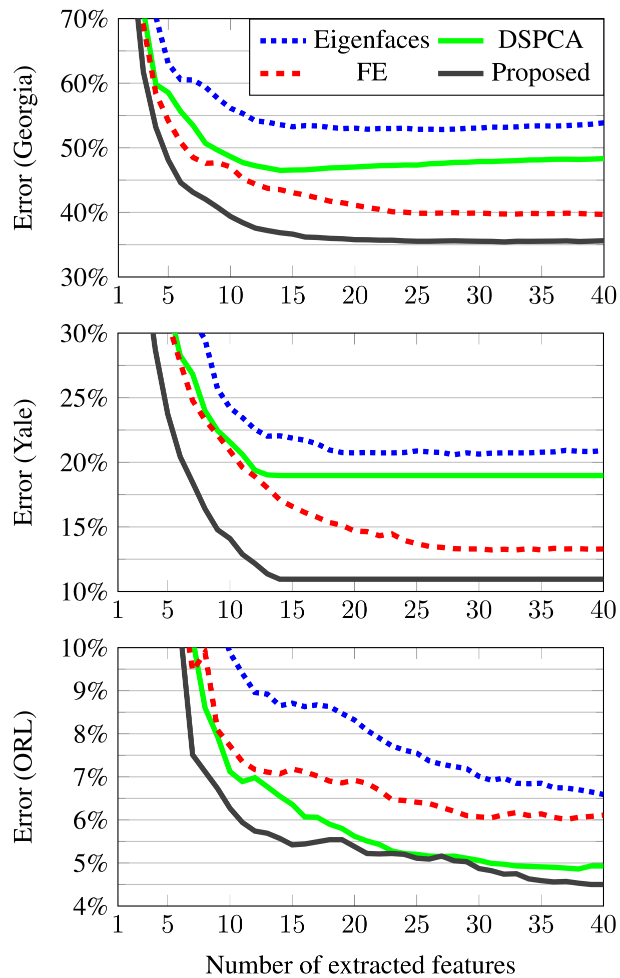

Fig. 1. Recognition error per number of extracted features for three face databases: Georgia Tech, Yale, and ORL. Using the 1-NN classifier, four feature extraction methods are compared: Eigenfaces, Dual Supervised PCA (DSPCA), Fractional Eigenfaces (FE), and the proposed method Supervised Fractional Eigenfaces (SFE).

FE, DSPCA and Eigenfaces, receptively. The smallest error was always obtained by SFE and the second smaller error by FE. For the Georgia dataset, SFE presented an error of about 5% smaller than the second smallest error. This difference is of 3% for the Yale face images, but SFE achieves the error of 11% for 15 extracted features, and is not able to extract more features, at this point the error for FE is 17%. Note that for more than 15 features the error is constant for DSPCA and SFE due to the restriction to extract more features than the number of classes. The differences of the error measures between FE and SFE is small for the ORL dataset, but it is of 0.5%. This difference can be greater for the other method and for up to 15 extracted features. The proposed SFE presents the smaller recognition error for all three datasets.

Table 1. Comparing the proposed method Supervised Fractional Eigenfaces (SFE) with Fractional Eigenfaces (FE) for three face datasets. Wins are the number of instances in 50 holdouts that SFE have smaller error than FE, Ties the number of equal errors, and Losses the number of greater errors. The p-value is the statistics for the hypothesis test.

	Wins	Ties	Losses	p-value
Georgia	46	0	4	1.8×10^{-15}
Yale	33	5	12	2.5×10^{-3}
ORL	40	1	9	9.3×10^{-6}

In order to evaluate the significance of the results a hypothesis tests were performed. For 40 extracted features the 50 error measures for SFE and FE were compared for each dataset. Only these methods were compared because they achieved smaller error than the other two methods. Results are described in Table 1. For each pair of error measured, if the error for SFE was smaller than the error for FE it was counted as a win, if equal a tie, and if greater a loss. The p-value for the sign test is shown in the last column of the table. For a level of significance of 5% the null hypothesis that the difference of the error measures between both methods has zero median is reject. SFE have significant smaller recognition error than FE for all three face datasets. Since the error difference for the DSPCA and Eigenfaces is higher we can conclude the same results for the other methods. Therefore, we can observe that the proposed method reduce significantly the recognition error.

5. CONCLUSIONS

The proposed method, Supervised Fractional Eigenfaces (SFE), is a direct extension of two state-of-the art methods: Fractional Eigenfaces (FE) and Dual Supervised PCA (DSPCA). All three methods are very relevant because they extend PCA, which is widely used in several different problems. SFE extends DSPCA [6] by using the fractional covariante matrix [7] and non-linear fractional data transformation [8], before the linear data projection.

For face recognition, both FE and DSPCA improve considerably the accuracy results. The proposed SFE method shows an even better performance, with a recognition error significantly smaller for all three face databases. The proposed method is able to generate nonlinear discriminant new features. For future work, an evaluation of the SFE on different kind of high dimensional data is of interest. Also it is intriguing to verify if is possible to improve SFE recognition rate by combining it with another state-of-the-art extension of PCA.

6. REFERENCES

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

[2] Feiyu Shi, Menghua Zhai, Drew Duncan, and Nathan Jacobs, "Mpca: Em-based pca for mixed-size image datasets," in *Image Processing (ICIP), 2014 21st IEEE International Conference on*, 2014.

[3] Jingang Shi and Chun Qi, "Face hallucination based on pca dictionary pairs," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, Sept 2013, pp. 933–937.

[4] Youngsang Woo, Cheongyong Yi, and Youngmin Yi, "Fast pca-based face recognition on gpus," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 2659–2663.

[5] Zhaojun Yuan, Xudong Xie, Xiaolong Ma, and Kin-Man Lam, "Color facial image denoising based on rpca and noisy pixel detection," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 2449–2453.

[6] Elnaz Barshan, Ali Ghodsi, Zohreh Azimifar, and Mansoor Zolghadri Jahromi, "Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds," *Pattern Recognition*, vol. 44, no. 7, pp. 1357 – 1371, 2011.

[7] 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.

[8] Tiago B. A. de Carvalho, Maria A.A. Sibaldo, Tsang I. Ren, George D. C. Cavalcanti, Tsang I. Jyh, and Jan Sijbers, "Fractional eigenfaces," in *Image Processing (ICIP), 2014 21st IEEE International Conference on*, 2014.

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