

Sparse Similarity-Based Fisherfaces

Jens Fagertun¹, David D. Gomez², Mads F. Hansen¹, and Rasmus R. Paulsen¹

¹ DTU Informatics, Image Analysis & Computer Graphics
Lyngby, Denmark

² Carlos III University, Department of Signal Theory and Communications
Madrid, Spain

Abstract. In this work, the effect of introducing Sparse Principal Component Analysis within the Similarity-based Fisherfaces algorithm is examined. The technique aims at mimicking the human ability to discriminate faces by projecting the faces in a highly discriminative and easy interpretative way. Pixel intensities are used by Sparse Principal Component Analysis and Fisher Linear Discriminant Analysis to assign a one dimensional subspace projection to each person belonging to a reference data set. Experimental results performed in the AR dataset show that Similarity-based Fisherfaces in a sparse version can obtain the same recognition results as the technique in a dense version using only a fraction of the input data. Furthermore, the presented results suggest that using SPCA in the technique offers robustness to occlusions.

Keywords: Face recognition, Sparse Principal Component Analysis, Fisher Linear Discriminant Analysis, Biometrics, Multi-Subspace Method.

1 Introduction

Recognizing a face is an important everyday task for human interaction. It is a skill we acquire before we can walk, and we are able to perform it with high accuracy using little or no effort. Due to the importance of this skill, machine aided face recognition is one of the most researched fields in image analysis. However, results reported in literature suggest that facial recognition still lacks the performance that a human operators can achieve.

In the last two decades, as advanced spectral techniques have emerged, there have been a gradual shift of the research on face recognition from geometrical towards spectral analysis. Such methods include the unsupervised method Eigenfaces [1,2] and the supervised method Fisherfaces [3]. Later on, related techniques have been proposed aiming at obtaining better classification results by developing more discriminative projections [4,5]. Most supervised methods in the literature today try to project the face representation into a subspace where a measure of global separation is maximized. However, the proposed algorithm in this paper tries to project the face representation into a series of one-dimensional subspaces where one person (class) is discriminated from all others in the population.

By solving the problem in a series of subspaces, this algorithm makes the enrollment or removal of a person easy. It is not needed to recalculate all of the existing individual subspaces when changes to the population are made. When a new person is introduced, simply a new individual subspace is added. If a person has to be removed from the database, it is only needed to remove the corresponding individual subspace.

In this article the algorithm of Similarity-based Fisherfaces proposed in [7] is extended by introducing Sparse Principal Component Analysis (SPCA) into the recognition algorithm. The terms *dense* and *sparse* will be used to differentiate between standard Principal Component Analysis (PCA) and SPCA.

The structure of the paper is as follows. Section 2 describes the algorithm to construct the sparse similarity-based face representation. Section 3 presents results that show the discriminative power of using SPCA versus PCA in Sparse Similarity-based Fisherfaces and its ability to discover the individuals most discriminative characteristics. Section 4 gives a discussion and conclusion for Sparse Similarity-based Fisherfaces.

2 Algorithm Description

The proposed algorithm builds upon SPCA [10] and Fisher Linear Discriminant Analysis (FLDA) [3].

However, unlike traditional FLDA, which maximizes a global measure of class separation¹, it obtains a one-dimensional individual linear subspace for each person (class) enrolled in a training database. The value of the projection of each sample (face representation) projected into this subspace aims at measuring the similarity of the sample with respect to that person for whom the subspace was created. Following, the proposed algorithm is briefly described. For a better understanding, Fig. 1 displays a diagram of the algorithm.

2.1 Obtaining the Texture Formulation

Face Recognition can be conducted using several types of features such as geometrical and textural². In this work, visible texture features are used. The texture features are obtained by a piece-wise affine warp based on the Delaunay triangulation of the mean shape. Hereafter, the texture is normalized to zero mean and unit variance.

When the facial feature representations has been obtained, these are projected into a SPCA feature space to remove redundancy. There are different ways to formulate the SPCA objective function [9,10]. In this study the generalized power method for SPCA formulation is used as described in [10]

$$\phi_{\ell_0}(\gamma) = \max_{x \in S^P} \sum_{i=1}^n [(a_i^T x)^2 - \gamma]_+. \quad (1)$$

¹ In a subspace of dimensionality "Number of classes" minus one.

² Textural features can be recovered from any spectral range (E.g. visible or infrared).

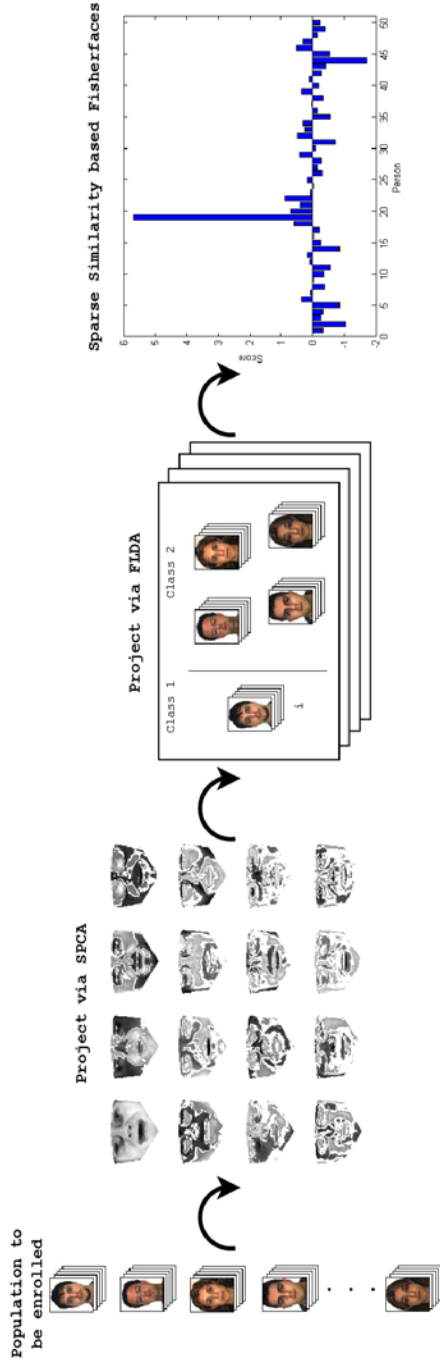


Fig. 1. Algorithm overview. SPCA and FLDA is used in turn on the population to be enrolled, in order to obtain the Sparse Similarity-based Fisherface representations.

Formulation (1) yields a single Sparse Principal Component (SPC). In order to obtain more than one SPC, (1) is used iteratively by means of deflation. Deflation is obtained as follows: when a new SPC is obtained its variation is removed from the populations variance before obtaining the next SPC. The advantage of this is that the resulting SPC number k will be nearly orthogonal to the $k - 1$ previous computed SPC.

The γ parameter controls the sparsity of the solution in such a way that the size of γ will move the SPCA solution between the two extremes: a full PCA and the zero solution. It depends on the application whether sparsity or explained variance is valued.

2.2 Creating the Individual Subspaces

After the face representations have been projected into the SPCA feature space, a second projection using FLDA for each individual in the database is conducted to build a personalized subspace for each of them. This subspaces are obtained discriminating each individual with respect to all others.

A standard FLDA projects the data samples into an F -dimensional subspace so that, it maximizes the ratio of the between-class scatter to the within-class scatter. The dimensionality F of this subspace is equal to the minimum of $n - 1$ and $m - 1$, where n is the number of variables and m is the total number of people (classes). The projection matrix W is found by maximizing the ratio

$$\frac{W^T S_B W}{W^T S_W W}, \quad (2)$$

where S_B and S_W are the between-class scatter and the within class scatter matrices, respectively. The projection vectors of this matrix W correspond to the eigenvectors associated to the non-zero eigenvalues of the matrix $S_W^{-1} S_B$.

In this case of two classes FLDA returns a one dimensional subspace for each individual. For a more formal discussion of FLDA and Individual Subspaces see [7].

2.3 Classification

To turn the obtained projections into measurements of similarity a standardization is applied. The standardization of model $i = 1, \dots, m$ is based on two assumptions. First, the number of observations for person i is much smaller than the number of the observations of all other people. Second, the projection of the other people follows a Gaussian distribution. These two assumptions imply that the distribution of all the projected facial images on a particular discriminative individual model is a Gaussian distribution with outliers. The standardization of model i is then achieved by transforming the projections into a standard Gaussian distribution, keeping the projections of the person i positive. Formally, let \bar{x}_i be the mean of the projections on model i , σ_i the standard deviation, and let $x_{i,j}$ be the projection of face representation j in the i th subspace. These projections are standardized by

$$\hat{x}_{i,j} = (x_{i,j} - \bar{x}_i) / \sigma_i. \quad (3)$$

If the standardized projection for the images corresponding to person i are negative, then $\hat{x}_{i,j}$ are replaced by $-\hat{x}_{i,j}$ for all projections. This causes the projection of the images corresponding to person i to be positive and far from the mean of the gaussian.

Once the model i is standardized, the probability of a projected image to belong to person i is given by the value of the standard normal cumulative function in the projected value. This fact is used to classify a given image. If it is assumed that the image belongs to a person from the data set, the image is projected by all the models and classified as belonging to the model that gives the largest probability. Moreover, it is also statistically possible to decide if a given person belongs to the data set or it is unknown. This can be achieved by comparing the largest projection obtained in all the models with a probabilistic threshold. For example if a 99.9% of probability is required, a given image will only be considered as belonging to the database if the projection in one of the individual models is higher than 3.1 standard deviations.

3 Experimental Results

In this article two experiments are presented. The first experiment aims at determining classification rates of sparse vs. dense versions of Eigenfaces, Fisherfaces and Similarity-based Fisherfaces, together with a visualization of Sparse Principal Components. The second experiment analyzes and visualizes which are the most discriminating pixels in a face image based on the algorithm for Sparse and Dense Similarity-based Fisherfaces.

3.1 Classification Accuracy

This experiment aims at comparing the performance of SPCA and PCA in the proposed method with respect to Fisherfaces and Eigenfaces method in terms of false classification rates. Both PCA and SPCA versions of Fisherfaces and Eigenfaces will be used in this experiment.

As data set for this study 50 persons (25 male and 25 female) was randomly selected from the AR face database [8]. The database is composed of two independent sessions recorded 14 days apart (Only images without occlusions is used). An example of the selected images for two persons is displayed in Fig. 2. All the images were manually annotated with 22 landmarks.

The data set was divided into two sets. The images of the first session were used to train the algorithms, whereas the images from the second session were subsequently used to test the performance. In order to obtain the texture representation of each face in the training set, the different images were warped with respect to the mean shape, represented by 41339 pixels. These representations was normalized to zero mean and unit variance.

In the Fisherface and the Eigenface algorithms the Nearest-Neighbor algorithm with Euclidean metric was used as classifier. In the proposed method the classification is performed in such a way; that a given face image is recognized as the person associated to the subspace that yields the highest probability.

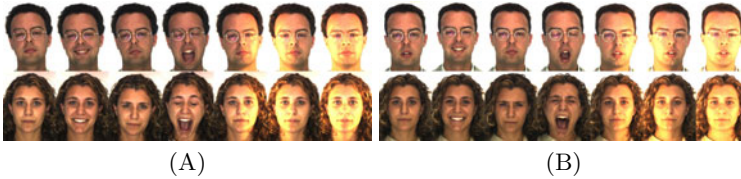


Fig. 2. The AR data set: (A) The seven images without occlusions from first session, (B) The seven images without occlusions from the second session.

The test was repeated a second time changing the roles of the training and the test sets: session two was used as training data and session one as test data.

Determine γ for Use in SPCA. Considering the formulation for SPCA (1) it can be determined that as γ goes to zero the SPCA goes to a full PCA solution. As described before, the process of choosing γ is not trivial and depends on the problem at hand, whether a more dense or sparse solution to SPCA is wanted. After performing exploratory experiments with different values of γ the value of $\gamma = 0.003$ was chosen as a good value for this problem. It is beyond the scope of this paper to analyze methods for determining an optimal γ for a given problem. The value of $\gamma = 0.003$ is used for the experiments in this study.

The false classification rates for the different techniques are shown in Fig. 3, where these rates are plotted as a function of using the first i PC/SPC. The figure shows rates up to the first one hundred PC/SPC.

To analyze the effect of the light variance, another test was conducted where the three pictures containing extreme light of each person in the test and training set was removed. This reduces the inner class variance, due to lighting noise. Any method for postprocessing could be used for removing light variation.

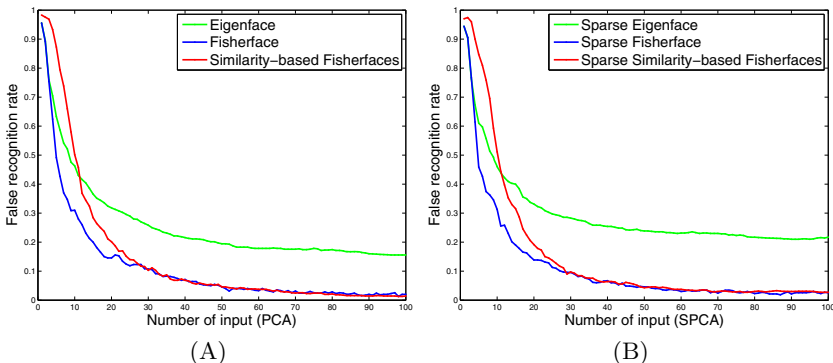


Fig. 3. False classification rates for Eigenfaces, Fisherfaces and Similarity-based Fisherfaces with extreme light variation. (A) displays the Dense versions, (B) displays the Sparse versions. The rates are plotted as a function of using the first i PC/SPC, $i = 1 \dots 100$.

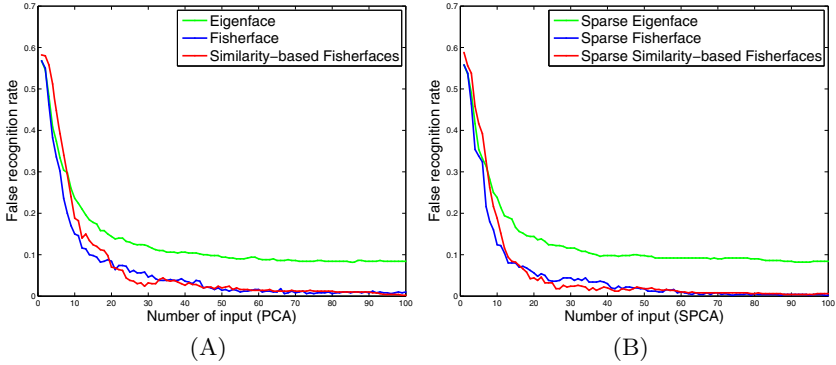


Fig. 4. False classification rates for Eigenfaces, Fisherfaces and Similarity-based Fisherfaces without extreme light variation. (A) displays the Dense versions, (B) displays the Sparse versions. The rates are plotted as a function of using the first i PC/SPC, $i = 1 \dots 100$.

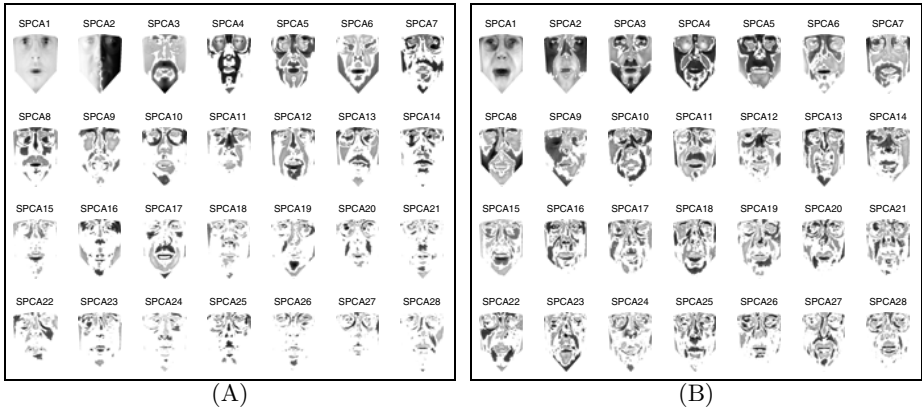


Fig. 5. Visualization of the first 28 Sparse Principal Components from the experiment (A) with extreme light variation (B) without extreme light variation.

However, for simplicity the images are simply removed in this study. The new false classification rates for the different techniques are shown in Fig. 4. The SPC can be seen in Fig. 5 for the two experiments with and without extreme light, respectively.

From Fig. 3 it can be seen that the Sparse versions of Fisherfaces and Similarity-based Fisherfaces obtain similar recognition results as the Dense versions with only using a fraction of the input data. In Fig. 4 it can be seen that Sparse Similarity-based Fisherfaces performs even better than the Dense version when large variance in the sample population due to inner class variance noise is minimized (removing light variation). The false classification rates for the first 25, 50 and 100 PC/SPC in Table 1 and Table 2 for the different techniques with and without extreme lighting, respectively.

Table 1. False classification rates for Eigenfaces, Fisherfaces and Similarity-based Fisherfaces with extreme light variation. The rates are shown for the first 25, 50 and 100 PC/SPC, respectively.

Method	Dense version			Sparse version		
	25 PC	50 PC	100 PC	25 SPC	50 SPC	100 SPC
Eigenfaces	29%	19,4%	15,6%	29,7%	23,9%	21,7%
Fisherfaces	11,9%	4,6%	2%	11,3%	4,6%	2,6%
Similarity-based Fisherfaces	14%	4,6%	1,4%	13,7%	4,3%	2,7%

Table 2. False classification rates for Eigenfaces, Fisherfaces and Similarity-based Fisherfaces without extreme light variation. The rates are shown for the first 25, 50 and 100 PC/SPC, respectively.

Method	Dense version			Sparse version		
	25 PC	50 PC	100 PC	25 SPC	50 SPC	100 SPC
Eigenfaces	13%	9,4%	8,4%	12,4%	9,8%	8,4%
Fisherfaces	5,8%	1,4%	1%	3,6%	1,8%	0,2%
Similarity-based Fisherfaces	3,8%	2%	0,2%	3,2%	1,8%	0,6%

3.2 Discriminative Pixels

An interesting property of the proposed algorithm is that it is possible to determine which are the most discriminative features of a given person. The 10, 15 and 25% discriminative pixels corresponding to the highest weights in the model are displayed (in red) in Fig. 6 for Sparse and Dense Similarity-based Fisherfaces respectively. It is clear that important discriminating features include eyes,

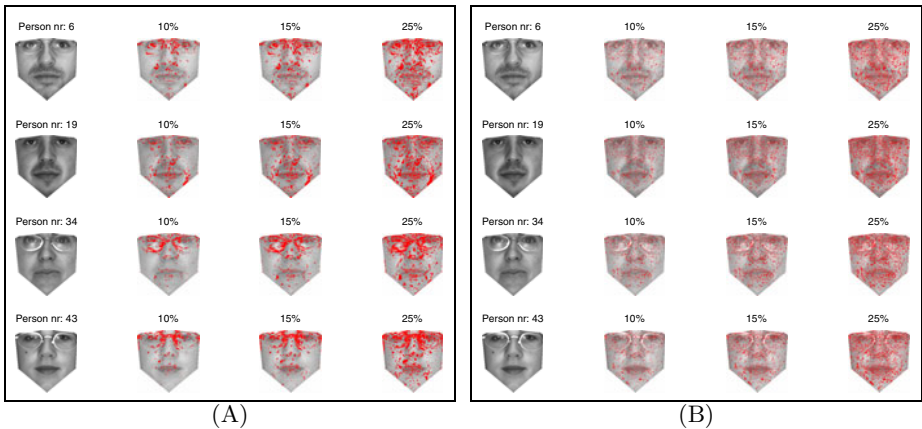


Fig. 6. The 10, 15 and 25% discriminative pixels displayed in red for (A) Similarity-based Fisher and (B) Sparse Similarity-based Fisherfaces

noses, glasses, moles and beards. Notice that the Dense algorithm detects the glasses and the mole of person 43 in Fig. 6 as discriminative features, whereas the Sparse algorithm does not do this to the same degree. This suggests that the Sparse method will be more robust giving the same result no matter if the person is wearing glasses in one picture and contact lenses in another.

4 Discussion and Conclusion

In this work, the effect of introducing Sparse Principal Component Analysis within the Similarity-based Fisherfaces algorithm are examined. Both the dense and sparse version of Similarity-based Fisherfaces aims at being a precise and robust algorithm that can be incorporated into biometrical security systems.

Experimental results in this study have shown that the technique in a sparse version can obtain the same recognition results as the technique in a dense version (presented in [7]) with only a fraction of the input data.

Furthermore, the presented results suggest that using SPCA in the technique offers robustness to occlusions. The discriminative pixels are not primarily fixed on naturally occurring occlusions in the face (Glasses, molds etc.), whereas for the dense version of the algorithm they are. This point needs to be examined in detail in future work.

Moreover, just as the dense version, the sparse version also allows for a simple interpretation of the results in the final one-dimensional individual subspace.

Another interesting property of the proposed algorithm is that by solving the problem in a series of subspaces, this algorithm makes the enrollment or removal of a person easy. It is not needed to recalculate all of the existing individual subspaces when changes to the population are made. When a new person is introduced, simply a new individual subspace is added. If a person has to be removed from the database, it is only needed to remove the corresponding individual subspace.

References

1. Kirby, M., Sirovich, L.: Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12, 103–108 (1990)
2. Turk, M., Pentland, A.: Eigenfaces for Recognition. *Journal of Cognitive Neuroscience* 3, 71–86 (1991)
3. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 711–720 (1997)
4. Cevikalp, H., Neamtu, M., Wilkes, M., Barkana, A.: Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 4–13 (2005)
5. Liu, Q., Huang, R., Lu, H., Ma, S.: Face recognition using Kernel-based Fisher discriminant analysis. In: 2002 Proceedings: Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 197–201 (2002)

6. Baddeley, A.D.: Essentials of Human Memory. Psychology Press, Taylor and Francis (1999)
7. Gomez, D.D., Fagertun, J., Ersbll, B., Sukno, F.M., Frangi, A.F.: Similarity-based Fisherfaces. *Pattern Recognition Letters* 30(12), 1110–1116 (2009)
8. Martinez, A.M., Benavente, R.: The AR Face Database. Technical Report, Computer Vision Center Purdue University (1998)
9. Zou, H., Hastie, T., Tibshirani, R.: Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics* 15, 265–286 (2006)
10. Journee, M., Nesterov, Y., Richtarik, P., Sepulchre, R.: Generalized Power Method for Sparse Principal Component Analysis. *The Journal of Machine Learning Research* 11, 517–553 (2010)