Recognition of Expression Variant Faces Using Weighted Subspaces

Yongbin Zhang and Aleix M. Martínez
Department of Electrical and Computer Engineering
The Ohio State University
{zhangyo, aleix}@ece.osu.edu

Abstract

In the past decade or so, subspace methods have been largely used in face recognition – generally with quite success. Subspace approaches, however, generally assume the training data represents the full spectrum of image variations. Unfortunately, in face recognition applications one usually has an under-represented training set. A known example is that possed by images bearing different expressions; i.e., where the facial expression in the training image and in the testing image diverge. If the goal is to recognize the identity of the person in the picture, facial expressions will be seen as distracters. Subspace methods do not address this problem successfully, because the feature-space learned is dependent over the set of training images available – leading to poor generalization results. In this communication, we show how one can use the deformation of the face (between the training and testing images) to solve the above defined problem. To achieve this, we calculate the facial deformation between the testing and each of the training images, project this result onto the (learned) subspace, and there weight each of the features (dimensions) inverse-proportionally to the estimated deformation. We show experimental results of our approach on those representations given by the following subspace techniques: Principal Components Analysis (PCA), Independent Components Analysis (ICA) and Linear Discriminant Analysis (LDA). We also present comparison results with a number of known techniques and show the superiority of our weighted LDA algorithm over the rest.

1 Introduction

In recent years, we have witnessed a significant success of appearance-based approaches especially when applied to face recognition problems [15, 16, 3, 7]. Appearance-based methods are attractive because the model of each class is directly defined by the selection of the sample images of that object, without the need to create precise geometrical or algebraic representations. The clear disadvantage of this definition is that any testing image with an appearance different to those used for training, will not generally be correctly classified – since it is known that images of different individuals obtained under similar imaging conditions are more alike than those images of the same person taken under different conditions [1]. In pattern recognition, it is generally argued that one should always use a minimum of samples equal to ten times the number of classes by the number of dimensions in the original feature space [9].[1] Unfortunately, it is rare the case when one has access to such a large number of training images per class in applications such as face recognition. And, even when one does have a large number of training images, these do not usually correspond to uncorrelated vectors.

The problem possed by subspace methods (as defined above) is illustrated in Fig. 1. In this example, we used one training images for each of the four classes to learn our low-dimensional image representation. All training images have a common facial expression: neutral. A testing image (with another facial expression) is latter projected onto this subspace (this image is marked with a green square in our Fig. 1). As the figure shows, simple Euclidean distances do not classify the test image into the correct class. In fact, it can be shown that any standard distance measure will generally fail in this task [13, 14].

The problem with subspace techniques is that the learned features (dimensions) do not only represent class discriminant information but are also tuned to those specific facial expressions of our training set. Our solution is to first learn which dimensions are *most* affected by this problem when comparing the testing image to each of the training images and, then, build a weighted-distance measure that gives less importance to those dimensions representing the facial expression changes and more importance to the rest. In our Fig. 1, we would need to assign a large weight to the second feature ($\mathbf{e}_2$) and a small weight to the first feature ($\mathbf{e}_1$).

In the rest of the paper we detail the formulation of our approach (Section 2) and compare it to a morphing alterna-

[1] Recall that in the appearance-based domain, the number of features in the original spaces equals the number of pixels, which makes this problem more challenging.

tive [5, 11, 17] (Section 3). Morphing and warping algorithms [5, 11, 10, 17] are used to warp training and testing samples to either neutral or shape-free image representations. However, morphing can fail due to occlusions (e.g., teeth and closed eyes), large deformations and textural changes due to the local deformation of the face. This last point is very important. Note that as the expression changes, so does the 3D position of several areas of the face which may vary the reflectance angle.This affect will obviously change the brightness of the image pixels (texture) in our image. Our approach addresses this problem by assigning low weights to those areas with large deformations and large weights to those areas that remain still.

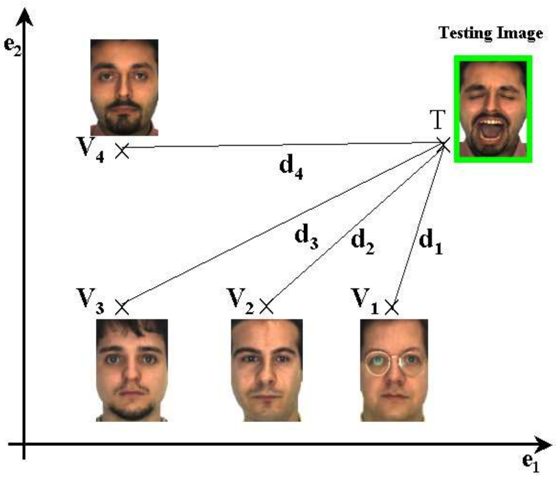


Figure 1. Depicted here is the problem possed by facial expression changes.

2 The Weighted Approach

We show derivations of our method on three popular subspaces, those given by Principal Components Analysis (PCA), Independent Components Analysis (ICA) and Linear Discriminant Analysis (LDA). We demonstrate the superiority of our weighted-LDA algorithm over the other two weighted subspaces and to a morphing approach.

2.1 Learning linear subspace representations

PCA, ICA and LDA are three of the most popular linear subspace methods and have been largely used in face recognition applications [15, 16, 3, 7, 2].

PCA projects the original space of d dimensions to that low-representation of p dimensions which minimizes the reconstruction error of the Gaussian distribution of the training data. To accomplish this, PCA uses the first and central moments of the data; i.e., the sample mean μ and the sample covariance matrix Σ_X.

While PCA only computes the first and central moments of the data, ICA uses higher moments of the data to find those feature vectors that are most independent from each other [2]. ICA does not have a general close-form solution but iterative methods are available. In our experimental results we will use the Infomax algorithm defined in [4].

While PCA and ICA are unsupervised learning algorithms, LDA is supervised. LDA selects those basis vectors that maximize the distance between the means of each class and minimize the distance between the samples in each class and their corresponding class means [8]. This can facilitate the task of feature extraction in some applications.

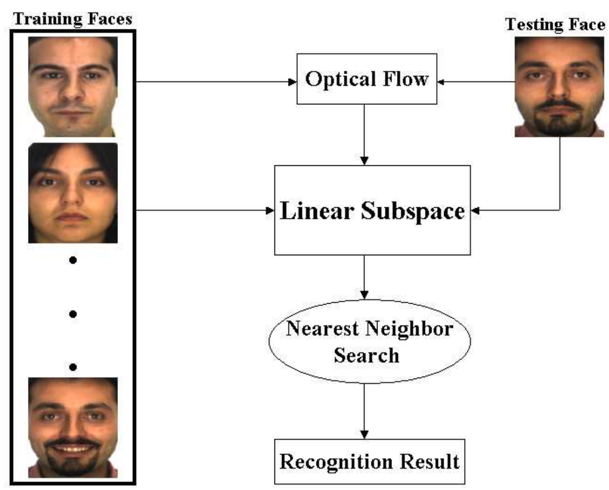


Figure 2. Schematic depiction of the proposed algorithm.

2.2 Weighted subspaces

We will refer to the projection matrix given by each of the methods described in the previous section as: Φ_{PCA}, Φ_{ICA} and Φ_{LDA}. These projection matrices are obtained with the help of a training set, $\mathbf{I} = \{\mathbf{I}_1, \ldots, \mathbf{I}_n\}$ (where n is the number of training images). Any new test image, $\mathbf{T}$, will then be classified as belonging to the class of the closest training sample given by the following weighted-distance equation:

$$\|\widehat{\mathbf{W}}_i \, (\hat{\mathbf{I}}_i - \hat{\mathbf{T}})\| \,, \tag{1}$$

where $\hat{\mathbf{I}}_i = \Phi * \mathbf{I}_i$ which is the i^{th} image projected onto the subspace of our choice of Φ ($\Phi = \{\Phi_{PCA}, \Phi_{ICA}, \Phi_{LDA}\}$), $\hat{\mathbf{T}} = \Phi * \mathbf{T}$, and $\widehat{\mathbf{W}}_i$ is the weighting matrix that defines the importance of each of the basis vectors in the subspace spanned by Φ.

Before one can use Eq. (1), we need to compute the weights of our matrix $\widehat{\mathbf{W}}$. While it may be very difficult to do that in the reduced space spanned by Φ, it is easy to calculate such values in the original space and then project these onto the subspace. For example, we can compute the value of the weights in the original space, $\mathbf{W}$, by means of the optical flow approach as follows (see Fig. 2). First we compute the optical flow between the testing image and each of the training samples,

$$\mathbf{F}_i = OpticalFlow(\mathbf{I}_i, \mathbf{T}). \tag{2}$$

Second, we calculate the weight in the original space. In this case we do that by assigning large weights to those areas with small deformations and low weights to those areas with large deformations. Formally,

$$\mathbf{W}_i = \mathbf{F}_{max} - \|\mathbf{F}_i\|, \tag{3}$$

where $\mathbf{F}_{max} = max_i \|\mathbf{F}_i\|$.

And, finally, we project this result onto our subspace,

$$\widehat{\mathbf{W}}_i = \Phi * \mathbf{W}_i. \tag{4}$$

A test image is classified by assigning to it the class label of the closest sample. Formally, find the closest sample,

$$s = argmin_i \|\widehat{\mathbf{W}}_i \left(\hat{\mathbf{I}}_i - \hat{\mathbf{T}}\right) \|, \tag{5}$$

and determine the class, c^*, of $\mathbf{I}_s$.

In our experimental results, described in the section to follow, we used Black and Anandan's optical flow algorithm [6], because it is robust to motion discontinuities such as the ones produced by facial muscles moving in opposite directions.

3 Experimental results

To test the proposed approach, we use the AR-face database [12]. From this dataset, we randomly selected a total of 100 people. Each individual consists of eight images with neutral, happy, angry and scream expressions taken during two sessions separated by two weeks time. These images for one of the people in the database are shown in Fig. 3.

Figure 3. Sample images from the AR database. (a-d) Neutral, happy, angry and scream expressions taken in the first session. (e-h) Same as (a-d) but from the second session.

All algorithms were tested using the leave-one-expression-out procedure. For example, when the happy face was used for testing, the neutral, angry and scream faces were used for training.

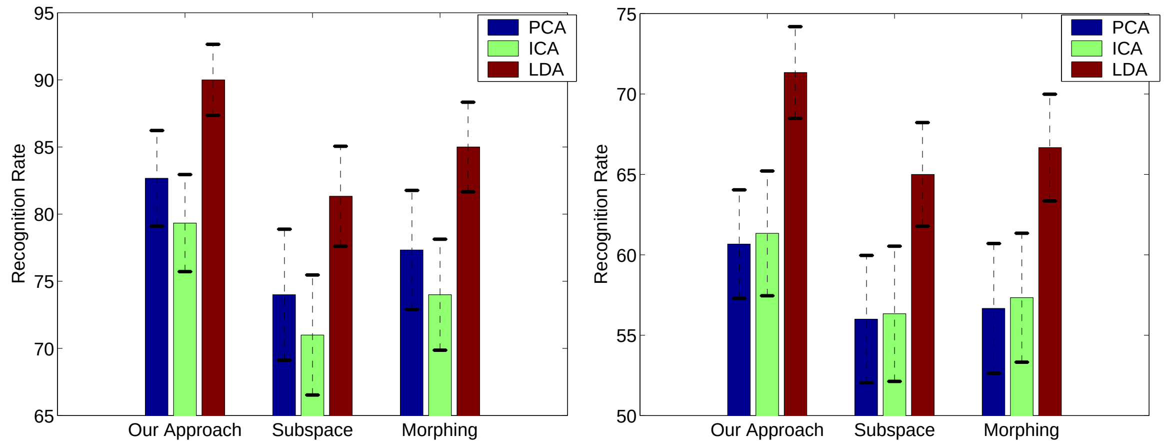
In our first experiment, only those images taken during the first session were used. The results obtained using the proposed weighted subspace approaches as well as those of PCA, ICA and LDA are shown in Fig. 4(a). In this figure we also show the results obtained by first morphing all faces to a neutral-expression face image and then building a PCA, ICA or LDA subspace. The bars in this figure show the average recognition rate for each of the methods. The standard deviation for the leave-one-expression-out test is shown by means of the small variance line at the top of each bar.

The second test was similar to the first one except that, this time, we used the images of the first session for training and those of the second session for testing. The results are summarized in Fig. 4(b).

It is worth mentioning that our weighted-LDA approach worked well for the scream face shown in Fig. 3(d) with a recognition rate of about 84%. Other methods could not do better than 70% in this particular case. In the figures shown above this is made clear by the small variance associated to our method as compared to the others.

4 Conclusion

In this paper, we have presented a weighed subspace approach to compensate for changes in expression. Three popular subspace methods (PCA, ICA and LDA) were explored, from which LDA has demonstrated the most promise. We compared our approach to PCA, ICA, LDA and a morphing technique and showed the superiority of the method proposed in this article. Smaller standard deviations in the results obtained by our method show that our algorithms are less sensitive to facial expression changes



(a) (b)

Figure 4. Recognition rates on the leave-one-expression-out test with: (a) images from the same session and (b) images from different sessions.

than others.

Acknowledgments

This research was financially supported in part by the National Institutes of Health.

References

[1] Y. Adini, Y. Moses and S. Ullman, "Face Recognition: The problem of compensating for changes in illumination direction," IEEE Trans. PAMI 19(7):721-732, 1997.

[2] M.S. Bartlett, "Face Image Analysis by Unsupervised Learning," Kluwer Academic,Dordrecht, 2001.

[3] P.N. Belhummeour, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs.Fisherfaces," IEEE Trans. PAMI 19(7):711-720,1997.

[4] A.J. Bell and T.J. Sejnowski,"An information-maximization approach to blind seperation and blind deconvolution," Neural Computation 7(1995) 1129-1159

[5] D. Beymer and T. Poggio,"Face Recognition from one Example View," ICCV 1995.

[6] M.J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," Comp Vis Image Under 63(1):75-104, 1996.

[7] K. Etemad and R. Chellapa, "Discriminant analysis for recognition of human face images," Journal of Optics of American A 14(8):1724-1733, 1997.

[8] R.A. Fisher, "The Statistical Utilization of Multiple Measurements," Annals of Eugenics,8:376-386,1938.

[9] K. Fukunaga, "Introduction to Statistical Pattern Recognition(second edition)," Academic Press, 1990.

[10] P.L. Hallinan, G.G. Gordon, A.L. Yuille, P. Giblin and D. Mumford, "Two- and Three-Dimensional Patterns of the Face," AK Peters, 1999.

[11] A. Lanitis, C.J. Taylor and T.F. Cootes, "Automatic Classification and Coding of Face Images Using Flexible Models", IEEE Trans.PAMI 19(7):743-756,1997.

[12] A.M. Martinez and R. Benavente, "The AR face database," CVC Tech.Rep. 24,June 1998.

[13] A.M. Martinez,"Recognizing Imprecisely Localized, Partially Occuded and Expression Variant Faces from a Single Sample per Class," IEEE Trans. PAMI 24(6):748-763.

[14] A.M. Martinez, "Matching Expression Variant Faces," Vision Research 43(9):1047-1060, 2003.

[15] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," Journal of the Optical Society of America A 4 (1987) 519–524

[16] M. Turk and A. Pentland, "Eigenfaces for recogntion," J. of Cog. Neuro. 3(1):71-86,1991.

[17] T. Vetter and T. Poggio,"Linear Object Class and Image Synthesis from single example view," IEEE Trans. PAMI 19(7):733-742, 1997.