

Smiling Faces are Better for Face Recognition

Yaser Yacoob and Larry Davis
Computer Vision Laboratory
University of Maryland
College Park, MD 20742

Abstract

This paper investigates face recognition during facial expressions. While face-expressions have been treated as an adverse factor in standard face recognition approaches, our research suggests that if a system has a choice in the selection of faces to use in training and recognition, its best performance would be obtained on faces displaying expressions. Naturally, smiling faces are the most prevalent (among expressive faces) for both training and recognition in dynamic scenarios. We employ a measure of Discrimination Power that is computed from between-class and within-class scatter matrices. Two databases are used to show the performance differences on different sets of faces.

1 Background

Face recognition has been widely studied within the research and development communities in the last decade. The objective has been to create a reliable and accurate biometric for human identification. Almost all research involved representing and recognizing faces in their neutral state. A study of the degradation of face recognition as faces deform during face expressions was reported in [9]. Feature and template-based approaches [6, 8] were evaluated with respect to recognition robustness as non-neutral expressions occurred in long image sequences. The authors found that recognition degrades as faces deformed. Also, it was observed that recognition parameters must be adjusted depending on the degree of neutrality of the face to ensure satisfactory performance as measured by an Receiver Operating Characteristics (ROC) curve.

The hypothesis in this paper is that faces are more recognizable if a system can observe them during face expressions. As faces deform they reveal more underlying information about the individual and as a result increase the identification ability

of the system. The revealed information reflects anatomic (bone, skin and muscle properties) as well as individual activation patterns.

We are unaware of any psychological studies to support the hypothesis that humans would perform better if training/recognition involved deforming faces. Nevertheless, we show in this paper that this hypothesis is well founded using a computational-statistical evaluation.

One challenging aspect of the face recognition problem is scaling up the performance so that tens of thousands and even millions of faces could be recognized. This challenge stems from the increased proximity of faces as the number of samples increases for a given representation. Ideally, one should seek a representation that maintains the separability of faces even as the number of faces increases. In this paper we show, on small data sets, that the scalability of the representation is generally better for expressive faces than it is for neutral faces.

2 Measures of Performance

Numerous algorithms for representing faces have been proposed (for a survey see [2]). A division of algorithms into feature and template-based approaches has been widely used. The performance of a face-recognition algorithm is clearly dependent on the representation chosen. In this paper we chose Principal Component Analysis (PCA) proposed by Turk and Pentland [8]. We omit the description of PCA due to its wide acceptance and use in the research community. The PCA representation imposes constraints on the recognition performance, but we focus on evaluating comparative performance.

We adopt the statistical measures used in [4] and discussed in [3, 5]. Assume that N face descriptors (e.g., a descriptor may be an intensity image, a vector of expansion coefficients, etc.) of L individuals are given for training ($N > L$). Each descriptor

is defined in an M space (e.g., M is equal to the number of pixels if an intensity image is used or the coefficient multipliers of the principal vectors of a PCA representation). Assume also that an equal number of samples of each face are given. In the case of PCA a set of expansion coefficients are assumed. Let V denote a representation that captures the N samples of faces in an R^M dimensional space. The objective is to define a measure of the discrimination power of V given the training set.

We consider all the descriptors of a particular individual’s face to be a class and the descriptors of all subjects to be in different classes. Therefore, we have L classes in V . The within and between-class scatter matrices are defined as follows

$$S_w^V = \sum_{i=1}^L P(C_i)\Sigma_i \quad (1)$$

$$S_b^V = \sum_{i=1}^L P(C_i)(\mu - \mu_i)(\mu - \mu_i)^T \quad (2)$$

where μ is the the overall mean of the vectors and S_w^V is the within-class scatter matrix showing the average scatter Σ_i of the sample descriptors of classes

$$\Sigma_i = E[(D - \mu_i)x(D - \mu_i)^T | C = C_i]. \quad (3)$$

where μ_i is the mean of class i , D are the sample vectors of different classes C_i and $P(C_i)$ is the probability of the i -th class. In the rest of the paper we assume that $P(C_i)$ is uniform. Quantifying the discrimination power of a representation can be done in several ways as described in [5]. In this paper we employ the separation matrix, which captures the relationship between the within and between class scatters of the descriptors in the representation space. Specifically, J^V

$$J^V = sep(V) = trace(S_w^{-1}S_b) \quad (4)$$

measures the discrimination power (DP) of a given representation V .

3 Experimental Objectives

We test our algorithms on two databases ([1, 7]). Both databases contain tens of subjects and therefore are relatively small. A more suitable database would have thousands of subjects with multiple instances of neutral and expressive faces, but unfortunately it is hard to create. The database in [1]

contains movie clips and therefore several instances of facial states are readily available. Notice that images taken from a sequence are very similar to each other and therefore the computation of S_w^{-1} can be computational unstable (small S_w values lead to large values in S_w^{-1}). On the other hand, the database provided by [7] has two samples of each facial states (some are taken in different imaging conditions) but these are taken days apart. Significant changes in appearance occurred for some subjects (hair cuts, facial hair, head orientation). As a result, the two samples used to constitute a class can be quite far apart. In some instances the two neutral images taken at different times are farther apart than the Smile and neutral faces taken at the same session. This makes the dataset very challenging and not perfectly suitable to evaluate the hypothesis of this paper, but nevertheless it is one of few that is available. To alleviate the within class distance variations described above, in the rest of the paper we will assume that the DP is computed as $trace(S_b)$ (which reflects the between-class scatter) whenever S_w involves a class that consists of one expression while if multiple expressions are involved in a class then computation proceeds based on Eq. 4.

Our experiments are designed to evaluate several issues associated with constructing a representation of faces for the purpose of recognition

- Assuming that samples of *all* possible faces of subjects are available to build the representation, what is the best choice to construct a representation for recognition?
- Assuming that samples of a *single facial state* of a person is available for all subjects, how do different representations of different facial states compare?
- Assuming that samples of only a few subjects are available, how good is each facial state representation when new faces are added?

In the rest of the paper we assume that the faces are normalized to a pre-set size (121x137 for [1] database and 128x160 for [7]). The images are also warped to align the eyes and mouth centers then cropped to capture the face area only. Figure 1 shows sample images of five subjects showing neutral and Smile expressions.

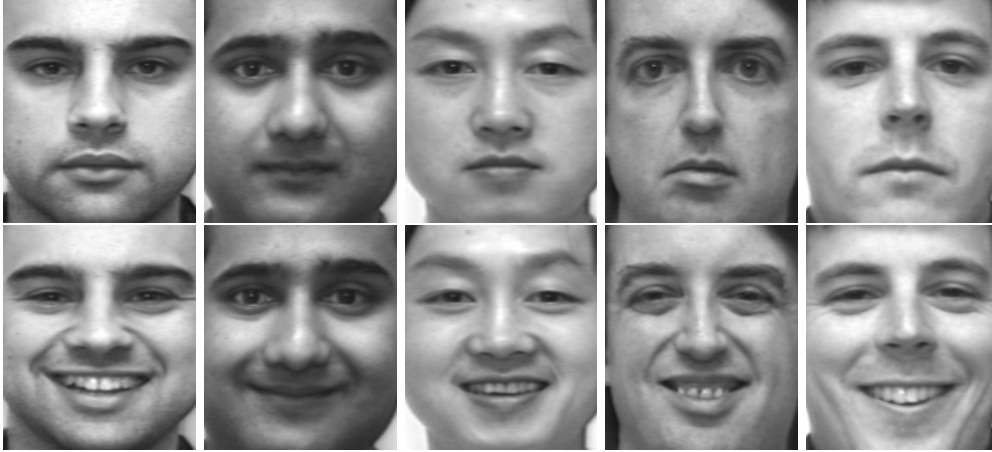


Figure 1: Sample images of 5 subjects during neutral and Smile (top row and bottom, respectively).

4 Discrimination Power of Representation of Facial States

In this section the discrimination power of a representation that includes both the neutral and expression images is evaluated. A dataset of images of 20 different individuals is used (taken from a video clip). For each individual 5 neutral and 5 peak expression images are selected in constructing a *joint* PCA representation. This joint representation allows us to compare the individual representations using the same bases. Since Smile and Anger turned out to be more easily conveyed by subjects in a laboratory set-up, we use these expressions in the rest of the paper. In the following experiment we assume that this common space is used to calculate the DP value.

Figure 2 shows the DP value as a function of the number of principal components. The graph (top) shows that the best discrimination power is achieved when a representation of face classes is constructed with Smile samples (here $DP = \text{trace}(S_b)$ to avoid computational instability in S_w). The bottom graph in Figure 2 shows simultaneous use of exemplars of neutral, Anger and Neutral expressions. The best DP representation is achieved by using the Neutral-Anger images as the basis for representation, followed by the Neutral-Smile representation (using Eq. 4) (here $DP = \text{trace}(S_w^{-1}S_b)$).

Next we construct an individual PCA representation for each set of Smile, Anger and neutral faces. Figure 3 shows the DP (here $DP = \text{trace}(S_b)$ to avoid computational instability in S_w) of several

representations:

- Neutral, Smile, Anger each computed with respect to its respective PCA.
- The DP of neutral projected onto the Smile and Anger PCA spaces.
- The DP of Smile and Anger projected onto the neutral PCA space.

Each DP is shown as a function of the number of eigenvectors used for the representation. The figure shows that neutral faces have generally a smaller DP in all spaces. In contrast, Smile and Anger faces have a large DP regardless of the space they are projected onto. However, DP of Smile and Anger calculated for projection onto their respective PCA spaces is better than projecting them onto neutral PCA space. The Figure shows that Anger and Smile faces have better DP values than Neutral and thus if these faces were available an improved performance should result.

Figures 4 and 5 show the results when the second dataset of 60 male-subjects [7] is used. The results are similar to the first dataset. Both Figures favor Smile over Neutral while Figure 5 also favors Smile over Neutral even if each was projected on the other space.

5 Degradation of performance with addition of subjects

In this section we evaluate the discrimination power of representations as new faces are added to the representation. A representation is superior if addition

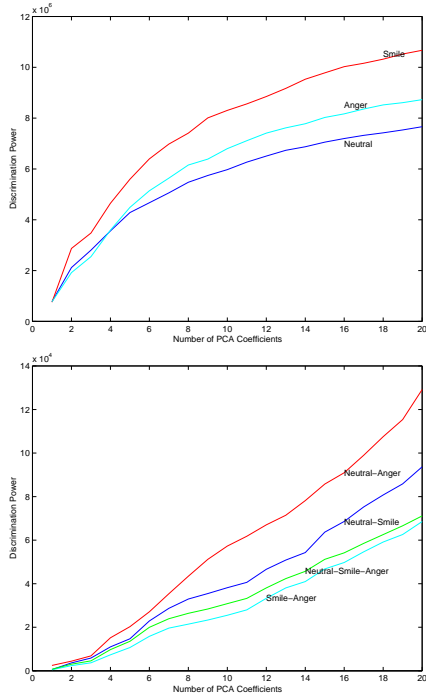


Figure 2: Discrimination power of representations computed from: neutral, Smile, Anger, joint Smile-neutral, neutral-Anger, and Smile-Anger

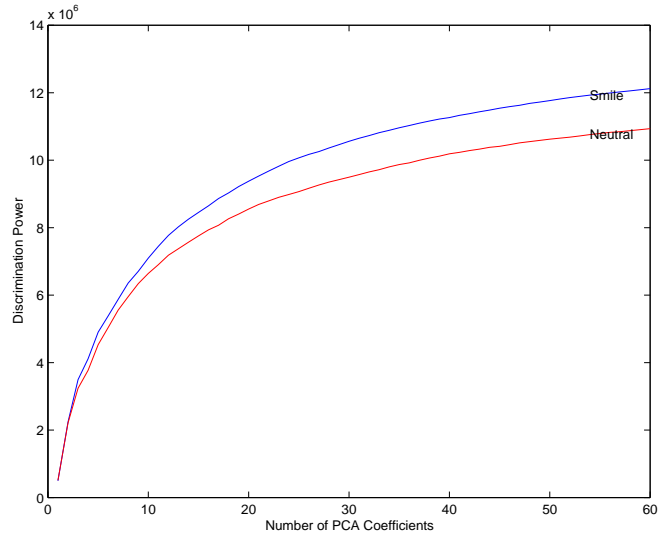


Figure 4: Discrimination power of representations computed from: neutral, Smile and joint Smile-neutral

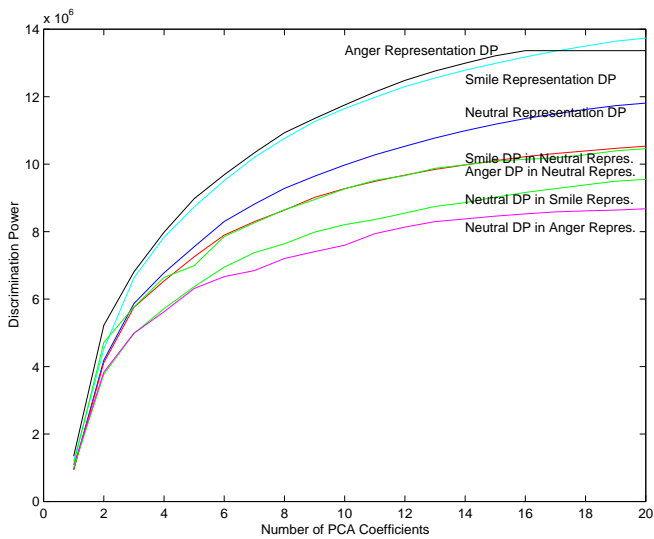


Figure 3: Discrimination power of representations computed from: neutral, Smile, Anger data in neutral space and neutral data in Smile and Anger space.

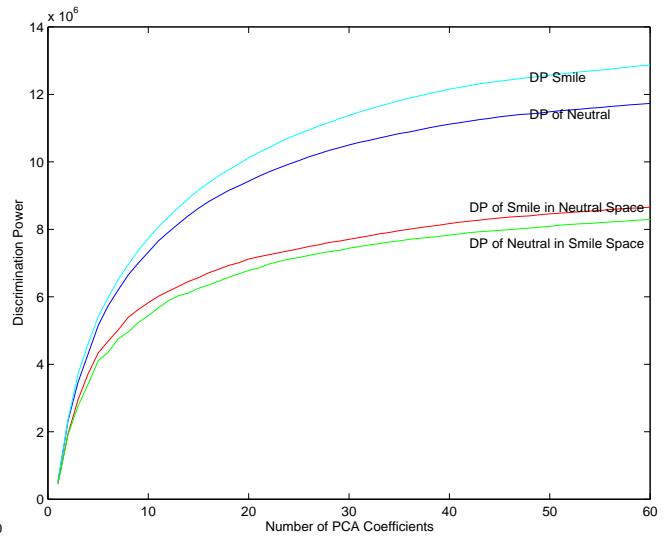


Figure 5: Discrimination power of representations computed from: neutral, Smile data in neutral space and neutral data in Smile space.

of faces maintains or perhaps increases the discrimination power. In fact, however, addition of new faces to an existing representation will degrade the discrimination power of the representation once a sufficiently large number of faces has been added. This is expected since the increase in the number of classes in a fixed dimensional representation leads to increased proximity between classes and a reduction in discrimination power.

We first consider the case of separate representations of neutral, Smile and Anger faces. For each representation we assume that 19 classes have been learned and that one subject is added to the representation. Figure 6 shows the DP as a function of the number of eigenvectors used for both the computed representation R^{19} of 19 neutral subjects and the DP for the representation for R^{19+1} which is the result of adding a new face without recomputing the PCA for the 20 faces. The DP for R^{19+1} is computed as the average of 20 DP experiments in which 19 faces are randomly selected for constructing a PCA and the 20th face inserted in the space. The DP increases with the increase of the number of eigenvectors for both representations. For a given number of eigenvectors, a small decrease in the DP occurs once a new face is added. Figures 7 and Figure 8 are the corresponding figures for Smile and Anger.

Figure 9 shows the average DP as a function of the number of added faces to a representation for neutral, Smile and Anger representations. The average was computed from 20 experiments in which 5 subjects were randomly selected for constructing the representation and the rest were inserted incrementally. Fifteen new faces are added in the neutral and Smile representation and 11 in the Anger representation. Overall, the Smile and Anger representations maintain a larger DP relative to the neutral, while it DP of Smile is slightly higher than Anger.

Figure 10 shows the average DP for 20 randomly selected faces in which another 40 faces are inserted (the second database is used). Here also, the average is for 20 repeated experiments. In this dataset we observe the same improved performance as seen in Figure 9.

6 Realism of expressions in verification and surveillance

Human recognition by biometric means has applications in access and surveillance applications. In

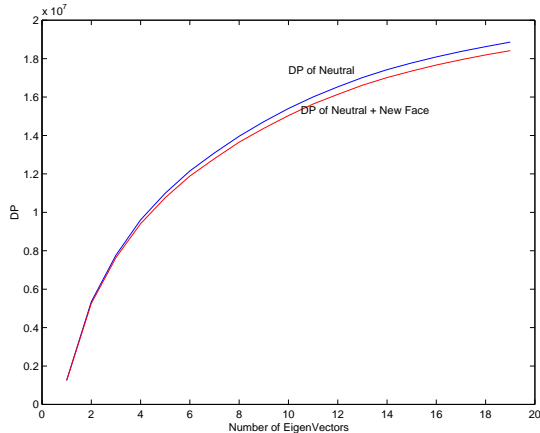


Figure 6: Discrimination power of neutral representation for 19 subjects and 19+one new subject as a function of the number of eigenvectors.

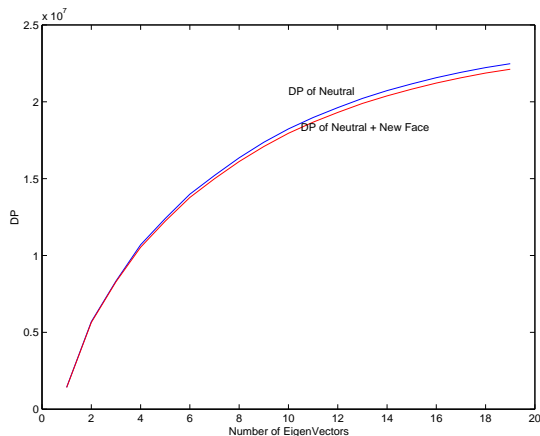


Figure 7: Discrimination power of SMILE representation for 19 subjects and 19+one new subject as a function of the number of eigenvectors.

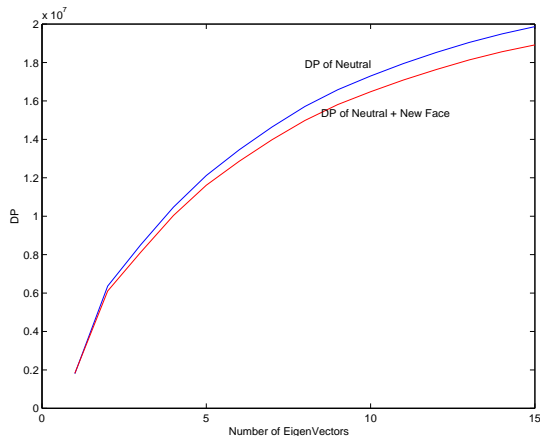


Figure 8: Discrimination power of representation for 15 subjects and 15+one new subject as a function of the number of eigenvectors.

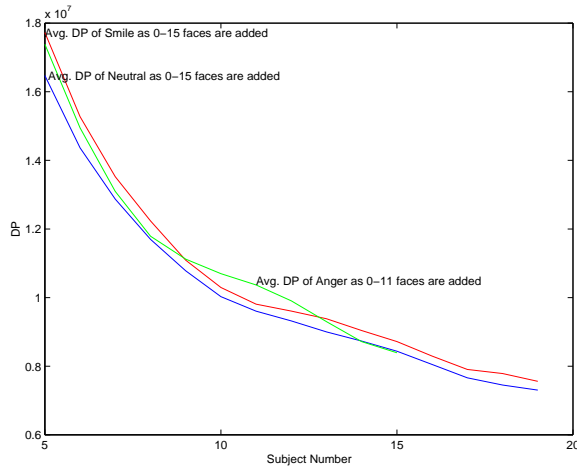


Figure 9: Discrimination power of representations computed from: neutral, Smile, Anger for newly inserted faces. While the representation is constructed using 5 faces, 15 new faces are added for neutral and Smile and 10 for Anger.

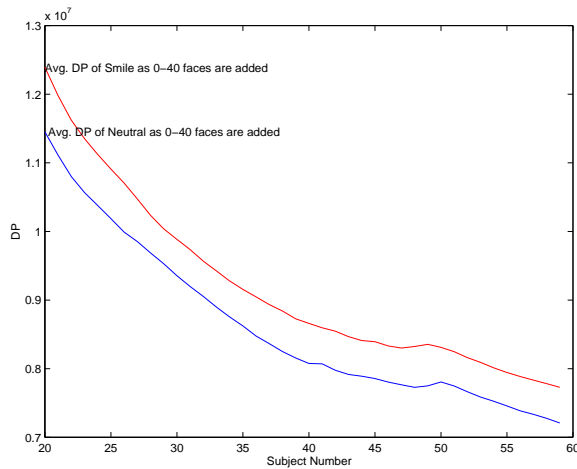


Figure 10: Discrimination power of representations computed from: neutral and Smile for newly inserted faces using the second database. While the representation is constructed using 20 faces, 40 new faces are added for neutral and Smile.

these situations the system has access to many images of the person. In access applications the user is cooperative and perhaps it is not unreasonable to ask the person to smile for the camera. On the other hand a surveillance system has to cope with uncooperative subjects and can only opportunistically use expressive images should these be viewed and detected. Face-expression recognition systems can be used for detection of the most “expressive” face in a video clip of a person.

Our research suggests that the more wildly expressive the face the better the derived representation. In our experiments Anger scored better than Smile. However, it is not generally realistic to assume that these expressive faces will commonly be encountered by the system.

We hope that this paper will stimulate research into “opportunistic face recognition” in which fleeting images that perhaps have higher information value are detected and used.

References

- [1] M. Black and Y. Yacoob. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motions. *IJCV*, 25(1), 1997, 23-48.
- [2] R. Chellappa, C.L. Wilson and S. Sirohey, “Human and Machine Recognition of Faces: A Survey” *Proceed. IEEE*, 83, 705-740, 1995.
- [3] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification* John Wiley and Sons, 2nd Edition, 2000.
- [4] K. Etemad and R. Chellappa, Discriminant analysis for recognition of human face images. *J. of Optical Society of America. A*, (14)8, 1724-1733, 1997.
- [5] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, New York, 1989.
- [6] M.S. Manjunath, R. Chellappa and C.v.d. Malsburg, A feature based approach to face recognition, *IEEE CVPR*, 373-378, 1992.
- [7] A.M. Martinez and R. Benavente, The AR Face Database. CVC Technical Report #24, 1998.
- [8] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, (3)1, 71-86, 1991.
- [9] Y. Yacoob, H. Lam and L.S. Davis, Recognizing Faces Showing Expressions, *International Workshop on Automatic Face- and Gesture-Recognition*, Zurich, 1995, 278-283.