Using an holistic method based on prior information to represent global and local variations on face images

Carlos E. Thomaz
Department of Electrical Engineering, FEI
São Bernardo do Campo, São Paulo, Brazil

Vagner do Amaral
Department of Electrical Engineering, FEI
São Bernardo do Campo, São Paulo, Brazil

Gilson A. Giraldi
National Laboratory for Scientific Computing
Petrópolis, Rio de Janeiro, Brazil

Duncan F. Gillies
Department of Computing, Imperial College London
London, UK

May 12, 2014

Abstract

Faces are familiar objects that can be easily perceived and recognized by ourselves. However, the computational modeling of such apparently natural human ability remains challenging. Recent studies in the literature have suggested that face processing is a cognition task composed of configural (or global) and featural (or local) sources of information, but with controversial arguments about the combination of these two types of information. In this work, we describe an holistic method that combines variance used in Principal Component Analysis (PCA) with some prior knowledge about the underlying visual perception task, including systematically the global and local information in the common multivariate representation of face images. We have showed that, with prior information, important local variations represented by principal components with small eigenvalues may not be discarded augmenting the classification accuracy of the first orthogonal basis vectors. Most interestingly, PCA with prior knowledge provides a specialized feature selection procedure, where the mapping of high-dimensional data into a lower-dimensional space has been able to handle local variations and capture not only the entire facial appearance but also some sample group facial features.

1 Introduction

Faces are one of the most familiar objects that can be perceived and recognized by humans [23, 15, 8, 24, 26]. However, the computational modeling of such apparently natural and heritable human ability [35, 17] remains challenging.

Recently, it has become a consensus, especially owing to biological and behavioral evidences [7, 15, 14, 1, 27], that face processing is a cognition task composed of configural and featural sources of information [4, 34, 16, 26], but with controversial arguments about the combination and relative interaction of these two types of information [26]. In other words, faces are expected to have not only a global and common spatial layout with all its parts such as eyes, nose and mouth arranged consistently [13], but also variations in these local features [18], which are fundamental to explain the singularity of each individual [25] or even samples of individuals with the same, for example, gender information or facial expression.

Several studies in the literature have suggested that this inherent relationship between the whole face and its constituent parts might rely on a holistic representation to appropriately describe the low-dimensional mechanism underlying our visual perception of faces [28, 2, 10, 19, 9, 13, 20, 6, 24]. In this context, Principal Component Analysis (PCA) [21, 12] has been the best known holistic representation used as a pre-processing step for automated face recognition systems [28, 33] as well as a conceptual framework for human face reasoning and coding [10, 19, 3, 5, 20, 14, 1]. Despite these well-known properties of PCA extensively applied in both computer vision and human perception communities, the issue of handling local variations differently in the common n-dimensional representation of face images with n pixels has not been addressed yet.

In this work, we describe and implement a priori-driven PCA that represents global and local variations on face images. In face recognition, with prior information, we show that important local variations represented by principal components with small eigenvalues may not be discarded augmenting the classification accuracy of the first orthogonal basis vectors. Analogously, PCA with prior knowledge is able to convey, using the same n-dimensional representation, the different visual cues that create, for instance, our distinct and selective perception for facial identity and expression [6] when reducing the dimensionality of the high-dimensional image inputs.

The paper is organized as follows. Next, in section 2, we translate in a systematic way the configural and featural sources of information for face processing on n-dimensional priori-driven principal components. Then, section 3 describes the face image datasets used to evaluate the effectiveness of the feature selection method proposed. All the analyzes of the experimental results carried out in this work have been explained in section 4. Finally, in section 5, we conclude the paper, discussing its main contribution and limitation.

2 Method

In this section, we describe the holistic method proposed that combines variance with prior information in order to include systematically the configural and featural sources of information in the n-dimensional representation of face images.

2.1 Combining variance with prior knowledge

Let an $N \times n$ data matrix X be composed of N face images with n variables (or pixels), that is, $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)^T$. This means that each column of matrix X represents the values of a particular variable observed all over the N signals. Let this data matrix X have covariance matrix

$$S = \frac{1}{(N-1)} \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T, \tag{1}$$

where $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T$ and $\bar{\mathbf{x}}$ is the grand mean vector of X given by

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n). \tag{2}$$

The well-known Pearson's sample correlation coefficient between the j^{th} and k^{th} variables is defined as follows [11]:

$$\begin{aligned} r_{jk} &= \frac{s_{jk}}{\sqrt{s_j}\sqrt{s_k}} \qquad (3) \\ &= \frac{\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2}\sqrt{\sum_{i=1}^{N}(x_{ik} - \bar{x}_k)^2}}, \end{aligned}$$

for $j = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, n$. Analogously to equation (3), we can describe the priori-driven sample covariance s^*_{jk} between the j^{th} and k^{th} variables by

$$\begin{aligned} s^*_{jk} &= (\sqrt{w_j}\sqrt{w_k}) s_{jk} \qquad (4) \\ &= \sum_{i=1}^{N} \sqrt{w_j}(x_{ij} - \bar{x}_j)\sqrt{w_k}(x_{ik} - \bar{x}_k). \end{aligned}$$

The spatial weighting vector (or spatial multivariate map)

$$\mathbf{w} = [w_1, w_2, \ldots, w_n]^T \tag{5}$$

is such that $w_j \geq 0$ and $\sum_{j=1}^{n} w_j = 1$, where each w_j measures the information power of the j^{th} variable. Thus, when n variables are observed on N samples, the weighted sample covariance matrix S^* can be described by

$$S^* = \left\{ s^*_{jk} \right\} = \left\{ \sum_{i=1}^{N} \sqrt{w_j}(x_{ij} - \bar{x}_j)\sqrt{w_k}(x_{ik} - \bar{x}_k) \right\}. \tag{6}$$

It is important to note that $s^*_{jk} = s^*_{kj}$ for all j and k and consequently the matrix S^* is a nxn symmetric matrix. Let S^* have respectively P^* and Λ^* eigenvector and eigenvalue matrices, as follows:

$$P^{*T} S^* P^* = \Lambda^*. \tag{7}$$

The set of k ($k \leq n$) eigenvectors of S^*, that is, $P^* = [\mathbf{p}^*_1, \mathbf{p}^*_2, \ldots, \mathbf{p}^*_k]$, which corresponds to the k largest eigenvalues, defines a new orthonormal coordinate system for the data matrix X called *priori-driven principal components* [31, 32].

2.2 The priori-driven weights

The issue here is how we can translate in a systematic way the configural and featural sources of information for face processing on a n-dimensional vector of priori-driven weights.

Since the holistic representation assumes that an input face image with n pixels can be treated as a point in an n-dimensional space (by concatenating the rows, or columns, of the image matrix), the layout of the whole face and its constituent parts like eyes, nose and mouth is already arranged by definition. As a consequence, any configural face variation related, for instance, to the spatial distances between facial parts would be intrinsically encoded in this holistic representation. The question is then how we can define the n-dimensional weighting vector to handle local variations and capture not only the entire facial appearance but also some individual or sample group facial features.

To do that we propose the idea of using a limited set of labeled samples and rearrange the data matrix X composed of N input images with n pixels on the following M ($M \leq N$) classification pairs: $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_M, y_M)$, where $\mathbf{x}_i \in \Re^n$ denote again the i^{th} face images and y_i are scalars that correspond to the human reasoning about the specific experimental task under investigation. For example, in a low-dimensional representation about gender facial differences, which involves two sample groups only (male versus female), $y_i \in \{-1, 1\}$.

The easiest multivariate linear method to calculate a spatial multivariate map $\mathbf{w}$ that discriminates labeled sample groups [30] is via computation of the spectral decomposition of the between-scatter matrix S_b given by:

$$S_b = \sum_{i=1}^{g} N_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T, \quad (8)$$

where N_i is the number of images from class i, $\bar{\mathbf{x}}_i$ is the unbiased sample mean of class i [11] and g is the total number of classes or groups. The spatial multivariate map $\mathbf{w}$ is simply the leading eigenvector of S_b, called here as the *1st order prior information*.

2.3 The Step-by-Step Algorithm

The main steps for calculating the priori-driven principal components $P^* = [\mathbf{p}_1^*, \mathbf{p}_2^*, \ldots, \mathbf{p}_k^*]$ of an $N \times n$ training set matrix X composed of N input images with n pixels can be summarized as follows:

1. Calculate the vector $\mathbf{w} = [w_1, w_2, \ldots, w_n]^T$ using the M classification pairs, where $M \leq N$, and the between-scatter matrix, as described in the previous sub-section;

2. Normalize $\mathbf{w}$ such that $w_j \geq 0$ and $\sum_{j=1}^{n} w_j = 1$, that is, replace w_j with $\frac{|w_j|}{\sum_{j=1}^{n} |w_j|}$;

3. Standardize all the n variables of the data matrix X such that the new variables have $\bar{x}_j = 0$, for $j = 1, 2, \ldots, n$. In other words, calculate the grand mean vector

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$$

   and replace x_{ij} with z_{ij} given by

$$z_{ij} = x_{ij} - \bar{x}_j$$

   for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, n$;

4. Spatially weigh up all the standardized z_{ij} variables using the normalized weighting vector $\mathbf{w}$ calculated in step 2, that is
$$z_{ij}^* = z_{ij}\sqrt{w_j};$$
5. The priori-driven principal components P^* are then the eigenvectors corresponding to the k largest eigenvalues of $(Z^*)^T Z^*$, where $Z^* = \{\mathbf{z}_1^*, \mathbf{z}_2^*, \ldots, \mathbf{z}_N^*\}^T$.

3 Experiments

The following experimental tasks have been performed using frontal and pre-aligned face images cropped to 128x128 pixels in size: (a) Gender experiments (female versus male samples); (b) Smiling experiments (neutral versus smiling samples). We chose to contrast these two experiments because the featural sources of information related to gender and smiling are both sparse, but distributed differently on face images.

Two publicly available data sets have been used to evaluate the low-dimension representation and interpretability of the priori-driven principal components: FEI [29] and FERET [22]. The FEI data set is composed of 200 subjects (100 men and 100 women). Each subject has two frontal images (one with a neutral or non-smiling expression and the other with a smiling facial expression). In total 400 images were used to perform the gender and expression experiments. In the FERET database, we have used 200 subjects (107 men and 93 women). Each subject has two frontal images (one with a neutral or non-smiling expression and the other with a smiling facial expression), also providing a total of 400 images to perform the gender and expression experiments.

4 Results

4.1 Prior information

The multivariate weighting maps that represent the difference between the sample group means are illustrated in Figure 1 for the gender and smiling variation tasks in both datasets. We can see clearly that the information varies depending essentially on the experimental task, consistent with the theoretical concept of a priori driven model. As expected, given the well-framed and pre-aligned face images, the spatial multivariate map of the gender experiments (right panel) is sparse and less localized whereas in the smiling task (left) the weights extracted by the 1st order prior information show group-differences present mainly on the mouth and areas nearby.

4.2 Dimensionality reduction

Figure 2 shows the total variance of the smiling (left panel) and gender (right) experiments using different number of principal components selected by the corresponding largest eigenvalues. It is possible to see that the combination of the variance criterion used in standard PCA with the prior knowledge about the experimental task has provided a more parsimonious feature extraction procedure, where the mapping of high-dimensional data into a lower-dimensional space used less features. For instance, to explain in both experiments 90% of the total variance, note that less priori-driven components would be necessary than standard ones in all experiments carried out.

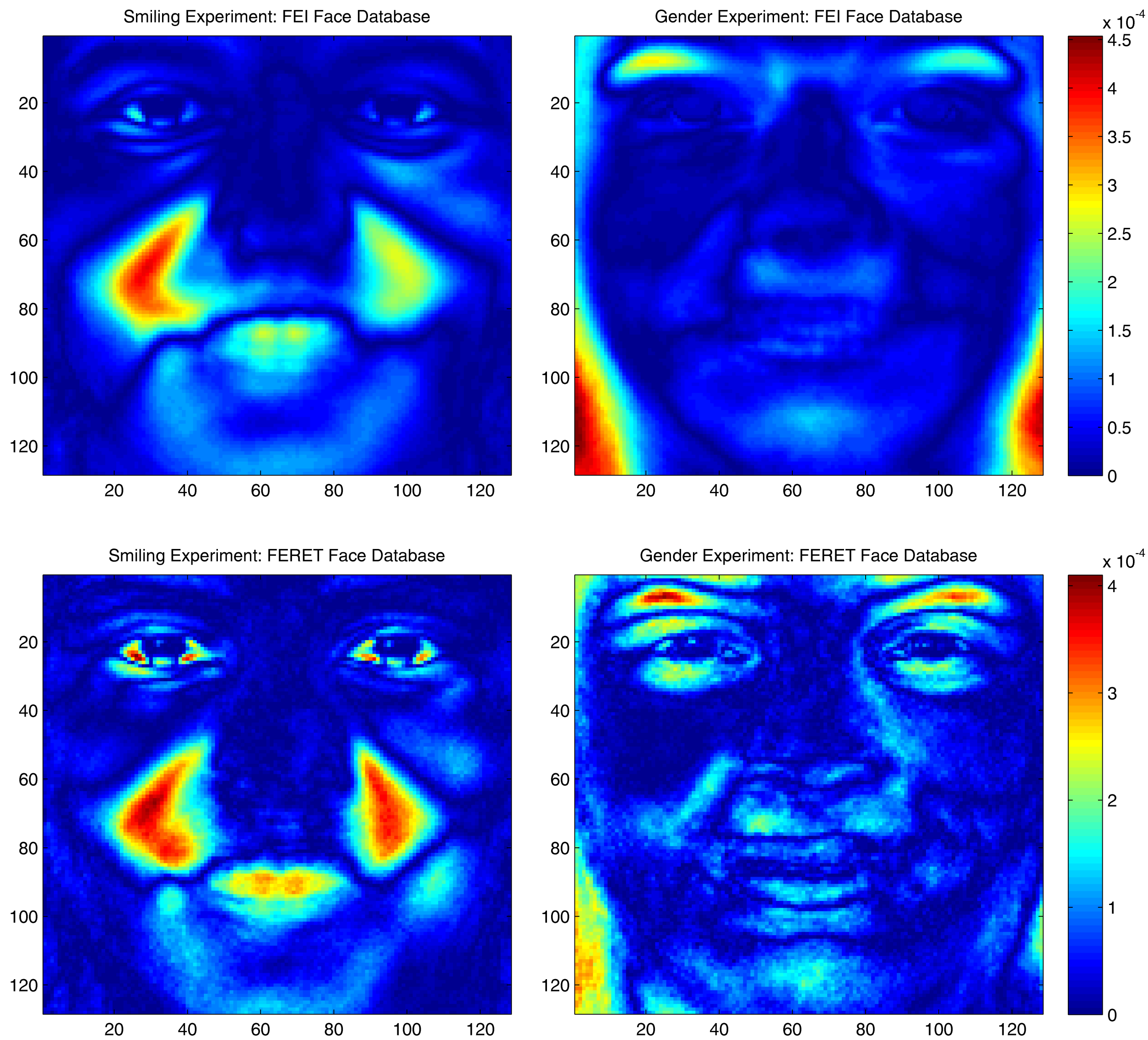

Figure 1: From top to bottom: spatial multivariate maps that represent sample mean differences extracted from the smiling (left) and gender (right) sample groups. Regions contained within the colored areas and closer to the spectrum of red represent pixels of relatively larger weights (in absolute values).

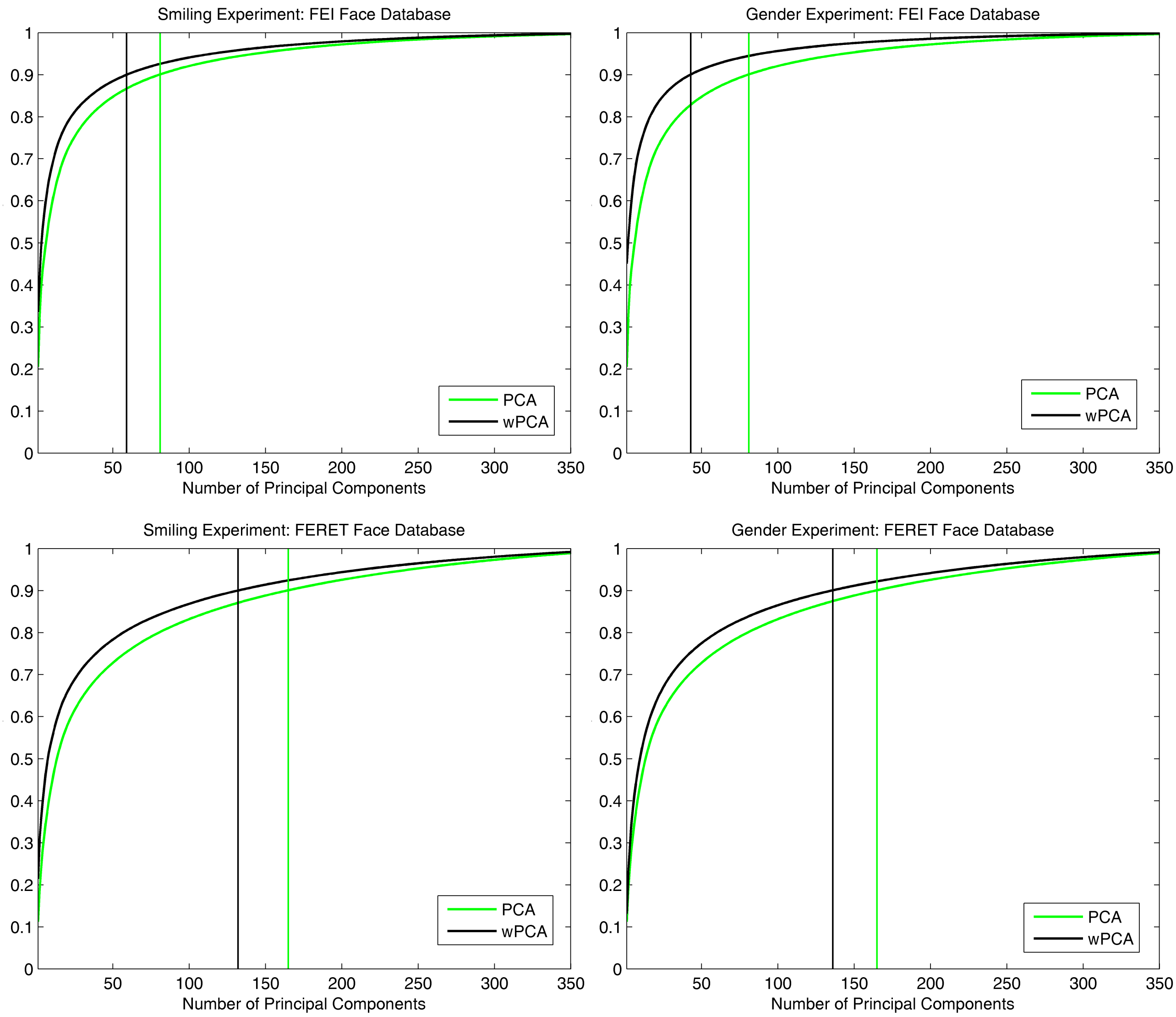

Figure 2: Total variance of the priori-driven PCA (wPCA) compared to the standard PCA for the smiling (left) and gender (right) experiments. The vertical lines denote the number of priori-driven and standard principal components necessary to explain 90% of the corresponding total variance information in each experiment.

Additionally, Figure 3 illustrates the inner product matrices of the first 40 standard (left) and priori-driven (right) principal components of the gender and smiling experiments using the FEI and FERET datasets. It is noteworthy that in contrast to the standard principal components that are identical (apart from the sign) in both datasets, there are clear changes in the information retained and the ordering of the priori-driven components, describing distinct patterns depending on the experimental tasks when reducing the dimensionality of the high-dimensional face image inputs.

4.3 Classification accuracy

We adopted a 10-fold cross validation method drawn at random from the gender and smiling corresponding sample groups to evaluate the classification accuracy of the priori-driven approach in comparison with the standard PCA. Throughout all the results, the Euclidean distance to the nearest neighbour sample has been used to assign a test image to either the male or female groups in the gender experiment, or to either the smiling or non-smiling group in the expression experiment. Figure 4 shows the average recognition rate using different number of principal components selected by the corresponding largest eigenvalues. It is clear that in both experiments and datasets the use of prior information improved the discriminant power of the principal components, especially in the first ones, allowing similar or higher average recognition rates with the same number of components. Most interestingly and, in fact, more importantly regarding an holistic representation to describe the low-dimensional mechanism underlying our visual perception of faces, the combination of the variance criterion used in standard PCA with the prior knowledge about the experimental task has provided a specialized and constrained feature selection procedure, where the mapping of high-dimensional data into a lower-dimensional space contained most of the relevant local information and used less features. For instance, note that in all results not only less priori-driven components would be necessary than standard ones to explain 90% of the total variance (vertical lines), but also such subsets of priori-driven components seem to be more appropriate to make judgments about the experiments because show higher classification accuracy.

5 Conclusion

In this paper we have described an holistic method that combines variance with prior information in order to include systematically global and local sources of information in the commonly used high-dimensional representation of frontal and pre-aligned face images. The advantage of the approach proposed is that it can be performed on the features of interest, generating simpler and easier low-dimensional representation of face images that allow subsequent retrieving or classification using fewer principal components. Its disadvantage is the dependence on some labeled samples pre-defined to handle the local variations and capture not only the entire facial appearance but also some sample group facial features. The rationale of this priori-driven multivariate approach is akin to the idea of investigating the link between low-level (or local) visual attributes, such as color, shape and texture, and high level (or global) ones, explained by semantic concepts of human reasoning, to extract and interpret the most informative features in face image analysis given by the data available.

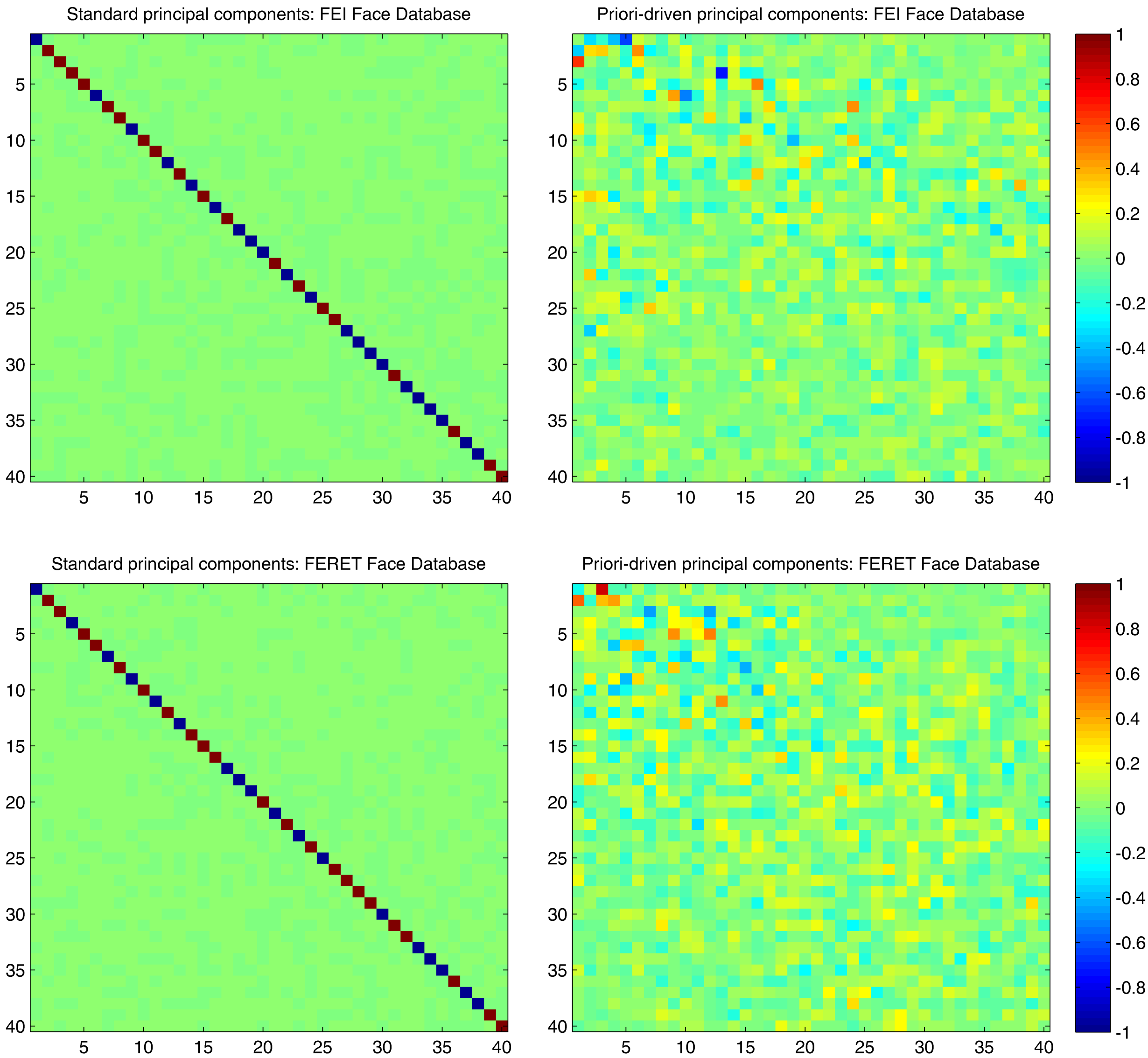

Figure 3: Inner product matrices of the first 40 standard (left) and priori-driven (right) principal components of the smiling and gender task experiments using the FEI (top) and FERET (bottom) datasets.

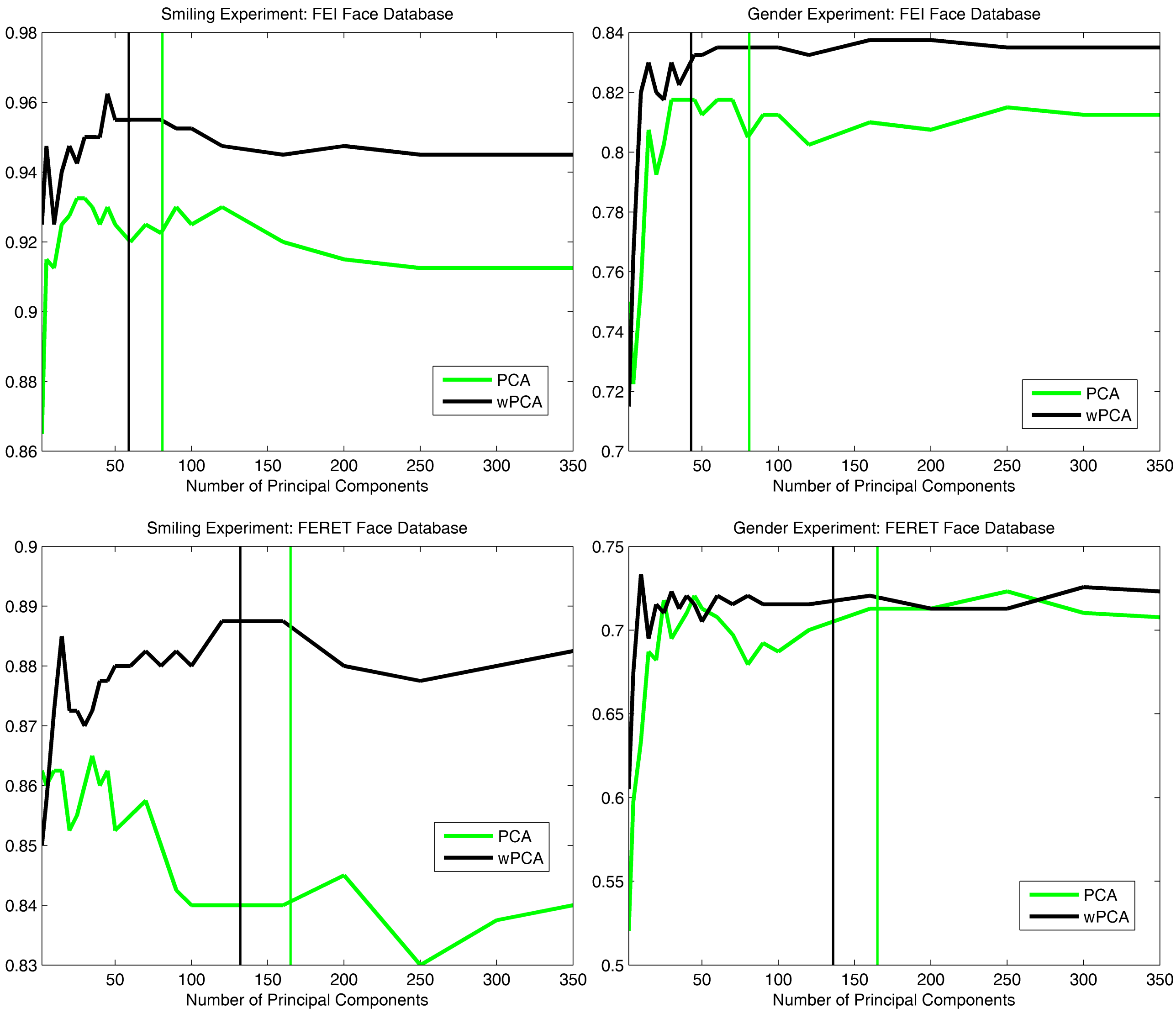

Figure 4: Average recognition rate of the priori-driven PCA (wPCA) compared to the standard PCA for the smiling (left) and gender (right) experiments. All the principal components retained have been selected by their corresponding largest eigenvalues. The vertical lines denote the number of priori-driven and standard principal components necessary to explain 90% of the corresponding total variance information in each experiment.

References

[1] L. Barrington, T. K. Marks, J. H. Hsiao, and G. W. Cottrell. Nimble: A kernel density model of saccade-based visual memory. *Journal of Vision*, 8(14):1–14, 2008.

[2] J. C. Bartlett and J. Searcy. Inversion and configuration of faces. *Cognitive Psychology*, 25:281–316, 1993.

[3] A. M. Burton, V. Bruce, and P. J. B. Hancock. From pixels to people: a model of familiar face recognition. *Cognitive Science*, 23(1):1–31, 1999.

[4] R. Cabeza and T. Kato. Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science*, 11(5):429–433, 2000.

[5] A. J. Calder, A. M. Burton, P. Miller, A. W. Young, and S. Akamatsu. A principal component analysis of facial expressions. *Vision Research*, 41:1179–1208, 2001.

[6] Andrew J. Calder and Andrew W. Young. Understanding the recognition of facial identity and facial expression. *Nature Reviews: Neuroscience*, 6:641–651, 2005.

[7] S. M. Collishaw and G. J. Hole. Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, 29(8):893–910, 2000.

[8] K. Crookes and E. McKone. Early mature of face recognition: no childhood development of holistic processing, novel face encoding, or face-space. *Cognition*, 111:219–247, 2009.

[9] S. Edelman and N. Intrator. Learning as extraction of low-dimensional representations. *Mechanisms of perceptual learning*, 36:353–380, 1997.

[10] Peter J. B. Hancock, A. Mike Burton, and Vicki Bruce. Face processing: Human perception and principal components analysis. *Memory and Cognition*, 24:26–40, 1996.

[11] R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall, 1998.

[12] I.T. Jolliffe. Principal component analysis. *Springer Series in Statistics*, 2002.

[13] R. M. Joseph and J. Tanaka. Holistic and part-based face recognition in children with autism. *Journal of Child Psychology and Psychiatry*, 44(4):529–542, 2003.

[14] J. Lacroix, J. Murre, E. Postma, and H. Herik. The natural input memory model. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, 2004.

[15] D. Maurer, R. L. Grand, and C. J. Mondloch. The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6):255–260, 2002.

[16] D. Maurer, K. M. OCraven, R. Grand, C. J. Mondloch, M. V. Springer, T. L. Lewis, and C. L. Grady. Neural correlates of processing facial identity based on features versus their spacing. *Neuropsychologia*, 45(7):1438–1451, 2007.

[17] E. McKone and R. Palermo. A strong role for nature in face recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11):4795–4796, 2010.

[18] S. Miellet, R. Caldara, and P. G. Schyns. Local jekyll and global hyde: The dual identity of face recognition. *Psychological Science*, 22(12):1518–1526, 2011.

[19] Alice J. O'Toole, Kenneth A. Deffenbacher, Dominique Valentin, Karen Mckee, David Huff, and Herv Abdi. The perception of face gender: The role of stimulus structure in recognition and classification. *Memory and Cognition*, 26:146–160, 1997.

[20] T. J. Palmeri and I. Gauthier. Visual object understanding. *Nature Reviews Neuroscience*, 5:291–303, 2004.

[21] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.

[22] P. J. Philips, H. Wechsler, J. Huang, and P. Rauss. The feret database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.

[23] A. Puce, T. Allison, M. Asgari, J. C. Gore, and G. McCarthy. Differential sensitivity of human visual cortex to faces, letterstrings and textures: a functional magnetic resonance imaging study. *The Journal of Neuroscience*, 16(16):5205–5215, 1996.

[24] J. J. Richler, O. S. Cheung, and I. Gauthier. Holistic processing predicts face recognition. *Psychological Science*, 22(4):464–471, 2011.

[25] Jeongin Seo and Hyeyoung Park. Robust recognition of face with partial variations using local features and statistical learning. *Neurocomputing*, 129:41–48, 2014.

[26] N. Y. Shin, J. H. Jang, and J. S. Kwon. Face recognition in human: the roles of featural and configurational processing. *Face Analysis, Modelling and Recognition Systems*, 9:133–148, 2011.

[27] Y. W. Shin, M. H. Na, T. H. Ha, D. H. Kang, S. Y. Yoo, and J. S. Kwon. Dysfunction in configural face processing in patients with schizophrenia. *Schizophrenia Bulletin*, 34(3):538–543, 2008.

[28] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of Optical Society of America*, 4:519–524, 1987.

[29] C. E. Thomaz and G. A. Giraldi. A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing*, 28(6):902–913, 2010.

[30] C. E. Thomaz, G. A. Giraldi, J. F. P. Costa, and D. F. Gillies. A simple and efficient supervised method for spatially weighted pca in face image analysis. Technical Report 13, Imperial College London, Department of Computing, www3.imperial.ac.uk/computing/research/techreports, 2010.

[31] C. E. Thomaz, G. A. Giraldi, J. P. Costa, and D. F. Gillies. A priori-driven pca. In *ACCV 2012 Workshops (LNCS 7729)*, pages 236–247, 2013.

[32] C. E. Thomaz, E. L. Hall, G. A. Giraldi, P. G. Morris, R. Bowtell, and M. J. Brookes. A priori-driven multivariate statistical approach to reduce dimensionality of meg signals. *Electronics Letters*, 49(18):1123–1124, 2013.

[33] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.

[34] C. Wallraven, A. Schwaninger, and H. H. Bulthoff. Learning from humans: Computational modeling of face recognition. *Network: Computation in Neural Systems*, 16(4):401–418, 2005.

[35] J. B. Wilmer, L. Germine, C. F. Chabris, G. Chatterjee, M. Williams, E. Loken, K. Nakayama, and B. Duchaine. Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11):5238–5241, 2010.