A Priori-Driven PCA

Carlos Thomaz[1], Gilson Giraldi[2], Joaquim Costa[3], and Duncan Gillies[4]

[1] Department of Electrical Engineering, FEI, Sao Paulo, Brazil
[2] National Laboratory for Scientific Computing, Rio de Janeiro, Brazil
[3] Department of Applied Mathematics, University of Porto, Portugal
[4] Department of Computing, Imperial College London, UK

Abstract. Principal Component Analysis (PCA) is a multivariate statistical dimensionality reduction method that has been applied successfully in many pattern recognition problems. In the research area of analysis of faces particularly, PCA has been used not only as a pre-processing step to produce accurate analytical model for automated face recognition systems, but also as a conceptual framework for human face coding. Despite the well-known attractive properties of PCA, the traditional approach does not incorporate high level semantics from human reasoning which may steer its subspace computation. In this paper, we propose a method that allows PCA to incorporate such semantics explicitly. It allows an automatic selective treatment of the variables that compose the patterns of interest, performing data feature extraction and dimensionality reduction whenever some high level information in the form of labeled data are available. The method relies on spatial weights calculated, in this work, by separating hyperplanes. Several experiments using 2D frontal face images and different data sets have been carried out to illustrate the usefulness of the method for dimensionality reduction, interpretation, classification and reconstruction of face images.

1 Introduction

Principal Component Analysis (PCA) [1, 2] is the best known multivariate statistical linear method for dimensionality reduction and has been applied successfully in many pattern recognition problems to reduce the computational costs, mitigate the curse of dimensionality and improve the classification performance.

In the research area of analysis of faces particularly, PCA has been used not only as a pre-processing step to produce accurate and computational efficient model for automated face recognition systems [3, 4], but also as a conceptual framework for human face reasoning and coding [5–8]. However, despite the well-known attractive properties of PCA in both computer vision and human perception communities, incorporating prior information in its process remains a challenge. In face recognition, without prior information important content-based features represented by principal components with small eigenvalues may be discarded reducing the accuracy of the automated representation. Analogously, PCA with no prior knowledge is a non-supervised algorithm unable to

J.-I. Park and J. Kim (Eds.): ACCV 2012 Workshops, Part II, LNCS 7729, pp. 236–247, 2013.

convey the different visual cues that create, for instance, a separate human perception coding for facial identity and expression [9].

The development of techniques that bring together dimensionality reduction and prior knowledge can be performed in the framework of supervised learning approaches. However, when only a small number of labeled samples are available, multivariate supervised dimensionality reduction methods tend to perform poorly due to overfitting [10]. Such problem has been addressed recently by a number of works related to semi-supervised dimensionality reduction methods [11–15]. A common issue to all semi-supervised learning techniques is how to optimize the regularization parameters necessary to blend supervised and non-supervised information often represented by local and global scatter matrices.

In this paper, we address this issue through separating hyperplanes. We propose a spatially weighted form of PCA that incorporates domain knowledge and generates an embedding space that preserves the optimality properties of dimensionality reduction and interpretability of the standard PCA. Unlike other similar projection approaches that either formulate their solutions in conjunction with parametric [16–18] or non-parametric [19] models or are restricted to the number of groups of patterns available [20], in our method a separating hyperplane, based on a discriminant criterion, is computed and discriminant weights are determined and used to generate the spatially weighted PCA subspace. In this sense, weights incorporate separately the prior knowledge extracted from the labeled data and can be systematically computed through any hyperplane direction. The approach is a simple way of allowing an automatic selective treatment of the variables that compose the patterns of interest, performing data feature extraction and dimensionality reduction whenever some high level semantics in the form of spatial weights are available.

2 A Priori-Driven PCA

Let an $N \times n$ training set matrix X be composed of N input samples (or patterns of interest, such as face images) with n variables (or attributes, such as 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 samples. 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.

It is a proven result that the set of m $(m \leq n)$ eigenvectors of S, which corresponds to the m largest eigenvalues, minimizes the mean square reconstruction error over all choices of m orthonormal basis vectors [21].

To note explicitly the spatial association between the j^{th} and k^{th} variables, we can rewrite the sample covariance matrix S described in equation (1) in order to indicate the position of each variable in the N samples. When n variables are observed on each sample, the sample variation can be described by the following sample variance-covariance equation [22]:

$$S = \{s_{jk}\} = \left\{ \frac{1}{(N-1)} \sum_{i=1}^{N} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \right\}, \tag{2}$$

for $j = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, n$. The covariance s_{jk} between the j^{th} and k^{th} variables reduces to the sample variance when $j = k$, $s_{jk} = s_{kj}$ for all j and k, and the covariance matrix S contains n variances and $\frac{1}{2}n(n-1)$ potentially different covariances [22].

It is clear from equation (2) that the variable deviations from the mean have the same importance in the standard sample covariance matrix S formulation. In other words, all the n variables are equally weighted. However, there are situations where this should not be the case, particularly in pattern recognition problems where some parts of the samples might be more informative than others.

2.1 Weighted Sample Covariance

The well-known Pearson's sample correlation coefficient between the j^{th} and k^{th} variables is defined as [22]:

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

for $j = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, n$. It is important to note that $r_{jk} = r_{kj}$ for all j and k.

From equation (3), it is clear that the sample correlation coefficient is a normalized version of the sample covariance, where the product of the square roots of the sample variances, known as the sample standard deviations, provides the spatial normalization of the sum of the variable deviations from the mean [22]. In other words, r_{jk} is a measure of the linear association between two variables that does not allow variables with larger variance or scale to dominate the corresponding deviations from the mean and, consequently, the subspace calculation of PCA.

In our method, we want to give higher importance to the variables which characterise a class of interest. However, the variables that vary most are not necessarily the ones that allow best interpretation of the sample groups. Therefore, we need to define a measure of association between variables, based on the Pearson's sample correlation coefficient, which uses the notion of spatial weights and is more or less dominant depending on the values of each spatial weight.

Extending equation (3), we can define a weighted sample covariance r^*_{jk} between the j^{th} and k^{th} variables by [23]

$$r^*_{jk} = \frac{(\sqrt{w_j}\sqrt{w_k})s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}} \tag{4}$$
$$= \frac{\sum_{i=1}^{N}\sqrt{w_j}(x_{ij}-\bar{x}_j)\sqrt{w_k}(x_{ik}-\bar{x}_k)}{\sqrt{\sum_{i=1}^{N}(x_{ij}-\bar{x}_j)^2}\sqrt{\sum_{i=1}^{N}(x_{ik}-\bar{x}_k)^2}},$$

for $j = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, n$. The spatial weighting vector

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

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

$$R^* = \{r^*_{jk}\} = \left\{ \frac{\sum_{i=1}^{N}\sqrt{w_j}(x_{ij}-\bar{x}_j)\sqrt{w_k}(x_{ik}-\bar{x}_k)}{\sqrt{\sum_{i=1}^{N}(x_{ij}-\bar{x}_j)^2}\sqrt{\sum_{i=1}^{N}(x_{ik}-\bar{x}_k)^2}} \right\}, \tag{6}$$

for $j = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, n$. The weighted sample covariance r^*_{jk} between the j^{th} and k^{th} variables is equal to w_j when $j = k$, $r^*_{jk} = r^*_{kj}$ for all j and k, and so the matrix R^* is a nxn symmetric matrix.

Let R^* have respectively P^* and Λ^* eigenvector and eigenvalue matrices, that is,

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

The set of m ($m \leq n$) eigenvectors of R^*, that is, $P^* = [\mathbf{p}^*_1, \mathbf{p}^*_2, \ldots, \mathbf{p}^*_m]$, which corresponds to the m largest eigenvalues, defines a new orthonormal coordinate system for the training set matrix X and is called here as the *spatially weighted principal components.*

In the last years, weighted PCA techniques have been proposed [24, 23, 25, 26] to obtain a consistent subspace representation of the original data in the presence of noise, outliers and missing data. However, a key remaining issue for the weighted PCA methods in general is how to automatically compute the optimal weights to combine low level features, such as colour, shape and texture inherent to problems like face image analysis, with high level semantics, such as labeled information from human reasoning. In other words, the remaining question is: how can we define spatial weights w_j to incorporate prior knowledge? Our approach is to define a systematic method to compute the weights from labeled data.

2.2 The Spatial Weights

We propose the idea of using the discriminant weights given by statistical separating hyperplanes as the spatial weights of the weighted sample correlation matrix defined in equation (6). The models need some labeled data of N pairs

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N), \tag{8}$$

where the $\mathbf{x}_i \in \Re^n$ denote the i^{th} training observations and y_i are scalars that correspond to the classification labels. For simplicity and without loss of generality, we concentrate on two-class problems, that is, $y_i \in \{-1, 1\}$.

One way to define the parametric spatial weights is provided by Linear Discriminant Analysis (LDA) [27, 21]. LDA depends on all of the data, even points far away from the separating hyperplane and its main objective is to find a projection vector $\mathbf{w}_{lda}$ that maximizes the Fisher's criterion [21]:

$$\mathbf{w}_{lda} = \arg\max_{\mathbf{w}} \frac{|\mathbf{w}^T S_b \mathbf{w}|}{|\mathbf{w}^T S_w \mathbf{w}|}. \tag{9}$$

The S_b and S_w matrices are the between-class and within-class scatter matrices. The vector $\mathbf{w}_{lda}$ defines the normal vector of the hyperplane that best separates the two classes.

Alternatively, to allow the investigation of spatial discriminant weights determined by non-parametric separating hyperplanes, we can use the Support Vector Machine method [28] based on the risk-minimization approach. The primary purpose of SVM is to maximize the width of the margin between two distinct sample classes [28]. Given a training set as described in the formulation (8), the SVM method seeks to find the hyperplane defined by

$$f(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{w}) + b = 0, \tag{10}$$

which separates positive and negative observations with the maximum margin. It can be shown that the solution vector $\mathbf{w}_{svm}$ is defined in terms of a linear combination of the training observations, that is,

$$\mathbf{w}_{svm} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i, \tag{11}$$

where α_i are non-negative coefficients obtained by solving a quadratic optimization problem with linear inequality constraints. Those training observations $\mathbf{x}_i$ with non-zero α_i lie on the boundary of the margin and are called support vectors [28].

2.3 The Step-by-Step Algorithm

The main steps for calculating the spatially weighted principal components $P^* = [\mathbf{p}_1^*, \mathbf{p}_2^*, \ldots, \mathbf{p}_m^*]$ of an $N \times n$ training set matrix X composed of N input samples with n variables can then be described as follows:

1. Calculate the spatial weighting vector $\mathbf{w} = [w_1, w_2, \ldots, w_n]^T$ using some labeled data and a separating hyperplane method, 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 $\bar{x}_j = 0$ and $s_j^2 = s_{jj} = 1$, 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 the vector of variances $(s_1^2, s_2^2, \ldots, s_n^2)$, where

$$s_j^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (x_{ij} - \bar{x}_j)^2,$$

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

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{s_j^2}}$$

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 spatially weighted principal components P^* are then the eigenvectors corresponding to the m largest eigenvalues of $(Z^*)^T Z^*$, where

$$Z^* = \{\mathbf{z}_1^*, \mathbf{z}_2^*, \ldots, \mathbf{z}_N^*\}^T .$$

3 Experimental Results

We have divided our experimental results into two parts. Firstly, we have investigated the usefulness of the priori-driven principal components in recognizing samples compared to the standard PCA and the corresponding separating hyperplanes. Then, in the second part, we have analyzed the effectiveness of the new principal components in reconstructing samples compared to the standard PCA.

The following two-group separation tasks have been performed using frontal face images: (a) Gender experiments (female versus male samples); (b) Facial expression experiments (non-smiling versus smiling). The goal of the gender experiment is to evaluate the method proposed on a discriminant task where the differences between the groups are evident. The facial expression experiment poses an alternative analysis where there are subtle differences between the groups.

In all experiments, the total number of training examples N is limited and significantly less than the dimension of the feature space, that is, $N \ll n$. To address this problem for the Fisher's criterion, we have calculated the leading eigenvector $\mathbf{w}_{lda}$ by using two different approaches. The first approach, based on the Zhu and Martinez method [29], replaces S_w with the nxn identity matrix and $\mathbf{w}_{lda}$ becomes simply the leading eigenvector of S_b. The other, based on the Maximum uncertainty Linear Discriminant Analysis (MLDA) proposed by Thomaz et al. [30], considers the issue of regularizing the S_w estimate with a multiple of the identity matrix.

3.1 Recognition Rate

We have used two publicly available data sets to evaluate the classification performance of the spatially weighted principal components: FEI [31] and FERET [32]. 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. We adopted the 10-fold cross validation method to evaluate the classification performance of all the methods. Throughout all the classification experiments, we have assumed that the prior probabilities and misclassification costs are equal for both groups. On the PCA subspace, the mean of each class has been calculated from the corresponding training images and the Mahalanobis distance from each class mean has been used to assign a test observation to either groups. In all the standard and weighted PCA experiments, we have considered different numbers of principal components to calculate the recognition rates of the corresponding methods implemented. Additionally, as benchmark measures, we have also calculated the classification performance of the separating hyperplanes on the corresponding original spaces.

Figure 1 shows the recognition performance of the 10-fold cross validation of the gender experiments using the FEI and FERET databases and different numbers k of principal components selected by the corresponding largest eigenvalues. The horizontal dashed lines denote the separating hyperplanes' classification accuracies using all the original features available without any dimensionality reduction. It can be seen that even in such experiments where the differences between the sample groups are not subtle, the use of prior information given by labeled samples improves the discriminant power of the principal components, allowing similar or higher average recognition rates with the same number of components. For instance, in the gender experiments using the FEI database, all the spatially weighted PCA methods consistently outperform the standard

PCA when the number of principal components retained has been higher than 10, that is, when $k \geq 10$. In the gender experiments using the FERET database, which is composed of frontal face images not as well aligned as in the FEI database, the superiority of the spatially weighted PCA is less evident, but still it is possible to see a better classification performance than the standard PCA when using few principal components, that is, when $5 \leq k < 40$. Additionally, it is possible to see on the boxplots of Figure 1 that the top recognition rates of the spatially weighted principal components are comparable to or higher than the separating hyperplanes, but less sensitive to the parametric (Zhu&Martinez and MLDA) or non-parametric (SVM) discriminant information used, particularly for the FEI experiments.

The importance of allowing a priori-driven treatment of individual pixels and, consequently, minimizing the potential problem of discarding information related to subtle group differences on the first components of the standard PCA can be seen in Figure 2. In both FEI and FERET face databases, the average recognition rates of the spatially weighted principal components are much higher than the standard ones when the original dimensionality of the data is considerably

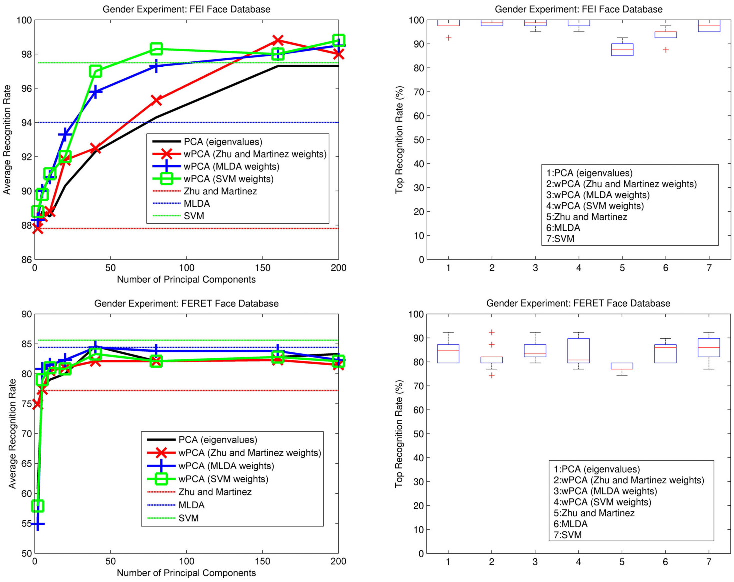

Fig. 1. Gender recognition performance of spatially weighted PCA (wPCA) compared to standard PCA using the FEI and FERET databases with 10-fold cross validation. On the left there are the average recognition rate curves using different numbers of principal components. As reference values, the horizontal dashed lines denote the corresponding separating hyperplane classification accuracies using all the original features without any dimensionality reduction. On the right there are boxplots of the top recognition rates achieved on a specific number of principal components for each method considered.

reduced. For example, when using only $k = 5$ spatially weighted principal components in the FEI face database, it is possible to achieve an average recognition rate of approximately 92% compared to 55% of the standard PCA. A significant improvement in classification performance is illustrated as well in the expression experiments using the FERET face database, where the spatially weighted and the standard principal components have achieved respectively approximately 71% and 57% with $k = 40$ components, for instance. In the expression experiments, more remarkably, the top recognition rates of the spatially weighted PCA are also comparable to or higher than the separating hyperplanes, but much less sensitive to the choice of using parametric or non-parametric spatial discriminant weights and consequently less prone to overfitting.

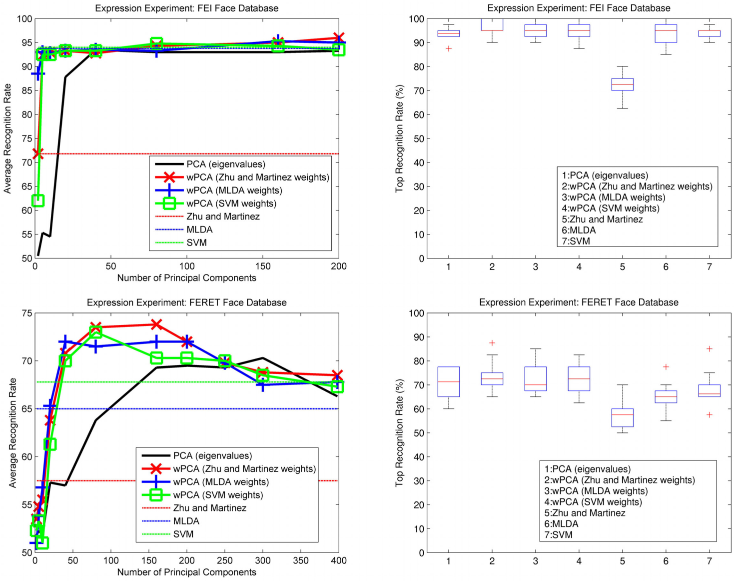

Fig. 2. Expression recognition performance of spatially weighted PCA (wPCA) compared to standard PCA using the FEI (top) and FERET (bottom) databases with 10-fold cross validation. On the left there are the average recognition rate curves using different numbers of principal components. As reference values, the horizontal dashed lines denote the corresponding separating hyperplane classification accuracies using all the original features without any dimensionality reduction. On the right there are the boxplots of the top recognition rates achieved on a specific number of principal components for each method considered.

3.2 Reconstruction

The reconstruction task cannot be performed by separating hyperplanes because both parametric and non-parametric classifiers retain only the information necessary to discriminate the classes, which is not enough to represent them back in the original feature space. In terms of the spatially weighted principal components, however, we can carry out an overall reconstruction process similar to unweighted PCA, but more efficient for making predictions especially on the major axes of projection because of the spatial weights control over the individual pixels within the face images.

Figure 3 shows two examples of the correlations between an image and its reconstruction. Data is given for the whole image and three smaller parts (eyes, nose and mouth) exclusively. The subjects are a male smiling (left on Figure 3) and a female with neutral expression (right on Figure 3) taken from the FEI database. The images are projected into the eigenspace and then reconstructed using 5, 10, 20, 40, 80, 160, 320 and all principal components. Three different methods were used for each experiment which are, from top to bottom: standard PCA, spatially weighted PCA using gender and spatially weighted PCA using expression. The spatial weights were calculated using the MLDA method and the other two separating hyperplanes considered in the previous subsection gave similar correlation results.

It is possible to see that the eigensubspace composed of the weighted principal components tends to reconstruct first the most informative parts of the face images for predicting differences relating to the choice of spatial weights. For example, on the left part of Figure 3 the image to be reconstructed is of a smiling face. Hence it is the region around the mouth that carries the most important discriminant information. Using spatially weighted PCA, with the expression discriminant weights we need only 5 weighted principal components to reconstruct the mouth with high correlation (> 0.7). Standard PCA needs at least 20 components to reconstruct the mouth correctly. The feature space weighted by gender information does not

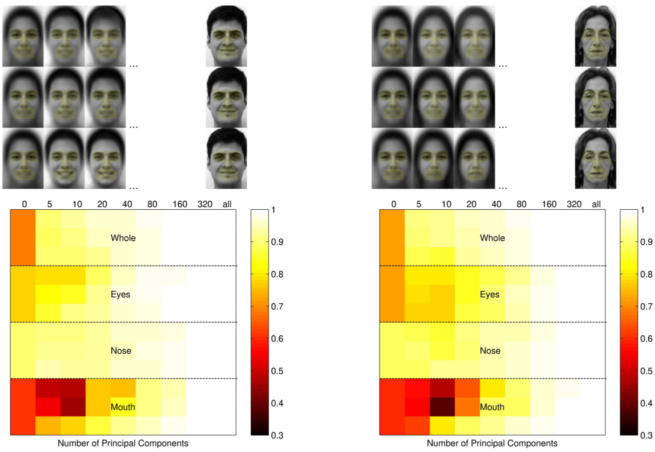

Fig. 3. Correlations between parts of smiling male (left) and non-smiling female (right) images and their reconstructions using different numbers of principal components and the following feature spaces (from top to bottom): standard PCA, PCA weighted by gender and PCA weighted by expression. The images show some of the partial reconstructions.

focus on large changes in the mouth, but contains information on other parts of the faces that better describe the main differences between male and female. It overtakes the standard PCA in reconstruction accuracy with around 40 components. A similar behavior can be observed on the right part of Figure 3, but now exemplifying a non-smiling sample reconstruction.

4 Conclusion

We have proposed a priori-driven PCA method using a modification of the Pearson's correlation formula that incorporates domain knowledge and generates an embedding space that preserves the properties of dimensionality reduction and interpretability of the standard PCA, without jeopardizing its inherent straightforward and simple calculation. This approach might be particularly useful for visual analytics and human perception experiments because it not only provides a more flexible form of data compression unlimited by the number of separating groups or classes, but also extracts relevant features in low dimension spaces providing better understanding and interpretation of the data for any specific *a priori* information of interest.

Acknowledgement. The authors would like to thank the support provided by LNCC, FAPESP (grants 2005/02899-4 and 2009/53556-0), CNPq (grants 473219/04-2 and 472386/07-7) and CAPES (grant 094/2007).

References

1. Pearson, K.: On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2, 559–572 (1901)
2. Jolliffe, I.: Principal component analysis. Springer Series in Statistics (2002)
3. Sirovich, L., Kirby, M.: Low-dimensional procedure for the characterization of human faces. Journal of Optical Society of America 4(3), 519–524 (1987)
4. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3, 71–86 (1991)
5. Hancock, P.J.B., Burton, A.M., Bruce, V.: Face processing: Human perception and principal components analysis. Memory and Cognition 24, 26–40 (1996)
6. O'Toole, A.J., Deffenbacher, K.A., Valentin, D., Mckee, K., Huff, D., Abdi, H.: The perception of face gender: The role of stimulus structure in recognition and classification. Memory and Cognition 26, 146–160 (1997)
7. Burton, A.M., Bruce, V., Hancock, P.J.B.: From pixels to people: a model of familiar face recognition. Cognitive Science 23, 1–31 (1999)
8. Calder, A.J., Burton, A.M., Miller, P., Young, A.W., Akamatsu, S.: A principal component analysis of facial expressions. Vision Research 41, 1179–1208 (2001)
9. Calder, A.J., Young, A.W.: Understanding the recognition of facial identity and facial expression. Nature Reviews: Neuroscience 6, 641–651 (2005)
10. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research, 1157–1182 (2003)
11. Zhang, D., Zhou, Z.H., Chen, S.: Semi-supervised dimensionality reduction. In: Proc. of the 2007 SIAM Intern. Conf. on Data Mining (2007)

12. Sugiyama, M., Idé, T., Nakajima, S., Sese, J.: Semi-Supervised Local Fisher Discriminant Analysis for Dimensionality Reduction. In: Washio, T., Suzuki, E., Ting, K.M., Inokuchi, A. (eds.) PAKDD 2008. LNCS (LNAI), vol. 5012, pp. 333–344. Springer, Heidelberg (2008)
13. Cai, D., He, X., Han, J.: Semi-supervised discriminant analysis. In: Proceedings of the International Conference on Computer Vision, ICCV 2007 (2007)
14. Lim, G., Park, C.H.: Semi-supervised dimension reduction using graph-based discriminant analysis. In: Inter. Conf. on Comp. and Inf. Tech., vol. 1, pp. 9–13 (2009)
15. Sun, D., Zhang, D.: A new discriminant principal component analysis method with partial supervision. Neural Processing Letters 30, 103–112 (2009)
16. Tipping, M.E., Bishop, C.M.: Mixtures of probabilistic principal component analysers. Neural Computation 11, 443–482 (1999)
17. Yu, S., Yu, K., Tresp, V., Kriegel, H.P., Wu, M.: Supervised probabilistic principal component analysis. In: Proc. of the ACM SIGKDD Intern. Conf. (2006)
18. Das, K., Nenadic, Z.: An efficient discriminant-based solution for small sample size problem. Pattern Recognition 42, 857–866 (2009)
19. Maszczyk, T., Duch, W.: Support Vector Machines for Visualization and Dimensionality Reduction. In: Kůrková, V., Neruda, R., Koutník, J. (eds.) ICANN 2008, Part I. LNCS, vol. 5163, pp. 346–356. Springer, Heidelberg (2008)
20. Wang, F., Wang, X., Li, T.: Beyond the graphs: Semi-parametric semi-supervised discriminant analysis. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 2113–2120 (2009)
21. Fukunaga, K.: Introduction to Statistical Pattern Recognition. Academic Press, New York (1990)
22. Johnson, R., Wichern, D.: Applied Multivariate Statistical Analysis. Prentice Hall, New Jersey (1998)
23. Costa, J.F.P., Silva, I., Silva, M.E.: Time dependent principal component analysis of time series data. In: IASC (2007)
24. Forbes, K., Fiume, E.: An efficient search algorithm for motion data using weighted pca. In: Proc. of the 2005 ACM SIGGRAPH/Eurographics Symp. on Comp. Anim., pp. 67–76 (2005)
25. Skocaj, D., Leonardis, A., Bischof, H.: Weighted and robust learning of subspace representations. Pattern Recogn. 40, 1556–1569 (2007)
26. Zhang, Y., Qiu, Z., Sun, D.: Palmprint identification using weighted pca feature. In: Proc. of the Inter. Conf. on Signal Processing, pp. 2112–2115 (2008)
27. Devijver, P., Kittler, J.: Pattern Classification: A Statistical Approach. Prentice-Hall (1982)
28. Vapnik, V.N.: Statistical Learning Theory. John Wiley & Sons, Inc. (1998)
29. Zhu, M., Martinez, A.M.: Selecting principal components in a two-stage lda algorithm. In: CVPR 2006, pp. 132–137 (2006)
30. Thomaz, C.E., Kitani, E.C., Gillies, D.F.: A maximum uncertainty lda-based approach for limited sample size problems - with application to face recognition. Journal of the Brazilian Computer Society 12, 7–18 (2006)
31. Thomaz, C.E., Giraldi, G.A.: A new ranking method for principal components analysis and its application to face image analysis. Image and Vision Computing 28, 902–913 (2010)
32. Philips, P.J., Wechsler, H., Huang, J., Rauss, P.: The feret database and evaluation procedure for face recognition algorithms. Image and Vision Computing 16(5), 295–306 (1998)