View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Servicio de Coordinación de Bibliotecas de la Universidad Politécnica de Madrid

Class-Conditional Probabilistic Principal Component Analysis: Application to Gender Recognition

Análisis de Componentes Principales Probabilístico Condicionado a la Clase: Aplicación al Reconocimiento de Género

Juan Bekios Calfa[1], José M. Buenaposada[2] and Luis Baumela[3]
[1]Dept. de Ing. de Sistemas y Computación, Universidad Católica del Norte,
Av. Angamos 0610, Antofagasta, Chile.
juan.bekios@ucn.cl
[2]Dept. de Ciencias de la Computación, Universidad Rey Juan Carlos,
Calle Tulipán s/n, 28933, Móstoles, Spain.
josemiguel.buenaposada@urjc.es
[3]Dept. de Inteligencia Artificial, Universidad Politécnica de Madrid,
Campus Montegancedo s/n, 28660 Boadilla del Monte, Spain.
lbaumela@fi.upm.es,
http://www.dia.fi.upm.es/~pcr

Article received on January 29, 2010; accepted on June 17, 2010

Abstract. This paper presents a solution to the problem of recognizing the gender of a human face from an image. We adopt a holistic approach by using the cropped and normalized texture of the face as input to a Naïve Bayes classifier. First it is introduced the Class-Conditional Probabilistic Principal Component Analysis (CC-PPCA) technique to reduce the dimensionality of the classification attribute vector and enforce the independence assumption of the classifier. This new approach has the desirable property of a simple parametric model for the marginals. Moreover this model can be estimated with very few data. In the experiments conducted we show that using CC-PPCA we get 90% classification accuracy, which is similar result to the best in the literature. The proposed method is very simple to train and implement.
Keywords: Gender classification, face analysis, class conditional PPCA.

Resumen. Este trabajo presenta una solución al problema del reconocimiento del género de un rostro humano a partir de una imagen. Adoptamos una aproximación que utiliza la cara completa a través de la textura de la cara normalizada y redimensionada como entrada a un clasificador Näive Bayes. Presentamos la técnica de Análisis de Componentes Principales Probabilístico Condicionado-a-la-Clase (CC-PPCA) para reducir la dimensionalidad de los vectores de características para la clasificación y asegurar la asunción de independencia para el clasificador. Esta nueva aproximación tiene la deseable propiedad de presentar un modelo paramétrico sencillo para las marginales. Además, este modelo puede estimarse con muy pocos datos. En los experimentos que hemos desarrollados mostramos que CC-PPCA obtiene un 90% de acierto en la clasificación, resultado muy similar al mejor presentado en la literatura. El modelo propuesto es muy sencillo de entrenar e implementar.
Palabras clave: Clasificación de género, análisis de caras, PPCA condicionado a la clase.

1 Introduction

Over the last years we have seen important advances in the areas of face detection and tracking, which have boosted a wealth of practical applications, such as face, smile or blink detection in digital cameras. In this paper we focus on automatic gender recognition from face images, a problem with applications in advanced human-machine interaction, image retrieval and surveillance.

Gender recognition has attracted the interest from researchers in computer vision and pattern recognition for years [Moghaddam and Yang, 2002; Baluja and Rowley, 2007; Andreu and Mollineda, 2008; Mäkinen and Raisamo, 2008; Mäkinen, 2008; Golomb, et al., 1990; Shakhnarovich, et al., 2002; Lapedriza and Marin-Jimenéz, 2006; Verschae, et. al., 2006], being SEXNET [Golomb, et al., 1990] the first attempt to recognize gender from faces. Solutions to this problem may be broadly grouped into *holistic approaches*, that use the cropped, re-sized and illumination normalized texture of the whole face as classification attribute, and *feature-based approaches*, that are based on extracting a set of discriminative face features.

The best holistic classification results are based on using Support Vector Machine (SVM) as classifiers [Moghaddam and Yang, 2002; Andreu and Mollineda, 2008; Mäkinen and Raisamo, 2008]. Moghaddam and Yang [Moghaddam and Yang, 2002] reported 96.6% recognition success in classifying 1775 images from the FERET database using automatically aligned and cropped images and a 5-fold cross-validation. Baluja and Rowley [Baluja and Rowley, 2007] reported a bias in the previous estimation caused the existence of subjects with the same identity in different folds. They achieved 93.5% success on a similar experiment with manual alignment and a proper cross-validation. Andreu and Mollineda [Andreu and Mollineda, 2008] used Principal Component Analysis (PCA) and SVM to classify. They also studied the contribution of different face parts to the solution. They achieved recognition rates around 80% using classifiers based on the eyes, nose, mouth and chin individually. Whereas the joint consideration of all face parts increases the recognition rate to 95.21% (92.37% considering only the inner part of the face) for the FERET database and 90.64% (90.79% inner face only) for the XM2VTS database. Recently, Mäkinen and Raisamo [Mäkinen and Raisamo, 2008] reported a large set of experiments using 411 images from the FERET database. The main conclusion from their work was that the automatic alignment should be very precise to be worthy. They got 86% success rate with non-aligned faces using SVM with RBF kernel.

Feature-based approaches use grey level pixel-wise differences [Baluja and Rowley, 2007], haar-like wavelets [Shakhnarovich, et al., 2002; Mäkinen and Raisamo, 2008; Mäkinen, 2008], multi-scale filter banks [Lapedriza and Marin-Jimenéz, 2006] or Locally Binary Patterns (LBP) [Mäkinen and Raisamo, 2008; Mäkinen, 2008] features to recognize the gender of a face. Shakhmarovich et al [Shakhnarovich, et al., 2002] achieved respectively 79% and 79.2% recognition accuracy in gender and ethnicity classification on a set of difficult images obtained from the web. They used Haar-like features within an AdaBoost-based approach which is various orders of magnitude faster than SVM. In contrast to Mäkinen [Mäkinen and Raisamo, 2008], Verschae et al. [Verschae, et. al., 2006] achieved similar results for the FERET database using manual (85.5%) and automatic (85.8%) image alignment. They also achieved similar recognition rates with AdaBoost and SVM with RBF kernel. Baluja and Rowley [Baluja and Rowley, 2007] used pixel-wise grey level comparisons as weak classifiers within an AdaBoost learning scheme. They used manually aligned images from “fa” and “fb” Color FERET database achieving 94% recognition rate. Their classifier is approximately 50 times faster than Moghaddan’s [Moghaddam and Yang, 2002] SVM solution. Lapedriza et al [Lapedriza and Marin-Jimenéz, 2006] performed a set of experiments using image features computed from multi-scale filter banks applied to internal and external face locations combined with various boosting classifiers and acquisition conditions. They obtained the best results using JointBoosting and combining internal and external face features. With this setup they achieved 96.7% and 91.7% recognition rate for images acquired respectively in controlled and uncontrolled conditions from the Face Recognition Grand Challenge data[1].

An important conclusion from the results reported in the literature is that it is quite difficult to compare between the different results since researchers did not use the same database or the same images within the same database. This is the reason for which Mäkinen and Raisamo did their full set of gender recognition experiments [Mäkinen and Raisamo, 2008; Mäkinen, 2008]. Nevertheless, they used very few testing and training images (411), from which definitive conclusions may not be extracted. In summary, state-of-the-art algorithms roughly achieve 90% recognition accuracy in laboratory conditions [Mäkinen and Raisamo, 2008; Mäkinen, 2008] when training with one database and testing on another.

In this paper we introduce a very simple gender recognition algorithm with performance comparable to the best in the literature. We use the simple but effective Naïve Bayes Classifier (NBC) for solving the classification problem. A major limitation of this classifier is that it assumes that attributes are independent from each other, something which is not true for our data. This problem has been previously addressed in the literature by relaxing the independence assumptions, e.g. Tree Augmented and General Bayes Nets [Cheng and Greiner, 1999], or by using preprocessing procedures to transform the data to satisfy the independence assumption [Prasad, et al., 2004; Bressan, et. al., 2001; Bressan and Vitrià, 2002]. Bressan and Vitrià achieved the independence using the Independent Component

[1] see *http://www.frvt.org/FRGC/*

Analysis (ICA) transformation, a linear projection of the data onto a low dimensional subspace in which the statistical dependence among the attributes is minimized.

Our proposal is directly related to the work of Bressan and Vitrià [Bressan and Vitrià, 2001; Bressan and Vitrià, 2002] that introduces the Class-Conditional ICA (CC-ICA) as a preprocessing step for reducing the dimensionality and enforcing the independence of the data attributes for each class-conditional distribution. CC-ICA has, nevertheless, some drawbacks. If the sample size is small we may not be able to compute ICA [Fan and Kim, 2007]. Also, estimating the density of each attribute marginal may be a complex task, since ICA maximizes the non-gaussianity of the marginals. We have to resort to kernel density estimation [Fan and Kim, 2007] or select among different parametric models depending on the sparsity of the result [Bressan, et. al., 2001].

In this paper we introduce the Class-Conditional Probabilistic Principal Component Analysis (CC-PPCA) representation. Like CC-ICA, it introduces a class-conditional transformation to achieve in-correlation in the lower-dimensional space. Contrary to ICA, PPCA provides Gaussian marginals and, for the Gaussian distribution, in-correlation and independence are equivalent properties. So, with CC-PPCA, like with CC-ICA, we look for a projection of the data that maximizes the independence of the marginals. The difference is that now we maximize the gaussianity of the marginals. This new approach has the desirable property that we always have a simple parametric model for the marginals that can be estimated with very few data.

2 Class Conditional PPCA (CC-PPCA)

2.1 Probabilistic Principal Component Analysis

Probabilistic Principal Component Analysis (PPCA) [Tipping and Bishop, 1999] is a probabilistic version of Principal Component Analysis based on a latent variable model that arises when latent variables are integrated out and the parameters of the model are optimized by maximum likelihood. Using PPCA we can model the relation between an observed vector, t, in a d dimensional subspace and a latent vector, x, belonging to a q dimensional subspace, given d > q, as:

$$t = Wx + \mu + \epsilon$$

where W is a $d \times q$ matrix that relates latent and observed subspaces, μ is the model mean, and ϵ is the noise vector. The latent variables are defined to be independent and Gaussian with unit variance, $x \sim \mathcal{N}(0, I)$. The error, or noise, model is an isotropic Gaussian, $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$ which induces Gaussian distributions on the observation vectors, $t \sim \mathcal{N}(\mu, WW^T + \sigma^2 I)$.

The maximum-likelihood estimator for μ is given by the sample mean, and for W and σ^2 the maximum-likelihood estimators are given by:

$$W_{ML} = U_q\left(\Delta_q - \sigma^2\right)^{1/2} R,$$
$$\sigma^2_{ML} = \frac{1}{d-q} \sum_{j=q+1}^{d} \lambda_j,$$

where the $d \times q$ matrix U_q stores the eigenvectors resulting from computing PCA on the set of sample vectors and their corresponding eigenvalues $\lambda_1, \ldots, \lambda_q$ are stored in the diagonal matrix Δ_q. The matrix R is an arbitrary rotation matrix that in practise can be assumed the identity. On the other hand σ^2 can be interpreted as the variance 'lost' in the projection on the latent subspace, averaged along the lost dimensions.

2.2 Dimensionality reduction in PPCA

The aim of Probabilistic Principal Component Analysis (PCA) is to transform observed data into a reduced dimensionality representation.
Following [Tipping and Bishop, 1999] we use the distribution of $p(x|t)$ to compute the projection of a vector t onto the lower dimensional latent subspace, x. Thus, we use the mean of $p(x|t)$ to estimate the low dimensional representation of any observed vector, t,

$$x = (W^T W + \sigma^2 I)^{-1} W^T (t - \mu) \quad (1)$$

2.3 Class conditional Naïve Bayes classification

A property of the variables in the PPCA latent subspace is that they are uncorrelated, Gaussian, and thus independent. We exploit this fact to enforce the independence requirement of Naïve Bayes by using the projected data as input to the classifier.

Let us assume that we have data from k classes, $C_1, \dots, C_k$. We estimate a PPCA model for each class. After that, by using equation (1), we project the sample data vector t, over each class-conditional PPCA model i obtaining a vector of latent variables x_i for each class i. Thus we can obtain $p(x_i|C_i)$ for each class C_i by using the well-known Naïve Bayes rule. We will classify t in the class given by:

$$C_i = arg \max_{i=1,\dots,k} p(x_i|C_i)p(C_i). \quad (2)$$

Assuming that all classes were equiprobable we can remove $p(C_i)$ from equation (2). The distribution of each marginal attribute of our classifier, $p(x_i^j|C_i)$, is a zero mean unit variance Gaussian from which we can get:

$$p(x_i|C_i) = v_i \prod_{j=1}^{q_i} p(x_i^j|C_i) = v_i \prod_{j=1}^{q_i} \mathcal{N}\left(x_i^j; 0,1\right), \quad (3)$$

where x_i^j is the j-th component of vector x_i and q_i is the number of dimensions of the PPCA latent subspace built for class i. Finally we use

$$v_i = \frac{1}{|T|} \sum_{\{t \in T\}} p(s_i|C_i),$$

a normalization constant so that $\int p(x_i|C_i)dx = 1$, and where T is the training set for all classes, $|.|$ is the cardinal operator, and s_i is the training vector t projected over class i latent subspace.

Using the log-likelihood in equation (3) we can transform products into sums and, considering again equiprobable classes, equation (2) is transformed into:

$$C_i = \arg \max_{i=1,\dots,k} (\log(v_i) + \sum_{j=1}^{q_i} \log(\mathcal{N}(x_i^j; 0,1)) \quad (4)$$

In the gender recognition case the classification rule given in equation (4) can be rewritten as:

$$C(t) = \begin{cases} \text{MALE} & , \quad \log p(x_m|C_m) > \log p(x_f|C_f) \\ \text{FEMALE} & , \quad \text{otherwise} \end{cases}$$

where, x_m is the projection of t in the male PPCA model and x_f is the projection of t in the female PPCA model.

3 Experiments

We have chosen the simple, but yet effective, CC-PPCA procedure to make gender classification on a single face image. As the input for our gender recognition algorithm we crop the images to 25×25 pixels by using a state-of-the-art face detector [Viola and Jones, 2004] (we simply re-size to 25×25 the image region marked by the detector as a face). Additionally, we apply simple histogram equalization in order to gain some independence to illumination changes and then we apply an oval mask in order to avoid background (see Fig. 1).

Fig. 1. The first row displays raw cropped face images using the face detector. The second shows equalized and masked images

We further reduce the dimensionality of the face images subspace in order to get a better separation between male and female classes. In our case we have chosen CC-PPCA to build two subspaces one for males and one for females and to project the input images on each subspace using equation (1). Our gender recognition approach consists on two steps:

- **Training:** We use standard face databases on which images are labeled with gender. We preprocess (by cropping, histogram equalizing and masking) all the images in this databases using face detection to get all the input faces for the training process. Using this preprocessed images we perform per-class PCA in order to construct the matrices needed for PPCA projection for the male and female subspaces.
- **Single image Classification:** Given a preprocessed face image I (cropped, histogram equalised and masked) we project it onto the male and female subspaces obtaining, x_m and x_f,

respectively as male and female projections. The classification is done using equation (4).

We compare our approach with one of the two algorithms that are the state of the art in gender recognition, the SVM with Radial Basis Functions kernel (SVM+RBF) used by Mogghaddam et al [Moghaddam and Yang, 2002]. We could compare against the Baluja and Rowley's algorithm but the main difference would be the speed and not the classification accuracy as shown in their work [Baluja and Rowley, 2007]. In our experiments we have used a non-public database from Universidad Católica del Norte in Chile with around 10,000 individuals, the UCN-DataBase, and two standard databases, the Color FERET [Phillips, et. al., 2000] with around 1,000 individuals and the Productive Aging Lab Face (PAL) database from the University of Texas at Dallas [Minear, et. Al. 2004].

The UCN-DataBase is a non-public database consisting on images of people from UCN (including students and lecturers). This data base is built with mug-shot images (one per individual), with different quality ranging from web-cam to digital still cameras used as imaging devices (see Fig. 2 for some sample images). All the faces are facing frontal to the camera with no occlusions. There are images from 10,700 individuals, from which 5646 are male and 5054 female. We used 5628 male and 5041 female images as the face detector missed some faces when preparing the data.

Fig. 2. Some cropped and re-sized images, after face detection, from the Chile Database

The Color FERET Face Database is a publicly available resource for face analysis research. In the database there are multiple images of 994 individuals, 591 male and 403 female. In our case, differently to Mogghadam et al [Moghaddam and Yang, 2002] and Baluja et al [Baluja and Rowley, 2007], we use only one image per subject from the FERET database fa gallery. In this gallery all the individuals are facing frontal to the camera with no occlusions (see Fig. 3 for some sample images). In our experiments we used 591 male images and as the detector missed one face, we use 402 female images.

Fig. 3. Some cropped and re-sized images, after face detection, from Color FERET database gallery fa

The PAL Database consists of frontal pictures of 576 individuals. The right profile and some facial expressions are also available for some subjects. There are 219 male and 357 female subjects divided into four groups depending on their age: 18-29, 30-49, 50-69 and 70-93 (see Fig. 4). All the faces are frontal and there are no occlusions.

Fig. 4. Some cropped and re-sized images, after face detection, from PAL database

3.1 Cross validation experiments over single databases

We performed experiments with SVM+RBF (we use WEKA Explorer[2] implementation of SVM+RBF) and CC-PPCA classifiers using different configurations in the training data. We have made experiments using a mask to remove the background effect and we have made different experiments using images directly from the face detector and images manually aligned (making use of eyes and mouth centre). In all cases we performed a 5 fold cross-validation scheme. In Table 1 we report average results for the 5 folds. Training images are cropped, re-sized to 25x25 pixels and equalized before masking.

[2] http://www.cs.waikato.ac.nz/ml/weka/

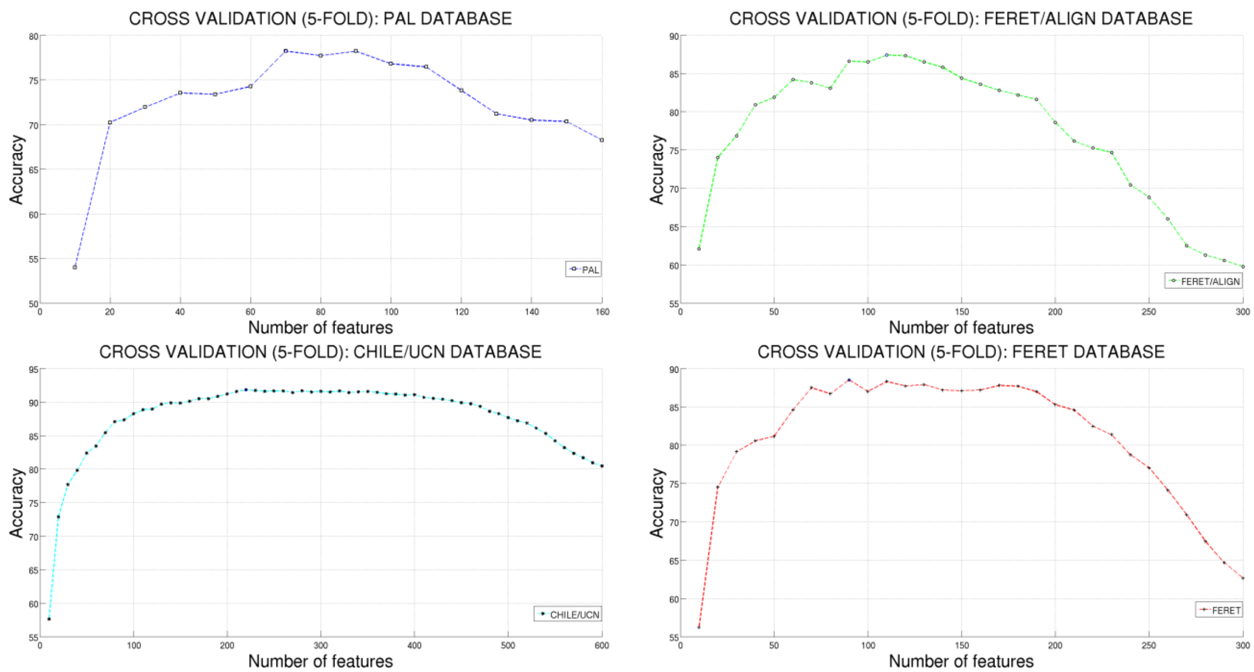

Fig. 4. Classification rate as a function of PPCA latent subspace dimensions

We have tested the CC-PPCA by increasing CC-PPCA latent subspace dimension of both classes, female and male, at the same time. In Fig. 4 we show respectively the results for the face detector (loosely aligned images) using UCN-DB and FERET databases, and also for manually aligned images, in this case, only for the FERET database (FERET Align). In Table Table we have summarized the best classification rates for all tests.

From Table 1 it is clear that SVM+RBF is the best performer but CC-PPCA get results quite similar to the SVM. The most interesting fact is that the CC-PPCA method is affected more by the size of the training sample set than SVM. In the case of CHILE-UCN with around 10,000 we get a recognition rate of 91.84% around 2% lower than the SVM result. On the other hand, the performance gap is bigger when the database has less samples as in the FERET case, around 3% lower recognition rate

for the CC-PPCA algorithm than SVM. In our case, contrary to Makïnen findings, the manual alignment did not improve the results in the FERET database. With face detectors as the one we use, the variance on eyes position over the cropped image is a function of the cropped image size. In other words, at a size of 25 pixels the face region is almost always aligned. This fact could explain why manual alignment does not improve over the face detection in our results on the FERET database.

The PAL database is especially difficult because it presents a greater variability in demography, with more race and age differences between the individuals. In this case, both algorithms SVM+RBF and CC-PPCA get lower recognition rate by more than a 10%.

If we now consider the execution time of the classifier, the state-of-the-art is the work done by Baluja et al. [Baluja and Rowley, 2007], since they use pixel-wise grey level differences, a feature that is quite fast to compute. On the other hand, from [Baluja and Rowley, 2007] it is not clear how well their algorithm will perform in face detector cropped images as they do not report results when training using this kind of images. Nevertheless, their system is able to get the same classification rate as the

SVM+RBF classifier with as few as 1000 pixel-wise comparisons (they report 19.53 μ sec execution time with 20×20 pixels images). On the other hand the SVM+RBF needs to project the input image (625 pixels) over each of the support vectors giving a total of 625x2997=1,873,125 pixel operations for the CHILE-UCN case. Finally, we need two PPCA projections of the input image, one for the female model and one for the male model, and therefore we need 2x625x220=275,000 pixel operations for the CHILE-UCN database (2x625x90=112,500 for the FERET database). Our execution time would be better than the SVM+RBF approach and worse than the Baluja et al. approach but still within the real-time range, always depending on the number of latent variable models chosen dimensions.

Table 1. Best classification rates for cross-validation experiments. In the SVM+RBF results we write in parenthesis the number of support vectors used in the tests. In the CC-PPCA results we show the dimensions of the latent variable subspace

DATABASES	CLASSIFIERS	
	SVM+RBF	CC-PPCA
PAL	83.77±1.61% (365)	78.24±3.64% (70)
CHILE-UCN	93.52±0.46% (2997)	91.84±0.66% (220)
FERET	91.54±3.08% (410)	88.51±2.33% (90)
FERET-ALIGN	92.15±2.81% (434)	87.42±1.79% (110)

3.2 Training and testing with different databases

With the experiments in this section we will show the generalization capabilities of the CC-PPCA classifier. We have trained each algorithm (5-fold cross-validation) using images from one database and then we have classified the images from another. In Table 2 we have summarized our results. In the CC-PPCA case (see Table 2) the optimal dimensions for both female and male classes, were found by cross-validation and are the ones shown in Table 1. In the experiment CC-PPCA* we show the best results that we can achieve by changing the dimensions of the PPCA latent subspace.

FERET and CHILE-UCN databases have a very similar demography (see Figs. 2 and 3). Therefore, when we use them in the cross database experiments we get similar results to the single database experiments but with an overall decrease in performance. The results for the experiments FERET/CHILE-UCN are worse than the CHILE-UCN/FERET. The smaller size of the FERET database has a negative impact in the generalization capabilities of the classifier.

On the other hand, cross-database experiments using FERET, PAL and CHILE-UCN are very important to test the generalization capabilities of the classifiers as PAL has very different demography to FERET and CHILE-UCN. Here the drop in classification performance is more important than in the single database experiments in Table 1. SVM+RBF is the worst classifier in the FERET/PAL experiment, the most difficult one given the size of the FERET database, and in the CHILE-UCN/PAL. These experiments prove that the CC-PPCA approach generalizes better than the non-linear approach of the SVM+RBF in the gender recognition problem.

Table 2. Best classification rates for cross database experiments. In the CC-PPCA* results we show the dimensions of the latent variable subspace

DATABASES	CLASSIFIERS		
	SVM+RBF	CC-PPCA	CC-PPCA*
FERET/CHILE-UCN	81.22%	79.50%	80.23% (99)
CHILE-UCN/FERET	90.23%	90.13%	90.43% (222)
FERET/PAL	58.89%	66.03%	69.51% (84)
CHILE-UCN/PAL	69.43%	74.39%	75.26% (211)

4 Conclusions

In this paper we have introduced a simple but effective algorithm for gender recognition. We use a Naïve Bayes classifier and introduce the CC-PPCA technique to reduce the dimensionality of our data and enforce the independence assumption of the classifier. An important conclusion from the results reported in the literature is that it is quite difficult to compare them, since researchers did not use the same database or the same images within the same database. Nevertheless, with this simple approach we achieve 90% recognition accuracy when training with one database and testing on another, which is similar to the state-of-the-art classification rates. Moreover, our approach is very easy to train, based on a class-conditional PCA and Naïve Bayes. On the other hand, the other competing approaches use SVM+RBF or AdaBoost algorithm, can take hours to train in a large database, like the UCN-database.

Acknowledgments

José M. Buenaposada and Luis Baumela gratefully acknowledge funding from the Spanish **Ministerio de Educación y Ciencia** under contract TIN2008-06815-C02-02. Juan Bekios-Calfa and Luis Baumela were also funded by Spanish research programme Consolider-Ingenio 2010: MIPRCV CSD2007-00018.

Referencias

1 **Andreu, Y. , & Mollineda, R. A. (2008).** The role of face parts in gender recognition. 5th *International Conference on Image Analysis and Recognition*. *Lecture Notes in Computing Science,* 5112, 945–954.

2 **Baluja, S., & Rowley, H. A. (2007).** Boosting sex identification performance, *International Journal of Computer Vision*, 71 (1), 111-119.

3 **Bressan, M., Guillamet, D. & Vitrià, J. (2001).** Using an ICA representation of high dimensional data for object recognition and classification. *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001) Kauai. Hawaii, 1004-1009.*

4 **Bressan, M., & Vitrià, J.** (**2002).** Improving naive bayes using class-conditional *ICA. Advances in Artificial Intelligence — IBERAMIA 2002, Lecture Notes in Computer Science*, 2527, 1–10.

5 **Cheng, J., & Greiner, R.** (**1999).** Comparing bayesian network classifiers, *Fifteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, Stockholm, Sweden, 101–108.

6 **Fan, L., & L. Poh, Kim, (2007).** A comparative study of PCA, ICA and class-conditional ICA. *9th International Work-Conference on Artificial Neural Networks, Lecture Notes in Computer Science*, 4507, 16–22.

7 **Golomb, B. A., Lawrence, D. T. & Sejnowski, T. J. (1990).** Sexnet: A neural network identifies sex from human faces. *NIPS-31990 conference on Advances in neural information processing systems 3 (NIPS-3)*, Denver, CO, USA, 572–577.

8 **http://www.cs.waikato.ac.nz/ml/weka/,** Waikato environment for knowledge analysis, weka software.

9 **Lapedriza, J. V. A., Marin-Jiménez, M. J. & Vitria, J. (2006).** Gender recognition in non controlled environments. *18th International Conference on Pattern Recognition (ICPR 2006)*, Hong Kong, China, 834–837.

10 **Mäkinen, E., & Raisamo, R.** (**2008).** Evaluation of gender classification methods with automatically detected and aligned faces, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30 (3), 541 – 547.

11 **Mäkinen, R. R. Erno , (2008).** An experimental comparison of gender classification methods, *Pattern Recognition Letters*, 29 (10), 1544–1556,

12 **Minear, M., & Park, D. C., (2004).** A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments and Computers,* 36 (4), 630-633.

13 **Moghaddam, B., & Yang, M.-H. (2002).** Learning gender with support faces, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (5), 707–711,

14 **Phillips, P., Moon, H., Rauss, P. & Rizvi, S.** (**2000).** The feret evaluation methodology for face recognition algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 (10), 1090–1104,

15 **Prasad, M. N., Sowmya, A. & Koch, I. (2004).** Feature subset selection using ICA for classifying emphysema in HRCT images, 17th International Conference on Pattern Recognition (ICPR 2004), Cambridge UK, 4, 515–518.

16 **Shakhnarovich, G., Viola, P. A. & Moghaddam, B.** (**2002).** A unified learning framework for real time face detection and classification, *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, Washington, D.C., 14-21.

17 **Tipping, M. E., & Bishop, C. M. (1999).** Probabilistic principal component analysers, *Journal of the Royal Statistical Society, Series B*, 61 (3), 611–622.

18 **Verschae, R., Ruiz-del-Solar, J. & Corea, M. (2006).** Gender classification of faces using adaboost, 11th *Iberoamerican Congress on Pattern Recognition,* Lecture Notes in Computer Science, 4225, 68–78.

19 **Viola, P., & Jones, M. J. (2004).** Robust real-time face detection, *International Journal of Computer Vision*, 57 (2), 137–154.

Juan Bekios Calfa

Computer Science Engineer from Universidad Católica Norte (Chile). Since 1996 he is lecturer at Universidad Católica del Norte within the Systems Engineering and Computation Department. He is currently working towards the PhD degree in the Artificial Intelligence Department, Universidad Politécnica de Madrid, and his research topic is automatic face analysis using Computer Vision and Machine Learning techniques.

José Miguel Buenaposada

BS and MS in 1999 and PhD in 2005 all in Computer Science from the Universidad Politécnica de Madrid. Since October 2003 he is lecturing at Universidad Rey Juan Carlos and since 2008 he is Associate Professor of Computer Science (Profesor Contratado Doctor). His research field is Computer Vision in general and automatic facial analysis in particular (tracking, detection, facial expressions recognition, etc.). He has published several research papers in Computer Vision.

Luis Baumela

BS and MS 1998 PhD 1995 all in Computer Science from the Universidad Politécnica de Madrid. From 1989 to 1992 he was an R&D engineer at Telefónica Investigación y Desarrollo. Since 1997 he is Associate Professor of Computer Science at the Facultad de Informática of the Universidad Politécnica de Madrid. He has lead 9 research projects and published more than 40 research papers in Computer Vision.