

A Gender Detection Approach

Marcos del Pozo-Baños, Carlos M. Travieso,
Jaime R. Ticay-Rivas, and Jesús B. Alonso

*IDeTIC - Institute for Technological Development and Innovation in Communications,
ULPGC - Universidad de Las Palmas de Gran Canaria
Spain*

1. Introduction

Which are the brain processes that underlie facial identification? What information, among the available in the environment, is used to elaborate a response on a subject's identity? Certainly, our brain uses all the information in greater or less extent. Just focusing on that present on the human face, the system can obtain knowledge about gender, age and ethnicity. This demographic data may not be enough for subject identification, but it definitely gives us some valuable clues. The same can be applied for computer systems. For example, having gender information into account, the system can reduce the pool of the possible identities considerably, making the problem easier and enforcing the final response. Moreover, raw gender information can also be used in fields such as micromarketing and personalized services.

A practical example of this can be found on the work presented by Peng and Ding (Peng & Ding 2008). These authors proposed a tree structure system to increase the successful rate of a gender classification. In particular, the system first classify between Asian and Non-Asian ethnicities. Then, two specialized gender classification systems are trained, one for each ethnicity. This resulted in an increase of around 4% over an ordinary system (gender classifier without ethnicity specialization).

Therefore, demographic classification systems are as much important and valuables as face identification systems themselves. This is why they have received increasing attention in the last years. In particular, this chapter focuses its attention in facial-base gender-detection systems. A summary of the problem's characteristics is first given in section 2, along with an overview of the state of art. Section 3 introduces the structure of the system used for experiments of section 4, where we check the effect of preprocessing variations on the systems performance. Finally, conclusions derived from the obtained results are presented in section 5.

2. Biometric gender classification

In general, biometric problems can be classified in two groups: classification and verification. In the former, samples from a number of well defined classes are given to the system for training. When a testing sample is presented, the system must classify it in the corresponding class. In other words, the system answers the "who is this?" question.

On the other hand, during a verification problem training samples are divided in classes "A" (called positive) and "others" (called negative). Then, a testing sample claiming to be of class "A" is presented to the system and the system must verify that this sample corresponds to the claimed class. Obviously, a classification system can be built upon a combination of verification systems.

Tools show different behaviors in different situations. Some perform better in classification problems while others perform better during verification. This is because differences between classification and verification are not just a matter of the number of classes, but of how classes are built. In a classification problem, classes are well defined patterns coming from a common thing. However, this cannot be expected for the negative class of a verification problem. Usually this class is too wide in the sample space to be represented with a reasonable amount of samples, and other representation techniques must be used.

This does not mean that representation techniques that work on classification problems cannot perform on a verification scenario or vice versa. But usually you cannot expect them to work as well. Therefore it is important to define the problem before decide the approaching technique and the tools to be used. Gender classification systems find themselves in a rather special situation, as they only define two classes (male and female). Therefore classification and verification techniques can be used without penalties in this case.

2.1 Facial image databases

Due to the gaining importance of face identification systems on the security field, a great deal of facial databases has appeared in the last years. As any other component of a biometric system, databases' technology has experienced an important step forward too.

Some of the first widely used databases were the Olivetti Research Laboratory database (Samaria & Harter 1994), also known as AT&T, and the YALE database. These databases consist of images taken from a frontal or almost frontal facial poses. As it can be seen in figure 1, subjects on the ORL database, presents only smiling / not smiling facial expressions and images with smooth lighting variations. On the other hand, the YALE database presents subjects with a number of different configurations such as center / left / right lighting, with / without glasses, and happy / normal / sleepy / surprised / wink expressions. Some examples are given in figure 2.

In both cases, few subjects and few images per subject are provided. Thus, they represent a problem which can fit some practical situations such as access control systems with few authorized persons, as for these systems it may be easier to control lightning and to obtain good images in terms of pose and expression, as subjects are willing to get recognized. However, they are impossible to use in problems such as gender or ethnicity classification, where big pools of samples are necessary in order to obtain reliable results.



Fig. 1. Some examples of the ORL database



Fig. 2. Some examples of the YALE database

However, more powerful and complex situations involve huge security systems installed in airports or public buildings. These systems face uncontrolled lightning conditions, non collaborative subjects, and vast pools of identities. New databases incorporate some or all of these characteristics in order to test system against such situations. For example, the YALEb database (Georghiades et al. 2001) (Lee et al. 2005); examples in figure 3, contains images coming only from 10 persons, but each seen under 576 viewing conditions coming from combinations of 9 poses and 64 illumination conditions.



Fig. 3. Some examples of the YALEb database.



Fig. 4. Some examples of the FERET database

The FERET database (Phillips et al. 2000) consist of images collected in a semi-controlled environment, from 1199 subjects, and for different facial poses. Some examples can be seen in figure 4. An interesting property of this database is that it provides an extensive ground truth data specifying coordinates of facial organs, ethnicity, gender, and facial characteristics such as moustache, beard, and glasses information. Therefore, the FERET database may be easily used in almost any experimental situation. In fact, we will be using it for further experiments in this chapter.

The Face Recognition Grand Challenge (FRGC) database (Phillips et al. 2005) is another complete database. As FERET, the FRGC database consist of high resolution images from a pool of more than 1000 subjects and complete ground truth information files. However, this database provides images of full body, taken in different scenarios, which implies important changes in background and lightning conditions. Some examples can be seen in figure 5.



Fig. 5. Some examples of the FRGC database

In short, when testing a system it is important to keep in mind what type of database is been used. Using different databases provides different conditions, which allows us to test the system against different problems.

2.2 State of the art

A priori, techniques used for face identification or verification can also be used for gender identification. Finding inspiration in the biological system, S.L. Phung and A. Bouzerdoun proposed a system implementing a pyramidal neural network (Phung & Bouzerdoun 2007). This structure combines 1D and 2D neural network architectures with a resilient backpropagation learning algorithm, in such a way that some interesting properties arise. For example, neurons from the first layer are directly connected to image pixels, and the net's structure implements local receptive fields that are slightly overlapped. These two properties are somehow similar to the human eye. Using a set of 1152 male and 610 female images from the FERET database was used to test the system, with which a best classification rate of 89.8% was obtained.

On the other hand, B. Moghaddam and Ming-Hsuan Yang asserted that the Support Vector Machine (SVM) pattern classification outperforms traditional classifiers such as linear, quadratic, nearest neighbor, and Fisher linear discriminant, as well as more modern techniques such as Radial Basis Function (RBF) and large ensemble-RBF neural networks (Moghaddam & Ming-Hsuan 2002). The authors used a set of 1044 male and 711 female images from the FERET database for the experiments, and obtained a lowest error rate of 3.38% using a Gaussian RBF kernel on their SVM.

For the characterization of images, A. Jain et al. combined the Independent Component Analysis (ICA) feature extract technique with the SVM classifier (Jain & Huang 2004). They

tested the system using a set of 250 male and 250 female images from the FERET database, obtaining a classification rate of 95.67%. Then, Zhen-Hua Wang et al. applied a Genetic Algorithm (GA) search over the feature found by ICA, improving the system performance on a 7.5% (Zhen-Hua & Zhin-Chun 2009). Moreover, we have shown in (del Pozo Baños et al. 2011) that the ICA approach named Join Approximate Diagonalization of Eigenmatrices (JADE-ICA) outperforms the fast-ICA method in both error rate and stability on the gender classification problem.

Another interesting point is which face area provides the best information for gender classification. M. Castrillón Santana and Q.C. Vuong presented a psychological study on this aspect (Castrillon-Santana & Vuong 2007). They showed that when humans have no face information, the neck of males and the long hair of females provide the most diagnostic information. Moreover, in order to compare human and artificial systems they performed a series of experiments using different face masks. The system based on Incremental Principal Component Analysis (IPCA) and support vector machine (SVM) performed surprisingly similar using only face information (no neck and no hair) and face with hair line information. In a similar approach, Jing-Ming Guo et al. proposed the use of a mask to remove those pixels that are not discriminative as they are common for both classes or come from the background noise (Jing-Ming et al. 2010). This mask was based on the difference between the mean male image and the mean female image. Pixels selected by the mask were then used as inputs to a SVM classifier. Experiments were performed using a set of 1713 male and 1009 female images from the FERET database, and an accuracy of 88.89% was reported. J.R. Lyle et al. studied the validity of periocular images (area around eyes) for gender and ethnicity classification (Lyle et al. 2010). Images were rescaled to 251x251 pixels, converted to gray scale and their histograms equalized. The parameterization relied on the Local Binary Pattern (LBP) (Tapi 2003) tool, and a SVM was applied for classification. Testing the system on the FERET database, the authors obtained a best accuracy of around 94%.

A more sophisticated system which performs score fusion of experts on different face areas is presented by F. Manesh et al. (Manesh et al. 2010). First, eyes and mouth coordinates are automatically extracted with the extended Active Shape Model (Milborrow & Nicolls 2008). The system aligns, crops, and rescaled face images to 80x85 pixels as a preprocessing stage. Faces are then divided in 16 regions based on a modification of the Golden ratio template proposed by K. Anderson et al. (Anderson & McOwan 2004). Each region has its own expert system. These experts use a family of Gabor filters (Gabor 1946) (Daugman 1980) with 5 scales and 8 orientations as a feature extractor method, and a SVM with a RBF kernel for classification. Score fusion is finally performed using the optimum data fusion rule, which weights the experts accordingly to their accuracy. For the experiments, a combination of 891 frontal images from the FERET database and 800 frontal images from the CAS-PEAL data base was used. This set was divided in 3 sub-sets labeled "training", "validation", and "test". Finally, the researchers reported an accuracy of 96% for the ethnicity problem (Asian vs. Non-Asian), and 94% for gender classification fusing the scores of eyes, nose, and mouth. S. Gutta et al. also highlighted how information such as gender, ethnicity or face pose can increase the performance of face identification systems (Butta et al. 2000). To automatically obtain this information from facial images, they proposed a mixture of experts' system, which uses the "divide-and-conquer" modularity principle. Therefore, the system is composed of several sub-systems or modules and it elaborates the final result based on the individual results. In particular, an architecture combining ensemble-RBF networks and

decision trees techniques was used for gender classification. Using a set of 1906 male and 1100 female images from the FERET database the authors obtained a gender recognition rate of 96%.

As in any other face identification system, the preprocessing step is crucial. E. Makinen and R. Raisamo performed a set of experiments to evaluate the effect of face alignment (Makinen & Raisamo 2008). They reported no improvement when automatic face alignment techniques were used. However, manual alignment did increase the systems performance by a small factor. Giving these results, authors concluded that alignment methods must be improved in order to be of some use in the gender recognition problem. As they tested different classification techniques, they obtained the best performance with the SVM classifier, a classification rate of 84.39% using a set of 411 images from the FERET database. However, Adaboost with haar-like features offered very close results, while it was faster and more resistant to the in-plane rotation variations. Moreover, Jian-Gand Wang et al. also reported no significant improvement in terms of performance between manual, automatic, and none face alignment (Jian-Gan et al. 2010). Surprisingly, not only face alignment have none or little effect on gender classification, but many works has reported the same affect between low and high resolution images (Moghaddam & Ming-Hsuan 2002) (Lyle et al. 2010). In addition, many authors have reported no significant changes in performance when different image qualities were used (Moghaddam & Ming-Hsuan 2002) (Makinen & Raisamo 2008).

As we have experienced the same effect when quite different preprocessing methods were used (del Pozo-Baños et al. 2010), we decided to run here a further experiment using a common database to reinforce these results.

3. The proposed system model

The system used in this study has a block diagram composed of three main blocks: preprocessing, parameterization, and classification. Four quite different components were implemented for the preprocessing block, while two tools were used on the parameterization block. Figure 6 shows the aspect of this architecture, where only one preprocessing and one parameterization can be active at the same time.

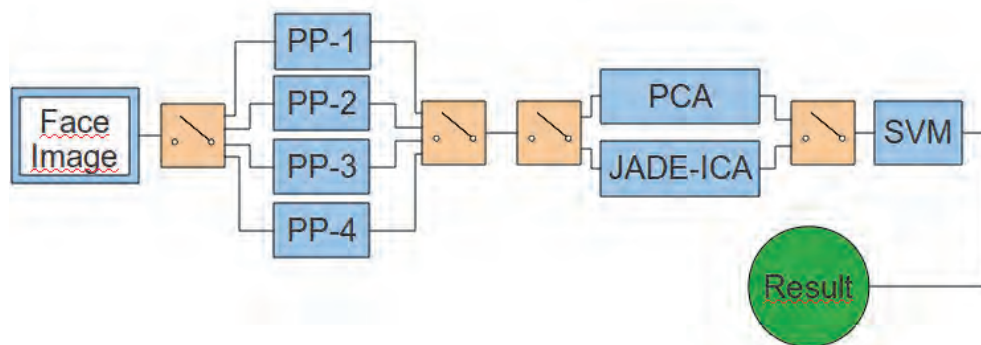


Fig. 6. Block diagram of the implemented system

3.1 Preprocessing methods

The first block at the system's entrance is the preprocessing block. This gets samples ready for the forthcoming blocks, reducing the noise and even transforming the original signal in a more readable one. Four different preprocessing components has been implemented.

- **PP-1.** This block normalizes the image histogram to a linear distribution before reducing its dimension to 15x20 pixels. Finally, an unsharpened filter is used to reduce the noise produced by the extreme reduction.
- **PP-2.** In this case, after histogram normalization a further local normalization (Xiong 2005) is performed. This normalization aims to reduce lighting effect through a double Gaussian filtering. Then, images are reduced to 15x20 pixels, and the unsharpened filter is applied.
- **PP-3.** The LBP (Topi 2003) is an invariant texture measure tool for gray scale images. When applied, it produces a matrix LBP where each point (x_c, y_c) corresponds to the differences between the centre pixel point (g_c) and its neighbours according to a given mask (g_m) . The mathematical definition is:

$$LBP(x_c, y_c) = \sum_{m=0}^{M-1} S(g_m - g_c) \times 2^m, S(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (1)$$

Here, the factor power of two makes the result of every possible combination unique, so that the LBP transformation is reversible. After applying the LBP, the PP-2 component reduces the resulting matrix dimension to 15x20, and applies the unsharpened filter.

- **PP-4.** This component is similar to the previous one, although in this case the image is first reduced to 15x20 and the filtered before apply the LBP transformation.

At the end of every preprocessing component, an elliptical mask is applied to remove peripheral noise located on corners. Images are then vectorized considering only pixels falling within the elliptical mask, which provides further reduction of samples dimensions. The effects of applying each preprocessing component can be seen in figure 7.

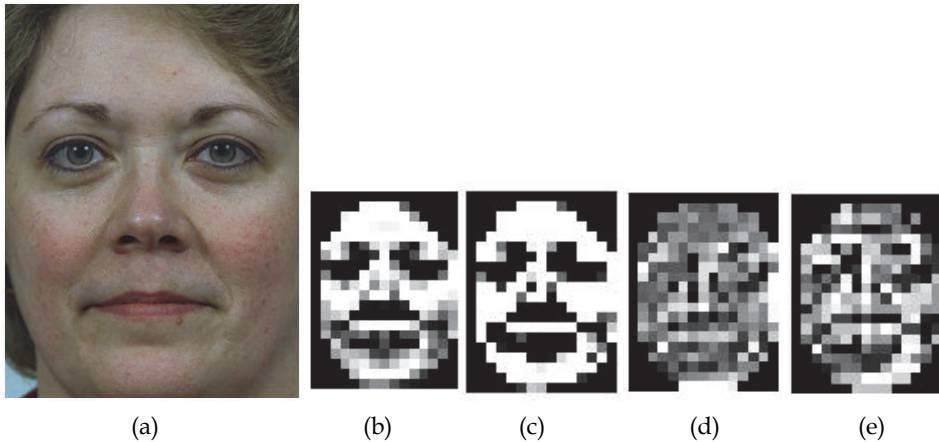


Fig. 7. Original image (A) and the resulting images for each preprocessing component: PP-1 (B), PP-2 (C), PP-3 (D), and PP-4 (E).

3.2 Parameterization techniques

The parameterization step analyzes samples and extracts relevant information. The proposed system uses both PCA and ICA appearance based methods.

3.2.1 Principal Component Analysis (PCA)

The PCA was introduced by Karl Pearson in 1901 (Jolliffe 2002), and then applied to face images by Kohonen (Kohonen 1989), Kirby y Sirovich (Kirby & Sirovich 1990). It was intended to extract information not viewable at first sight by projecting samples to a new space which maximizes variance. Moreover, by keeping only the first N coordinates of the new space, also called principal components (PCs) the system reduces sample dimensions keeping the most valuable information. Let X be a matrix of vectors x_i , each with p variables. PCA results in a set of projecting vectors α_i such that the transformation:

$$z_i = \alpha_i X = \sum_{j=1}^p \alpha_{ij} x_j \quad (2)$$

obtains a new set of vectors z_i representing the original x_i in a space maximizing its variance. Moreover, vectors α_i are uncorrelated to each other, so that new vectors appear in decreasing variance value order. By keeping the first N vectors z_i , the system remove redundant information and obtain an smaller representation of the data.

Projecting vectors α_i are computed by the eigenanalysis of the covariance matrix of X , referred to as S . Therefore, vector α_i corresponds to the i -th eigenvector of S , which when chosen to have unit length ($\alpha_i' \alpha_i = 1$) proves to provide a vector z_i with variance equal to the corresponding i -th eigenvalue of S .

3.2.2 Joint Approximate Diagonalization of Eigen-matrices Independent Component Analysis (JADE-ICA)

ICA is a particularization of PCA to extract components that are, at the same time, non-gaussian and statistically independent (Hyvärinen 2000). When used on images, ICA obtains independent base images which are not necessarily orthogonal. Application of these base images extracts between pixels information related to high order statistics.

In this study, an approach named JADE-ICA has been used to implement this tool. JADE-ICA is based on joint diagonalization of cumulant matrices. For simplicity, the case of symmetric distributions is considered, where the odd-order cumulants vanish. Let $X_1, \dots, X_4$ be random variables, and defined $X_i^* = X_i - E(X_i)$. The second order cumulants can be written as:

$$C(X_1, X_2) = E(X_1^*, X_2^*) \quad (3)$$

And the fourth-order cumulants as:

$$\begin{aligned} C(X_1, X_2, X_3, X_4) = & E(X_1^*, X_2^*, X_3^*, X_4^*) - E(X_1^*, X_2^*)E(X_3^*, X_4^*) - \\ & E(X_1^*, X_3^*)E(X_2^*, X_4^*) - E(X_1^*, X_4^*)E(X_2^*, X_3^*) \end{aligned} \quad (4)$$

In addition, the definitions of variance and kurtosis of a random variable X are:

$$\sigma^2 = C(X, X) = E(X^{*2}) \quad (5)$$

$$\text{kurt}(X) = C(X, X, X, X) = E(X^{*4}) - 3E^2(X^{*2})$$

Now, under a linear transformation $Y = AX$, the cumulants of fourth-order transformation became:

$$C(Y_i, Y_j, Y_k, Y_l) = \sum_{p,q,r,s} a_{ip} a_{jq} a_{kr} a_{ls} C(X_p, X_q, X_r, X_s) \quad (6)$$

, with a_{ij} the i-th row and j-th column entry of matrix A. Since the ICA model ($X = AS$) is linear, using the assumption of independence by $C(S_p, S_q, S_r, S_s) = kurt(S_p) \delta_{pqrs}$ where:

$$\delta_{pqrs} = \begin{cases} 1 & \text{if } p = q = r = s \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

and S has independent entries:

$$C(Y_i, Y_j, Y_k, Y_l) = \sum_{m=1}^n kurt(S_m) a_{im} a_{jm} a_{km} a_{lm} \quad (8)$$

, the cumulants of the ICA model are obtained.

Given any $n \times n$ matrix M and a random $n \times 1$ vector X, we consider a cumulant matrix $Q_x(M)$ defined by:

$$Q_x(M) = \sum_{m=1}^n C(X_i, X_j, X_k, X_l) M_{ki} \quad (9)$$

If X is centered, the definition of (4) shows that:

$$Q_x(M) = E\{(X^T M X^T) X X^T\} - R^X tr(M R^X) - R^X M R^X - R^X M^T R^X \quad (10)$$

, where $tr(B)$ denotes the trace of matrix B and $[R_x]_{ij} = C(X_i, X_j)$.

The structure of a cumulant $Q_x(M)$ in ICA model is easily deduced from (9) as:

$$Q_x(M) = A \Delta(M) A^T \quad (11)$$

With:

$$\Delta(M) = diag(kurt(S_1) a_1^T M a_1, \dots, kurt(S_n) a_n^T M a_n) \quad (12)$$

, where a_i is the i-th column of A.

Now, let W be a whitening matrix and $Z = WX$. Let us assume that the independent sources matrix S has unit variance, so that S is white. Thus $Z = WX = WAS$ is also white, and the matrix $U = WA$ is orthonormal. Similarly, the previous techniques can be applied into (13) for any $n \times n$ matrix M.

First, the whitening matrix W and the cumulant matrix Z are estimated. Then, the estimation of an orthonormal matrix U, denoted by U, is calculated. Therefore, an estimated matrix A denoted by A is obtained from $W^{-1}U$, and the sources matrix S is calculated by $A^{-1}X$.

To measure non-diagonality of a matrix B, $off(B)$ is defined as the sum of the squares of the non-diagonal elements:

$$\text{off}(B) = \sum_{i \neq j} (b_{ij})^2 \quad (13)$$

, where b_{ij} are elements of the matrix B . In particular $\text{off}(U^T Q_Z(M_i)U) = \text{off}\Delta_i = 0$ since $Q_Z(M_i) = U\Delta_i U^T$ and U is orthonormal. For any matrix set M and orthonormal matrix V , the joint diagonality criterion is defined as:

$$D_M(V) = \sum_{M_i \in M} \text{off}(V^T Q_Z(M_i)V) \quad (14)$$

, which measures diagonality far from the matrix V and bring the cumulants matrices from the set M .

3.3 Pattern classification

At this point, the system has retrieved and processed as much useful information from the input images as PCA or JADE-ICA can. Now, the classification component uses this information to take a decision on behalf the gender of the input face. To do so, this work uses the well known SVM (Schölkopf & Smola 2002).

The SVM is a structural risk minimization learning method of separating functions for pattern classification, that was derived from the statistical learning theory elaborated by Vapnik and Chervonenkis (Vapnik 1995). In other words, SVM is a tool able to differ between classes characterized by parameters, after a training process.

What makes this tool powerful is the way it handles non-linearly separable problems. In these cases, the SVM transforms the problem into a linearly separable one by projecting samples into a higher dimensional space. This is done using an operator called kernel, which in this study is set to be a Radial Basis Function (RBF). Then, efficient and fast linear techniques can be applied in the transformed space. This technique is usually known as the kernel trick, and was first introduced by Boser, Guyon y Vapnik in 1992 (Yan et al. 2004).

For simplicity, we configure the SVM to work as a verification system. In this particular case, the negative class (-1) corresponds to males and the positive class (1) to females. As a result, the classifier answers the "is this female?" question. The output of the SVM is a numeric value between -1 and 1 named score. A threshold has to be set to define a border between male (-1) and female (1) responses.

However, if all samples are used for training, there are no new samples for setting the threshold, and using the training samples for this purpose will lead to bad adjustments. Therefore, a 20 iterations hold-4-out (2 from each class) cross-validation procedure is used over the training samples to obtain 80 scores. These scores are then used to set the system's threshold to the equal error rate (EER) point, which is the point where False Acceptance Rate (FAR) and False Rejection Rate (FRR) coincide. The system's margin, defined as the distance of the closest point to the threshold line, is also measured. All these measures are referred to as validation measures.

When the threshold is finally set, the SVM is trained using all available samples. Because no big differences exist in the number of training samples used for this final training and the validation, we can expect the system to have a very similar threshold than that computed before.

In particular, the Least Squares Support Vector Machines (LS-SVM) implementation is used (Suykens 2002). Given a training set of N data points $\{y_i, x_i\}_{i=1}^N$, where x_i is the i -th input sample and y_i its corresponding produced output, we can assume that:

$$\begin{cases} w^T \phi(x_i) + b \geq 1 & \text{if } y_i = +1 \\ w^T \phi(x_i) + b \leq -1 & \text{if } y_i = -1 \end{cases} \quad (15)$$

where ϕ is the kernel function that maps samples into the higher dimensional space. The LS-SVM solves the classification problem:

$$\min L_2(w, b, e) = \frac{\mu}{2} w^T w + \frac{\zeta}{2} \sum_{i=1}^N e_{c,i}^2, \quad (16)$$

where μ and ζ are hyper-parameters related to the amount of regularization versus the sum square error. Moreover, the solution of this problem is subject to the constraints:

$$y_i [w^T \phi(x_i) + b] = 1 - e_{c,i}, \quad i = 1, \dots, N \quad (17)$$

3.4 System optimization

Because every preprocessing, parameterization, and classification technique may have its own optimal point in terms of configurable parameters, the system optimizes itself automatically using the validation results. Three parameters need to be optimized every time the system is trained: the number of kept principal/independent components for the parameterization component, and the regularization and kernel parameter for the LS-SVM. An exhaustive search is done along a configuration volume looking for the optimal point, defined as the point which provided a lower validation error rate and a larger validation system's margin.

4. Experiments and results

In order to obtain more reliable results, a 10-Folds cross-validation procedure was run on the experiments. Frontal facial images were taken from the FERET database, and cropped manually using ground information provided by the database. An example of this crop can be seen in figure 7. For each iteration the system was optimized and trained as it was explained in the previous section. Moreover, eight different systems, made out of every possible configuration between the four preprocessing components and the two parameterization tools were tested. All these facts made the experimental time impractical when the whole FERET database was used. To reduce this time a set of 1600 images (800 males and 800 females) were randomly selected.

Figures 8 and 9 show the results obtained when PCA and JADE-ICA were applied using all different preprocessing components. Numbers specified on the legend represent the areas under the curves. The EER points are given in table 1. In general terms, all preprocessing techniques provide similar behaviors, although differences were magnified when JADE-ICA was used. In particular, PP-1 and PP-2 performed almost identical.

A second experiment was run combining the scores obtained with each preprocessing technique. In particular, the sum and the product score fusion techniques were applied. The former combines the scores by a sum before apply the decision threshold. The later performs

a product after sifting the scores to range $[1 \ 3]$ instead of $[-1 \ 1]$, and then apply the threshold. The obtained results can be seen in table 1, and in figures 10 and 11.

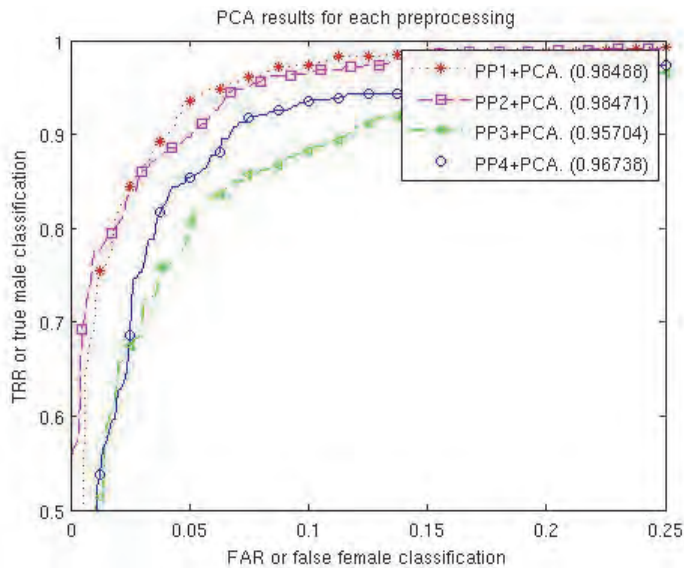


Fig. 8. Curves obtained for each preprocessing and PCA. In legend numbers represents the area under the curve.

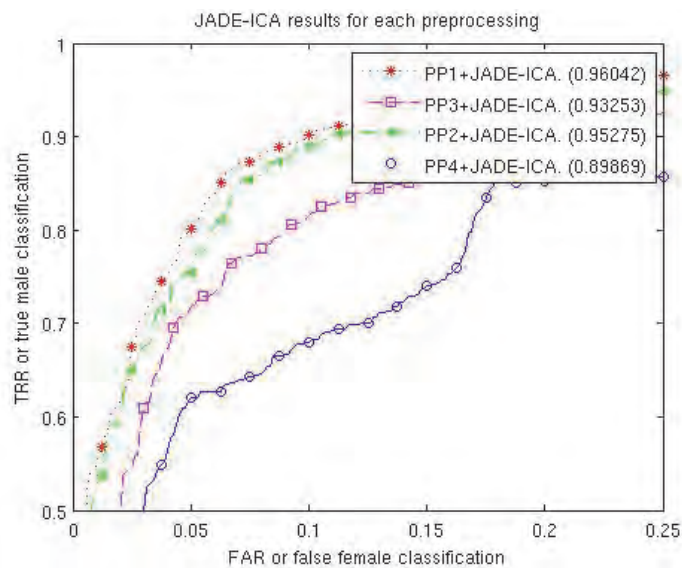


Fig. 9. Curves obtained for each preprocessing and JADE-ICA. In legend numbers represents the area under the curve.

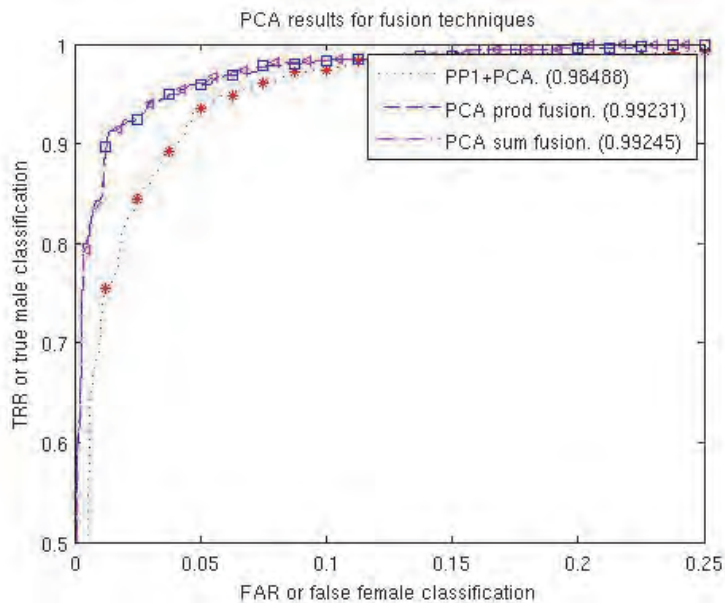


Fig. 10. Curves obtained for each score fusion of each preprocessing technique and PCA. In legend numbers represents the area under the curve.

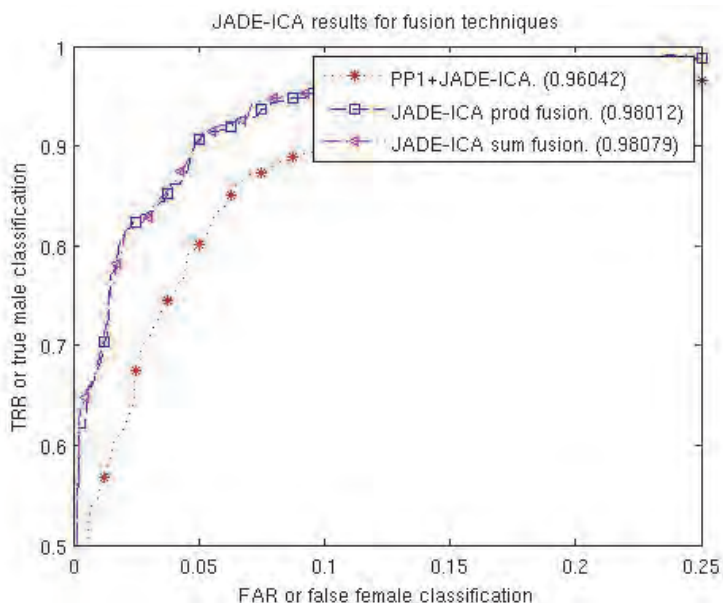


Fig. 11. Curves obtained for each score fusion of each preprocessing technique and PCA. In legend numbers represents the area under the curve.

Preprocessing	Parameterization	
	PCA	JADE-ICA
PP-1	5.54%	9.97%
PP-2	6.44%	10.54%
PP-3	10.83%	14.65%
PP-4	7.96%	16.50%
All preprocessing fusion		
Sum fusion	4.37%	6.99%
Prod fusion	4.44%	7.12%

Table 1. EER points for every combination of preprocessing and parameterization methods.

5. Conclusion

In this chapter, we have introduced the gender classification problem, from which a system automatically determines whether an input face corresponds to a female or a male. We have overview its characteristics as a bi-class problem and its relevance within the biometrics field. We have also introduced a biometric system with a simple architecture based on four preprocessing blocks, PCA and JADE-ICA parameterization, and an LS-SVM classifier. This system was used to test the variations of system's performance produced by wide changes on the preprocessing stage. The obtained results were consistent with other works, showing that in general there is little or no effect on the system's performance when these changes are applied.

Why do these big changes on the preprocessing stage provide similar results? Do they enhance different qualities of the facial images with similar level of discrimination? In a willing to through clarity on this intriguing characteristic of the gender recognition problem, we performed another experiment fusing scores obtained for each preprocessing technique. Both add- and product-fusion methods produced a small improvement in the system's behavior, of around 2% of reduction of EER comparing to the best preprocessing block.

This may suggests that all configurations are performing basing on the same or very alike information. Considering the massive differences between images resulting from each preprocessing block (figure 7), it is possible that this discriminant information is mostly related to very global and salient facial features, such as facial shape. This possibility is also consistent with the fact that image size does not affect gender classification performance. In fact, if facial shape is to be used, it does not matter whether the system has information coming from the inside of the face, or not.

6. Acknowledgment

This work has been partially supported by by "Cátedra Telefónica ULPGC 2009-10", and by the Spanish Government under funds from MCINN TEC2009-14123-C04-01.

7. References

- Du, P.; & Xiaoqing, D. (2008). The Application of Decision Tree in Gender Classification. Congress on Image and Signal Processing, vol.4, no., pp.657-660, 27-30 May 2008.

- Phung, S.L.; & Bouzerdoum, A. (2007). A Pyramidal Neural Network for Visual Pattern Recognition. *IEEE Transactions on Neural Networks*, vol.18, no.2, pp.329-343, March 2007.
- Schölkopf, B.; & Smola, A.J. (2002). *Learning with Kernels. Support Vector Machines, Regularization, Optimization, and Beyond*. Published by The MIT Press, 2002.
- Moghaddam, B.; & Ming-Hsuan, Y. (2002). Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.5, pp.707-711, May 2002.
- Hyvärinen, A. (2000). Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(4-5): 411-430, 2000.
- Jain, A.; & Huang, J. (2004). Integrating independent components and support vector machines for gender classification. *Proceedings of the 17th International Conference on Pattern Recognition*, vol.3, no., pp. 558- 561 Vol.3, 23-26 Aug. 2004.
- Zhen-Hua, W.; & Zhi-Chun, M. (2009). "Gender classification using selected independent-features based on Genetic Algorithm," *Machine Learning and Cybernetics*, 2009 International Conference on , vol.1, no., pp.394-398, 12-15 July 2009.
- del Pozo-Baños, M. ; Travieso, C.M. ; Alonso, J.B. ; & Ferrer, M.A. (2011) Gender Verification System based on JADE-ICA. *International Conference on Bio-inspired Systems and Signal Processing*, vol.7, no., pp. 570-576, ISSN: 978-989-8425-35-5, Rome, Italy, January 2011.
- Castrillon-Santana, M ; & Vuong, Q.C. (2007). An Analysis of Automatic Gender Classification. In *proceedings of 12th Iberoamerican conference on Progress in Pattern Recognition, Image Analysis and Applications (IbPRIA)*, pp.39-46, 2007.
- Jing-Ming, G.; Chen-Chi, L.; & Hoang-Son, N. (2010). Face gender recognition using improved appearance-based Average Face Difference and support vector machine. *2010 International Conference on System Science and Engineering (ICSSE)*, vol., no., pp.637-640, 1-3 July 2010.
- Lyle, J.R.; Miller, P.E.; Pundlik, S.J.; & Woodard, D.L. (2010). Soft biometric classification using periocular region features. *2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)*, vol., no., pp.1-7, 27-29 Sept. 2010.
- Topi, M. (2003) *The Local Binary Pattern Approach to Texture Analysis - Extensions and Applications*. Infotech Oulu and Department of Electrical and Information Engineering, University of Oulu, P.O.Box 4500, FIN-90014 University of Oulu, Finland. 2003.
- Manesh, F. Saei; Ghahramani, M.; & Tan, Y. P. (2010). Facial part displacement effect on template-based gender and ethnicity classification. *2010 11th International Conference on Control Automation Robotics & Vision (ICARCV)*, vol., no., pp.1644-1649, 7-10 Dec. 2010.
- Milborrow, S.; & Nicolls, F. (2008). Locating Facial Features with an Extended Active Shape Model. *Presented at the Proceedings of the 10th European Conference on Computer Vision: Part IV*, Marseille, France, 2008.
- Anderson, K.; & McOwan, P. W. (2004). Robust real-time face tracker for cluttered environments. *Computer Vision Image Understanding*, vol. 95, pp. 184- 200, 2004.
- Gabor, D. (1946). Theory of communication. *Journal of the Institute of Electrical Engineers*, 93, 429-457.

- Daugman, J. (1980). Two-dimensional analysis of cortical receptive field profiles. *Vision Research*, 20, 846–856.
- Gutta, S.; Huang, J.R.J.; Jonathon, P.; & Wechsler, H. (2000). Mixture of experts for classification of gender, ethnic origin, and pose of human faces. *IEEE Transactions on Neural Networks*, vol.11, no.4, pp.948-960, Jul 2000.
- Makinen, E.; Raisamo, R. (2008). Evaluation of Gender Classification Methods with Automatically Detected and Aligned Faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.30, no.3, pp.541-547, March 2008.
- Jian-Gang, W.; Hee, L.W.; Myint, Y.; Wei-Yun, Y. (2010). Real-time gender recognition with unaligned face images. *2010 the 5th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, vol., no., pp.376-380, 15-17 June 2010.
- del Pozo-Baños, M. ; Travieso, C.M. ; Alonso, J.B. ; & Ferrer, M.A. (2010) A Gender classifier Problem. 1er Workshop de Tecnologías Multibiométricas para la Identificación de Personas, vol.1, no., pp. 25-29, ISSN: 978-84-693-3389-1, Las Palmas de Gran Canaria, Spain, July 2010.
- Xiong, G. (2005). A Matlab implementation of the “localnormalize” function is available at <http://www.mathworks.com/matlabcentral/fileexchange/8303-local-normalization>. Last updated in 2005.
- Jolliffe, I.T.; (2002). Principal Component Analysis. 2nd ed. Springer Series in Statistics. 2002.
- T. Kohonen, (1989). Self-organization and Associative Memory, Springer-Verlag, Berlin, 1989.
- Kirby, M.; & Sirovich, L. (1990) Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, 1990.
- Vapnik, V. (1995) The Nature of Statistical learning Theory. Springer Verlag, New York, 1995.
- Yan, F.; Qiang, Y.; Ruixiang, S.; Dequan L.; Rong, Z.; Ling, C.X.; & Wen, G. (2004) Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry. *Bioinformatics*, Vol. 20, Issue 12, pp. 1948-1954, March 25, 2004.
- Georgiades, A.S.; Belhumeur, P.N.; & Kriegman, D.J.; 2001, From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, n. 6, pp: 643-660, 2001.
- Samaria, F.; & Harter, A. (1994). Parameterisation of a stochastic model for human face identification. *2nd IEEE Workshop on Applications of Computer Vision*. December 1994, Sarasota (Florida).
- Phillips, P. J.; Moon, H.; Rizvi, S. A.; & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
- Phillips, P. J.; Flynn, P. J.; Scruggs, T.; Bowyer, K. W.; Chang, J.; Hoffman, K.; Marques, J.; Min, J.; & Worek. (2005) Overview of the Face Recognition Grand Challenge. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- Suykens, J.A.K.; Van Gestel, T.; De Brabanter, J.; De Moor, B.; & Vandewalle, J. (2002) Least Squares Support Vector Machines. World Scientific, ISBN 981-238-151-1, Singapore, 2002.