

Measuring External Face Appearance for Face Classification

David Masip¹, Àgata Lapedriza² and Jordi Vitrià²

¹Universitat de Barcelona

²Universitat Autònoma de Barcelona
Spain

1. Introduction

Face classification can be defined as the problem of assigning a predefined label to an image or subpart of an image that contains one or more faces. This definition comprises many sub-disciplines in the visual pattern recognition field: (i) face detection, where the goal is to detect the presence of a face on an image, (ii) face recognition, where we assign an identifier label to the detected face, (iii) face verification, where the identity of the subject is given, and we should assure its truthfulness and (iv) gender recognition where the label *male* or *female* is assigned to each face image.

The information source of a facial image can be divided into two sets, depending on the zone of the face. The internal information is composed by the eyes, nose and mouth, while the external features are the regions of the hair, forehead, both laterals, ears, jaw line and chin.

Traditionally, face recognition algorithms have used only the internal information of face images for classification purposes since these features can be easily extracted. In fact, most of these algorithms use the aligned thumbnails as an input for some feature extraction process that yields a final feature set used to train the classifier. Classic examples of this approach are the *eigenfaces* technique (Turk & Pentland, 1991), or the use of Fisher Linear Discriminant Analysis (Hespanha Belhumeur & Kriegman, 1997). Moreover, in the face classification field, there are a lot of security-related applications where the reliability obtained by the internal features is essential: notice that the external information is more variable and easier to imitate. For this reason, the use of external features for these security-related tasks has often been ignored, given their changing nature. However, with the advances of technology in chip integration, small embedded computers are more integrated in our everyday life, favouring the appearance of new applications not directly related to security dealing with face classification, where the users do not make specific efforts to mislead the classifier. Typical examples are embedded camera-devices for human user-friendly interfaces, user profiling, or reactive marketing. In these cases we consider the external features as an extra source of information for improving the accuracies obtained using only internal features. Furthermore, notice that this consideration can be specially beneficial in natural and uncontrolled environments, where usually artefacts such as strong local illumination changes or partial occlusions difficult the classification task.

The use of external features has been seldom explored in computational face classification. Although there exists a plethora of methods to find the center pixel of each eye in order to put in correspondence each face image, the external regions are more difficult to align given that:

- External information does not have the same size in different persons. The hair volume can differ considerably between subjects. Pixel values at certain position do not mean the same depending on the sample.
- There is a lack of alignment on the features, given that there are no points of reference between samples from different subjects, or even between the same subject with different hairstyle.

In this context, the main motivation of this chapter is to provide a set of techniques that allow an efficient extraction of the external features of facial images. Commonly the extraction of internal information is faced using bottom-up techniques. In the case of external features, this strategy is not suitable due to the problems mentioned above. We propose a new algorithm to follow a top-down procedure to extract the external information of facial images, obtaining an aligned feature vector that can be directly used for training any standard pattern recognition classifier.

The chapter is organized as follows: in the next section we briefly review some psychological studies that support the usefulness of the external features in face classification in the normal human behaviour. Section 3 defines the feature extraction algorithm proposed for the external regions. Section 4 shows some results obtained in different face classification problems, using two publicly available face databases, and finally, section 5 concludes this chapter.

2. Motivation

In order to understand the human visual system's proficiency at the task of recognizing faces, different psychological experiments have been performed (Sinha & Poggio, 2002), (Fraser et al., 1990), (Haig, 1986), (Bruce et al. 1999), (Ellis, 1986), (Young et al., 1986). The results showed internal features to be more useful than external ones for recognition of familiar faces. However, the two feature sets reverse in importance as resolution decreases and also for recognition of unfamiliar faces.

Image resolution is an important factor to take into account when we try to characterize face recognition performance. Changes in the image information caused by increasing viewing distances, for instance, are very strong. See for example figure 1, where this fact is illustrated: comparing the internal features in low resolution of these faces we can see how difficult is to recognize them. However, when we add the external information the recognition task becomes easier.

Understanding recognition under such adverse conditions is of great interest given their prevalence in the real world applications. Notice that many automated vision systems need to have the ability to interpret degraded images, since in several cases they are acquired in low resolution due both to hardware limitations and large viewing distances. For this reason, Jarudi and Sinha (Jarudi & Sinha, 2003) performed an study with thirty subjects, ranging in age from 18 to 38. They were randomly placed in four non-overlapping groups corresponding to four experiments:

- Experiment A: recognition using the internal features of the face placed in a row.

- Experiment B: recognition using the internal features for each face in their correct spatial configuration.
- Experiment C: recognition using the external features alone with the internal features digitally erased.
- Experiment D: recognition using the full faces, including both internal and external features.



Figure 1. The low resolution problem

The mutual exclusion was enforced to prevent any transfer of information from one condition to another.

In each case, different images from famous people were presented sequentially, proceeding from the most degraded to the least degraded conditions. The results show that the full-face condition (D) is remarkably robust to reductions in image quality and declines gradually with increasing image blur, while performance in condition (A) is in general modest. However, when the internal features are placed in their correct spatial configuration (condition B), performance improves relative to condition A, but continues to be extremely sensitive to the amount of blur applied to the stimulus images. We can deduce then that the absence of external features damages the recognition performance in condition (B) relative to condition (D). Finally, the percentage of correct recognition in condition (C) is in general higher than in condition (A), and it is lower than in condition (B) only when the resolution is high enough.

There are more studies on the role of internal and external features that demonstrate the usefulness of the external information in face recognition. For instance, (Fraser et al., 1990) show that certain features are more important to face recognition than others. In particular, a feature hierarchy was observed with the head outline as the most significant, followed by the eyes, mouth and then nose.

Other works using different techniques have supported this general pattern of results suggesting that for the recognition of unfamiliar faces external features are more important than internal features (Bruce et al., 1999). A visual illusion shown in figure 2 justifies this hypothesis: internal face features in both portraits are exactly the same, but very few human observers are aware of this after an attentive inspection of the images if they are not warned of this fact. This is because the external features of these girls are contributing a priori more information than the internal features.

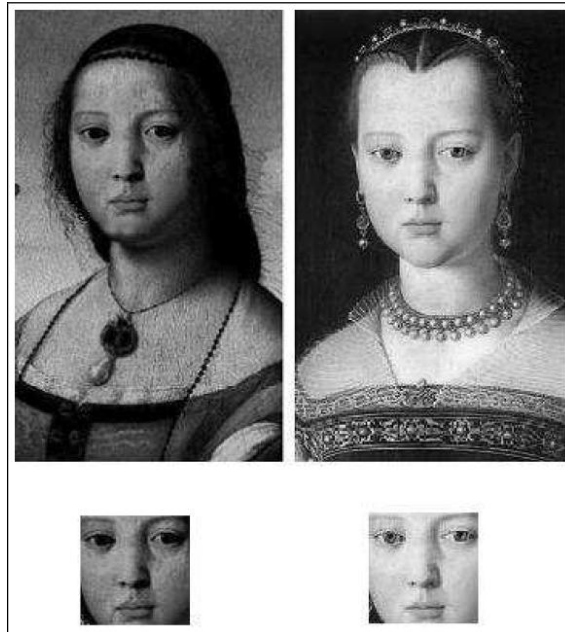


Figure 2. The Portraits illusion

The studies of (Jarudi & Sinha, 2003) suggest also that it is not the internal or external configuration on their own that serve recognition, but rather measurements corresponding to how internal features are placed relative to the external features. Thus, external features, even though poor indicators of identity on their own provide an important frame of reference for analyzing facial configuration. A visual illusion that was developed a few years ago serves to illustrate this idea: figure 2 shows what appears to be a picture of former US ex-President Bill Clinton and ex-Vice President Al Gore. Upon closer examination, it becomes apparent that Gore's internal features have been supplanted by Clinton's ones (in the configuration that they have on Clinton's face). If the appearance and mutual configuration of the internal features were the primary determinants of facial identity, then this illusion would have been much less effective. Then, it seems valid to conclude that

external features play a very significant role in judgments of identity. Furthermore, their contribution becomes evident only in concert with the internal features, because on their own, they do not permit reliable recognition. Thus, it seems valid to conclude that external features play a very significant role in judgments of identity, since their contribution becomes evident only in concert with the internal features in this case.

These psychological studies have motivated our interest for the usefulness of external features for automatic face classification.



Figure 3. The presidential illusion. These examples have been extracted from (Sinha & Poggio 1996) and (Sinha & Poggio 2002)

3. External Feature Extraction

The extraction of external information has two important drawbacks: the diverse nature and high variability of the external regions, and the lack of alignment of the external information. Therefore, most of the bottom-up approaches applied to internal feature extraction fail in obtaining an aligned D-dimensional feature vector from the external regions. Linear transformations such as PCA or FLD can not be directly applied. In this chapter we propose a top-down feature extraction algorithm that solves the alignment problems stated, by building a global fragment-based model of the external information (Lapedriza et al., 2005). The technique contains two main steps:

- Learning the model: Given the training examples, find an optimal set of fragments that constitute the model. This step is performed off line, and it is usually the most computationally intensive.
- Encoding of new images: Reconstruct a new unseen face image according to the fragments from the model that best fit with the image, and obtain a new aligned feature vector.

3.1 Learning the model

The proposed algorithm is based on previous works from the field of image segmentation (Borenstein & Ullman 2002) and (Borenstein, Sharon, & Ullman 2004). In the learning stage, the goal is to build a model of the external facial regions by selecting a subset of fragments from the external zone of the training images.

Image patches or fragments have been used as visual features in many recognition systems. These patches are normally extracted from images based on local properties such as similarity. A particularly useful set of features are the intermediate size image fragments, that arise in a natural way when searching for a set of features having as information as possible of the class (Ullman et al., 2002). Thus, intermediate sized image fragments are selected from a large set of candidates to represent a model of the class. This selected set is called here the *Building Blocks*.

The learning model algorithm receives as input a training set C with images that should contain clearly visible and diverse enough external features. From this training set, we generate all the possible sub images F_i of predefined sizes and store them. This step is computationally demanding, and can be optimized by selecting fragments only from specific zones of the image. Figure 4 shows the suggested regions of interest selected in this chapter. These surrounding regions can be easily isolated with the information provided at the previous face detection step.



Figure 4. Interest's regions of the external face features

Each fragment F_i is candidate to belong to the final model. Nevertheless, the size of the candidate set grows quadratically with the image dimensions, and the redundancy on the fragments is significant, being necessary a selection of the most representative fragments that account for the maximum diversity on the external regions.

Given the facial set C and a large set of non face images \bar{C} the goal is to find the fragments that are more representative of the external information of faces. The global selection criterion applied is to add to the model those fragments that can be found with high probability in face images but with low probability in non face images. The cross-correlation measure is used to determine whether a given Fragment F_i is similar to any part p from image I , and is defined as:

$$NCC(p, F_i) = \frac{\frac{1}{N} \sum_{x,y} (p(x,y) - \bar{p})(F_i(x,y) - \bar{F}_i)}{\sigma_p \sigma_{F_i}}$$

where N is the number of pixels in F_i , \bar{p} and \bar{F}_i are the means of p and F_i respectively, and σ_p and σ_{F_i} are their standard deviations.

For each fragment F_i , we compute the maximum values of the normalized cross-correlation between F_i and each possible sub image \bar{p} of $I \in C(NCC_i(C))$, and in $\bar{C}(NCC_i(\bar{C}))$.

Given a number of false positives α that can be tolerated for a fragment in \bar{C} we can compute a threshold value θ_i in order to assure that $P(NCC_i(\bar{C})) < \alpha$. This value can be used for determining whether a given fragment is present in an unseen image. Correlations below this threshold mean that not enough similarity has been found between the fragment and the image.

Finally, the K fragments with highest $P(NCC_i(C)) > \alpha$ are selected. These fragments have the highest probability to appear in the face set, and not to appear in the non face set (Borenstein & Ullman 2002, Sali & Ullman 1999). The complete algorithm is detailed in Table 1.

To ensure additional diversity on the fragment set we impose a geometrical constraint on the location of the fragments. The face image is divided in 3 independent regions: frontal part, left side and right side (see figure 4). The fragment selection process is run independently on each region, avoiding an important fragment concentration on small zones that would yield poor global reconstruction of new unseen images.

The algorithm takes as input:

- The face images set C
 - The set \bar{C} of non face images
 - The possible sizes of the fragments to analyze $S_i \in \{S_1, \dots, S_s\}$
 - The maximum number of fragments K that will be considered as building blocks.
 - The predefined threshold of false positives α .
1. For each fragment size S_i
 - Extract all the possible sub images F_i of size S_i from the set C using a sliding window procedure.
 - Add each sub image to the candidate fragments set.
 - Calculate and store the normalized correlation between each candidate fragment S_i and each image from C and \bar{C} .
 2. Compute the threshold θ_i for each fragment F_i that allows at most an α false positive ratio from the training set, $(P(NCC_i(\bar{C})) > \theta_i) < \alpha$.
 1. Compute the probability (frequency) of each fragment to describe elements from class C using the threshold θ_i , $(P(NCC_i(C)) > \theta_i)$.
 2. Select the K fragments with highest value $P(NCC_i(C)) > \theta_i$.

Table 1. Building blocks learning algorithm

3.2 Encoding the aligned External Information from new unseen images

Provided the learned fragment-based model of the external features, and supposing that the face detector has located a face in an image (internal features), the first step to extract the external face information is to cover the surrounding of the face area with the set of building blocks. To achieve this goal a function $NC(I, F_i)$ is defined as the pixel coordinates where the maximum $NCC(p, F_i)$ for all the possible sub images $p \in I$ is reached. Therefore, for each building block the place where the normalized cross-correlation value is maximum is computed, and then, the most appropriated covering is defined as an additive composition of the fragments that yields an optimal reconstruction of the surroundings of the detected internal face features.

The main goal of the proposed technique is to obtain an aligned feature vector of the external information, which can be used as an input for any traditional classifier designed for learning the internal features. The following three steps perform the feature extraction:

1. Given a new unseen image \mathbf{x} , we compute the normalized correlations between each fragment composing the model and the area of the image that surrounds a face. We store also the position of the maximum correlation $NC(I, F_i)$ for each fragment.
2. Using the optimal position for each fragment, a set of basis vectors \mathbf{B} are constructed as follows: for each fragment an image of the same size as the original image is generated with the fragment set at the optimal position (obtained in the first step), and the rest of the pixels set to 0.
3. Given \mathbf{B} , we find the coefficients \mathbf{S} that best approximate the linear transform:

$$x \approx BS$$

To compute the set of coefficients \mathbf{S} we use the Non Negative Matrix Factorization (NMF) algorithm (Lee & Seung, 1999). The NMF algorithm fulfils the three constraints inherent to this problem:

- The combination of coefficients must be additive, given that each fragment contributes to the reconstruction of the external features.
- The reconstruction error of the image external features using the fragments of the model must be minimized, and the NMF minimizes the reconstruction error (Lee et al., 2000) in the mean squared error sense.
- The fragment set is diverse, given that is extracted from different subjects in order to model the variability of the problem. Therefore, only a small part of the fragments from the general model can be useful to reconstruct a specific face. This fact implies that the coefficients \bar{S} must be sparse, and only a small part of the fragments of the model should be activated for each face.

There exist several implementations of the NMF algorithm, among them we have chosen the version developed by Patrick Hoyer (Hoyer 2004), fixing the bases matrix \bar{B} to constraint the NMF to our model (Lapedriza et al., 2006). This implementation has the advantage that the sparseness coefficient can be adjusted, in order to allow or restrict the amount of fragments that take part on the reconstruction. Figure 5 shows an example of the whole feature extraction process. We use only the external regions marked to be put in correspondence with the fragment model. The reconstructed image is expressed as a linear combination between the fragments placed at the optimal position and the NMF weights. Notice that the reconstruction is not used for classification purposes. The NMF coefficients

encode the aligned feature vector that represents the external information of the face. Each feature represents the contribution of one fragment from the model in the reconstruction of the original external region.

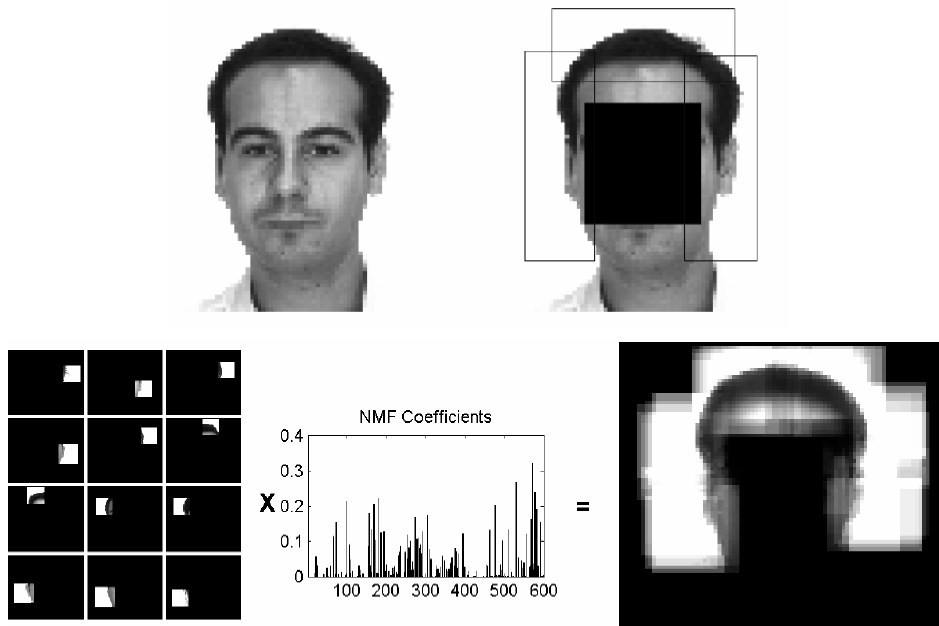


Figure 5. Example of the reconstruction process of the external information. In the first sample the original image is plotted, and the regions where the external information is marked. Some of the fragments from the model are plotted at the optimal position under the NC criterion, and the feature extraction process is illustrated as a linear combination of this fragment-basis and a set of NMF coefficients. The resulting reconstructed image is shown.

4. Face Classification Experiments

To test this external feature extraction system we have performed different experiments, including gender, verification and subject recognition, using different classifiers. In this section we present some results showing that the proposed method allows to obtain significant information from the external zones of the face.

Two publicly available databases have been used: the AR Face Database and FRGC (Face Recognition Grand Challenge, <http://www.bee-biometrics.org/>).

The AR Face Database is composed of 26 face images from 126 different subjects (70 men and 56 women). The images have uniform white background. The database has from each person 2 sets of images, acquired in two different sessions, with the following structure: 1 sample of neutral frontal images, 3 samples with strong changes in the illumination, 2 samples with occlusions (scarf and glasses), 4 images combining occlusions and illumination changes, and 3 samples with gesture effects. One example of each type is plotted in figure 6.

From the FRGC we have used the set of still high resolution images, which consists of facial images with 250 pixels between the centers of the eyes on average. This set includes 3772 images from 275 different subjects. There are from 4 to 32 images per subject. In our experiments done using this database, we have excluded 277 images where the external features were partially occluded. Figure 7 includes some examples of images from the FRGC and figure 8 shows some of the excluded images.

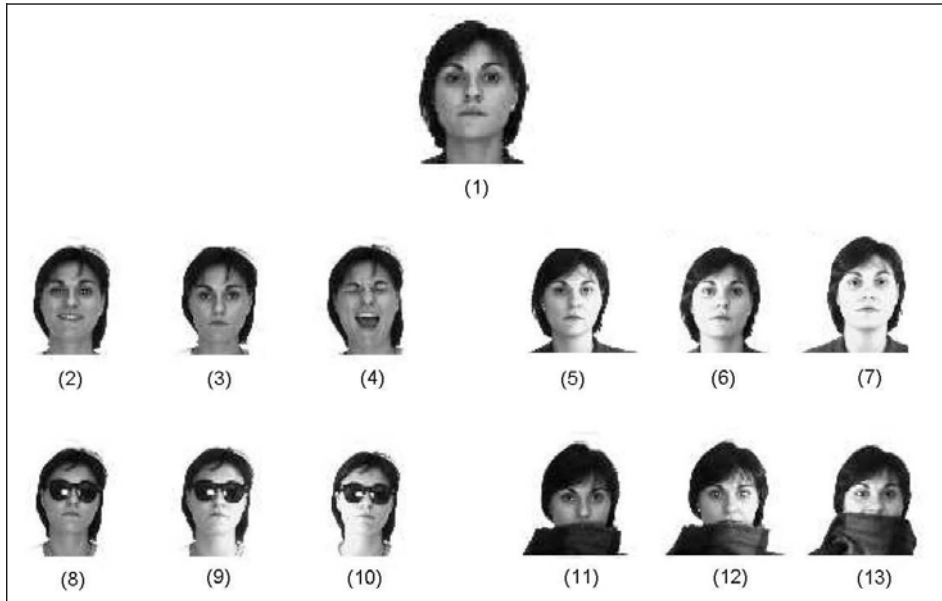


Figure 6. One sample from each of the image types in AR Face Database. The image types are the following: (1) Neutral expression, (2) Smile, (3) Anger, (4) Scream, (5) left light on, (6) right light on, (7) all side lights on, (8) wearing sun glasses, (9) wearing sunglasses and left light on, (10) wearing sun glasses and right light on, (11) wearing scarf, (12) wearing scarf and left light on, (13) wearing scarf and right light on.

In these experiments we need to work always in the same resolution, since the proposed external face features extraction system is a fragment-based method. For this reason, all the images in both databases have been resized to ensure that the between eyes distance is 16 pixels.

To evaluate the results in a reliable way we normally use a k-fold cross validation system. Moreover we compute also the radius of the confidence interval as follows:

$$r = \frac{1.96\sigma}{\sqrt{k}}$$

where σ is the standard deviation of the obtained results by these k performances. Then, the confidence interval is

$$I = [m-r, m+r]$$

where m is the mean of the results obtained by the experiments.

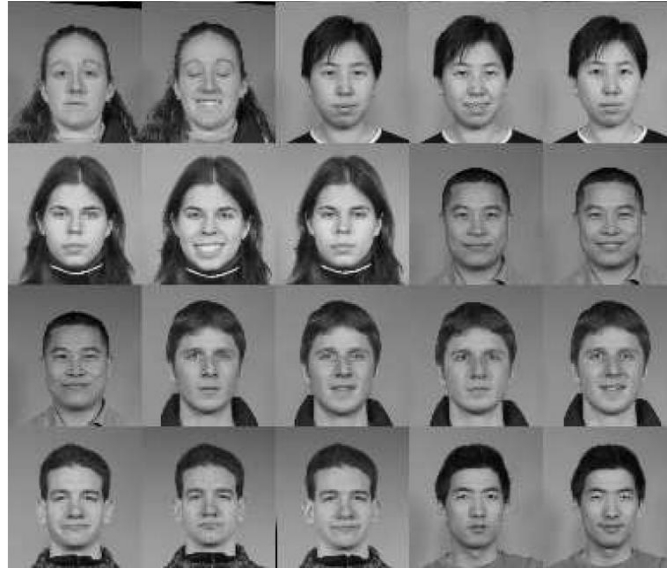


Figure 7. Examples of images in the FRGC Database



Figure 8. Some examples of excluded images

4.1 External Features' model construction

For all the experiments we need to construct a model for the external information to extract these features according to the method presented above. We have randomly selected some subjects from each database to construct the corresponding building blocks set. Notice that these databases have different illumination conditions and for this reason we construct two separated models, one per database.

The subjects considered to construct these Building Blocks are not considered in any classification experiments to ensure that the reconstruction of an image never takes use of fragments extracted from itself or from the same person.

The construction of the Building Blocks' set follows the same protocol in both cases:

- We use 80 face images from the database to extract the fragments, 40 male and 40 female.
- We have automatically extracted 24 fragments of each image to construct the set of candidate fragments and run the selection algorithm explained in section 3.1, using the parameters: $\alpha = 0.1$ and $K=200$.
- A hundred of natural images (with no faces) extracted from the web have been selected for the \bar{C} set.

4.2 Testing the external information

Here we present some results to show that the external face features contribute notably in face classification purposes.

First we perform Gender recognition experiments, where the data set has been split in a training set containing the 90% of the samples and a test set with the remaining 10% from the FRGC Database. The presence of male and female samples on each set has been balanced.

We have performed 50 iterations of the NMF algorithm.

In these experiments we have used 5 different classifiers: Maximum Entropy, Support Vector Machines, Nearest Neighbour, Linear and Quadratic classifiers, and the accuracies have been computed in all cases as the mean of 100 repetitions (using a cross-validation strategy). The mean results of the 5 classifiers are shown in table 2.

Algorithm	ME	SVM	NN	Linear	Quadratic
Accuracy	83.24	94.19	92.83	88.75	88.32
Confidence Interval	0.43	0.27	0.26	0.37	0.38

Table 2. Gender recognition experiment (FRGC)

To ensure the relevance of these external features we also perform two subject recognition experiments. In this case we have considered the same set of 2640 images as in the experiment described above. However, for the subject recognition experiment the data set has been organized as follows: a training set containing 10 randomly selected images from each subject and a test set with the remaining images.

We have performed also a discriminant feature extraction on the encoded external information (NMF coefficients) and then have classified using the Nearest Neighbour (NN). The used discriminant feature extraction technique is based on the Adaboost algorithm (Masip & Vitrià, 2005) and (Masip & Vitrià, 2006).

The accuracy obtained with the NN classifier directly applied on the NMF coefficients was 43% while the best accuracy obtained using the NN classifier on the extracted features was obtained in dimensionality 315, having 56% of correctly classification. Note that these results are relevant taking into account that we consider more than 200 classes. Therefore, this shows that the external features contain enough relevant information for classification purposes.

4.3 Comparing external and internal features

To compare the contribution of external features and internal features in automatic face classification field, we perform gender recognition and subject verification experiments.

First, we have selected 2210 images from the AR Face database, discarding subjects with missing images and balancing the presence of male and females. The error rates shown in this work were obtained repeating 100 times the next experimental protocol: data has been randomly split in a training and a testing set, we have used 90% of the data for training and 10% for testing; the splitting has been performed taking into account the person identity, so all samples from the same person must be in only one set to avoid person recognition instead of gender recognition.

We perform the test using the maximum entropy classifier (ME).

We have performed the same gender classification experiments using only the internal features (1188 pixel values) and using only the external features (600 NMF coefficients) and the obtained rates are shown in table 3. The results obtained in the external case slightly better than the ones obtained with the internal features. This fact can be justified by the loss of information in the internal part of the face caused by the partial occlusions and the high local changes in the illumination (almost a half of the AR Face database images have occlusion artefacts, see figure 6).

	ME (Confidence Interval)
Internal Features	80.4 (0.63)
External Features	82.8 (0.57)

Table 3. Gender recognition using the AR Face Database

On the other hand we have performed different subject verification experiments using internal and external features separately. The selection of the different data sets used in this experiment is based on the Sep96 FERET testing protocol (Phillips et al., 1996). In this protocol, two sets are distinguished: a target set (T), composed by the known facial images, and the query set (Q), including the unknown facial images to be identified. Two subsets should be selected from each of these sets: a gallery $G \subset T$ and a probe set $P \subset Q$. After this selection, the performance of the system is characterized by two statistics. The first is the probability of accepting a correct identity. This is referred as the verification probability, denoted by P_V (also referred to as the hit rate in the signal detection literature). The second is the probability of incorrectly verifying a claim, that is called the false-alarm rate and is denoted by P_F .

We have used the LBDP (Masip & Vitrià, 2005) method to reduce the dimensionality of the data vectors, working after that in a 300-dimensionality vectorial space.

We perform 2 the verification experiments using both FRGC and AR Face databases. The details of the sets used in these experiments are specified in table 4. They have been chosen following the scheme of the Lausanne Protocol configuration-1 (Kang & Taylor, 2002) and there are two kinds of subjects: clients (known by the system) and impostors (unknown by the system).

Set	Subjects	N. Images	Total images
Training (G)	Client : 100	3	300
Testing (P)	Client: 100	2	200
Testing (P)	Impostor: 50	5	250

Table 4. Configuration of the Gallery and the Probe sets in the face verification experiments

The first experiment has been made using the images from the FRGC database. The second one has been performed with a subset of the ARFace database, including only images having partial occlusions and high local changes in the illumination. The results obtained are shown in figure 9 and figure 10 respectively. In the FRGC experiments, the internal information outperforms the external one. Nevertheless, as in previous experiments, in presence of occlusions and illumination artefacts (AR Face data set), the external information becomes more relevant.

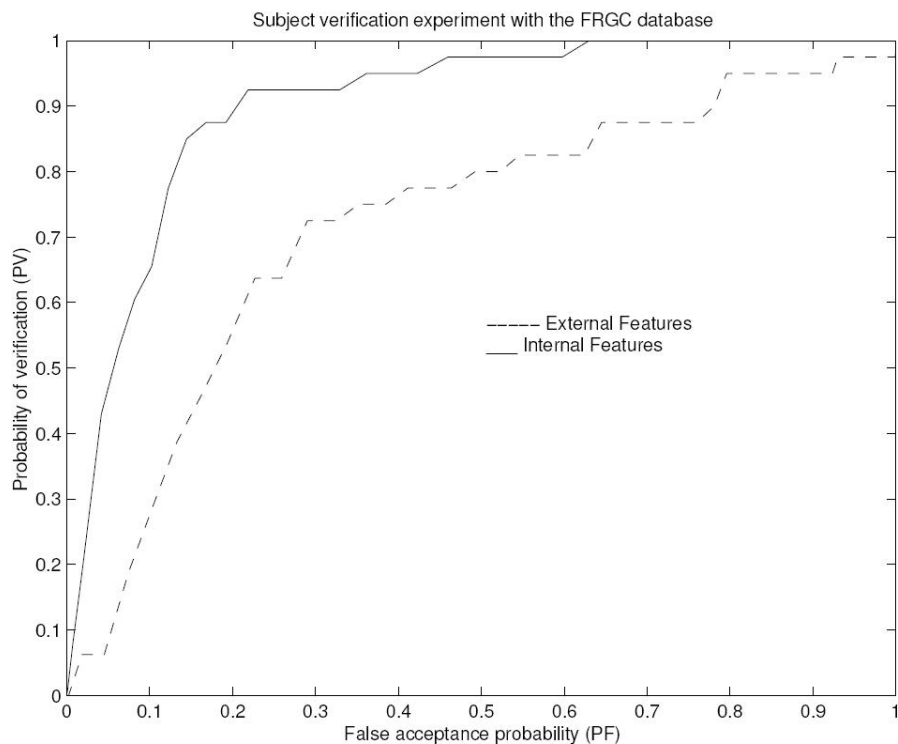


Figure 9. Subject verification using the FRGC Database

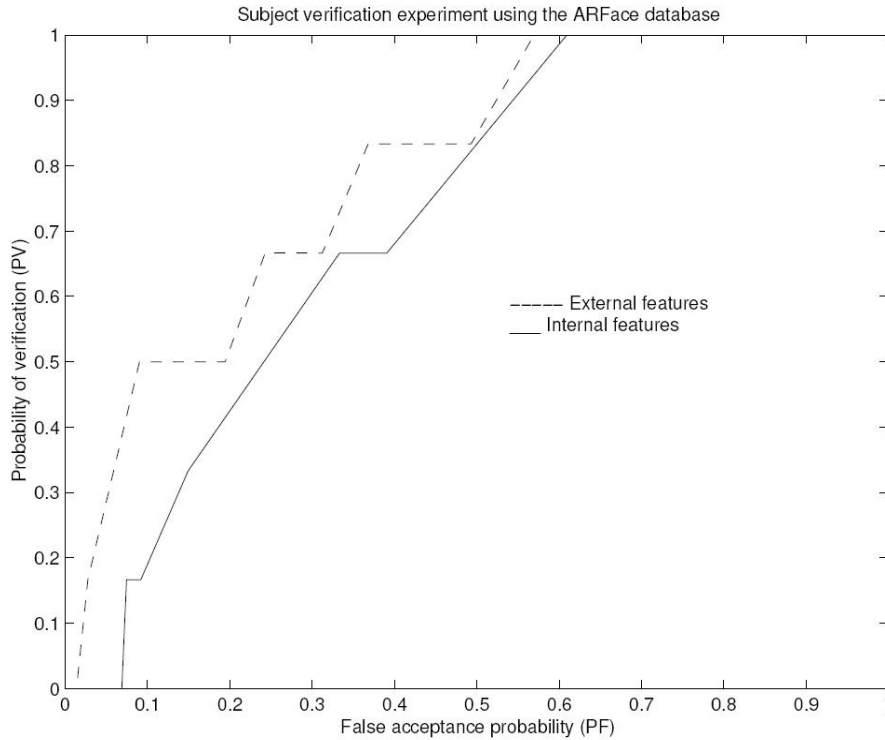


Figure 10. Subject verification using the AR Face Database

4.4 Combining external and internal information

As can be observed from the psychological results explained in section 2, the parallel use of the internal features and the external features is an important issue to be studied. However, to combine these sources of information is not a trivial task, given that the nature of these features is very different (notice that we consider the values of the pixels as internal features while the external ones are obtained using the presented fragment based method).

Nevertheless, we present here a first combination approach that consists in joining directly each information source and classifying the faces using this larger feature vector.

To appreciate the contribution of each feature set and the combination of external and internal information we perform gender classification experiments using the AR Face Database, and present the obtained rates in each image's type's set.

The error rates shown in this work were obtained repeating 100 times the next experimental protocol: (i) data have been randomly split in training and a testing set. There have been used 90% of the data for training and 10% for testing from each of the image's type's set ; (ii) the splitting has been performed taking into account the person identity, so all samples from the same person must be in only one set to avoid person recognition instead of gender recognition.

The results are detailed in figure 11. The best accuracy is marked with an '*', and the methods whose confidence intervals overlap with the best results are shown in boldface. Notice that the use of the combined feature vector (external and internal) obtains the best accuracies in almost all the cases, being the contribution of the external information more significant in presence of occlusion problems.














ME	 AR01	 AR02	 AR03	 AR04	 AR05	 AR06	 AR07	 AR08	 AR09	 AR10	 AR11	 AR12	 AR13
Int	83.7	84.4	85.6	82.8	84.3	92.3*	91.5*	87.5	88.3	89.8*	57.3	59.3	72.5
	±0.6	±0.6	±0.5	±0.6	±0.7	±0.5	±0.5	±0.4	±0.5	±0.4	±0.9	±1.1	±1.0
Ext	82.7	81.8	87.2	79.6	83.6	81.1	78.3	85.3	83.8	79.8	69.1	68.5	71.1
	±1.4	±1.2	±1.3	±1.2	±1.2	±1.1	±1.1	±1.6	±1.6	±1.6	±1.4	±1.8	±1.4
Comb	88.5*	86.7*	91.0*	85.4	86.8	87.3	85.3	89.9*	90.9*	88.7	72.1*	69.7	72.0
	±1.2	±2.2	±2.4	±2.8	±1.2	±2.2	±2.0	±2.1	±1.5	±1.5	±3.0	±1.9	±2.2
NN													
Int	82.8	81.6	80.4	82.1	85.1	89.6	89.0	86.1	87.6	88.1	67.5	69.9	70.6
	±1.9	±2.1	±1.9	±2.2	±1.9	±1.6	±1.8	±1.6	±1.6	±1.8	±2.7	±2.4	±2.4
Ext	64.5	66.2	65.5	62.9	66.4	65.6	64.1	68.9	65.8	68.0	57.1	58.1	63.1
	±2.4	±2.6	±2.5	±2.8	±2.6	±2.3	±2.3	±2.6	±2.4	±2.6	±2.3	±2.3	±2.3
Comb	85.3	85.4	83.7	85.5*	87.5*	90.9	91.0	87.9	89.7	89.2	71.4	74.9*	73.4*
	±1.9	±2.0	±2.4	±2.3	±1.9	±1.9	±1.8	±1.7	±1.6	±1.7	±2.3	±2.2	±2.4

Figure 11. Gender recognition using the AR Face Database

5. Conclusion

In this chapter we introduce the importance of the external features in face classification problems, and propose a methodology to extract the external features obtaining an aligned feature set. The extracted features can be used as input to any standard pattern recognition classifier, as the classic feature extraction approaches dealing with internal face regions in the literature. The resulting scheme follows a top-down segmentation approach to deal with the diversity inherent to the external regions of facial images.

The proposed technique is validated using two publicly available face databases in different face classification problems: gender recognition, face recognition and subject verification. In a first approach, we show that the external features encoded in the NMF coefficients yield enough useful information for classification purposes. Then we compare the information contributed by the external features and the internal features. Finally, the last step is to combine the information provided by the external and the internal features. We show that both kinds of information are complementary, providing an extra information cue that can improve the classification results in presence of occlusions and local changes in the illumination.

6. Future Work

The proposed method can be improved at three different levels: firstly the learning of the building blocks model could benefit from using some kind of normalization on the fragments generation. In particular, we propose the use of techniques of ridges and valleys detection to filter the images as a previous step on the feature extraction. In a second level, we plan to improve the selection of the fragments that compose the building blocs by adding a diversity measure that could model a larger rank of hairstyles. And in a third stage, we need to define a more robust combination rule of the internal and external

information. The use of ensembles of classifiers seems to be a natural continuation of this combination. For instance, the Adaboost (Freund & Schapire, 1996) algorithm can be studied for this purpose.

7. References

- Borenstein, E. & Ullman, S. (2002). Class-specific, top-down segmentation. *Proceedings of the 7th European Conference on Computer Vision- Part II*, pp 109-124, London, UK, 2002. Springer-Verlag.
- Borenstein, E.; Sharon, E. & Ullman, S. (2004) Combining top-down and bottom-up segmentation. *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, pp 46, Washington, DC, USA, 2004. IEEE Computer Society.
- Bruce, V.; Henderson, Z.; Greenwood, K.; Hancock, P. J. B.; Burton, A. M. & Miller, P. (1999) *J. Exp. Psychol. Applied* **5**, 339-360.
- Ellis, H.D. ; Shepherd, J.W. & Davies , G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, 8:431-439, 1979.
- Ellis, H. D. (1986). Face recall: A psychological perspective. *Human Learning*, 5, 1986. pp 189-196.
- Fraser, I.H.; Craig, G.L. & Parker; D.M. (1990) Reaction time measures of feature saliency in schematic faces. *Perception*, 19, 1990, pp 661-673.
- Freund, Y. & Schapire R. E. (1996). Experiments with a New Boosting Algorithm. *Proceedings of the International Conference on Machine Learning*. pp 148-156. Bari, July 1996. ISBN 1-55860-419-7, Morgan Kaufman.
- Haig, N.D. (1986). Exploring recognition with interchanged facial features. *Perception*, 15, 1986, pp 235-247.
- Hespanha, J.P; Belhumeur, P.N. & Kriegman, D.J. (1997). Eigenfaces vs fisherfaces: Recognition using class specific linear projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 7 ,July 1997, 711-720.
- Hoyer, P.O. (2004). Non-negative Matrix Factorization with sparseness constraints. *Journal of Machine Learning Research* 5 ,2004, pp 1457-1469.
- Jarudi, I. N. & Sinha, P. (2003). Relative Contributions of Internal and External Features to Face Recognition. *Technical Report 225. Massachusetts Institute of Technology, Cambridge, MA., March 2003.*
- Kang T.H. & Taylor C.J. (2002). A Comparison of Face Verification Algorithms using Appearance Models. *Proceedings of the 2nd Workshop on Biologically Motivated Computer Vision*. vol 2, pp 477-486. Tübingen, Germany. November 2002. Springer Verlag.
- Lapedriza, A.; Masip, D. & Vitrià J. (2005). Are External Face Features Useful for Automatic Face Classification?, *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*. pp 151-159. June 2005. 0-7695-2372-2-3. San Diego, USA. IEEE Computer Society. Washington, DC, USA.
- Lapedriza, A.; Masip, D. & Vitrià J. (2006). On the Use of External Face Features for Identity Verification, *Journal of Multimedia*. 1,4, July 2006, pp 11-20. ISSN : 1796-2048.

- Lee, D. & Seung, H. S. (1999). Learning the parts of objects with nonnegative matrix factorization, *Nature*, 401, July 1999, 788-791.
- Lee, D. & Seung, H. S. (2000). Algorithms for Non-negative Matrix Factorization, *Proceedings of Neural Information Processing Systems (NIPS)*, pp 556-562, (2000).
- Martinez A. & Benavente, R. (1998). The AR Face database, *Computer Vision Center, Tech. Rep. 24*, June 1998.
- Masip, D.; Kuncheva, L.I. & Vitrià, J. (2005). An ensemble-based method for linear feature extraction for two-class problems. *Pattern Analysis and Applications*, 8,3, pp 227-237 December 2005. Springer Verlag.
- Masip, D. & Vitrià, J. (2006). Boosted discriminant projection for nearest neighbor classification. *Pattern Recognition*, 39, 2, February 2006, 164-170.
- Peng W.; Matthew B.; Qiang Ji & Wayman J. (2005). Automatic Eye Detection and Its Validation, *Proceedings of IEEE Workshop on Face Recognition Grand Challenge Experiments (with CVPR)*, June 2005, San Diego, CA.
- Phillips P.J; Rauss, P.J & Der S.Z. (1996). FERET (Face Recognition Technology) Recognition Algorithm Development and Test Results. *Technical report 995. Army Research Lab.* October 1996.
- Sali E. & Ullman S. (1999). Combining Class-Specific Fragments for Object Classification. *Proceedings of the 10th British Machine Vision Conference (BMVC)*, 1, 1999.
- Sinha, P. & Poggio, T. I. (1996). Think I know that Face, *Nature*, 384, 6698, 1996, pp 404.
- Sinha, P. & Poggio, T. (2002) United we stand: The role of head-structure in face recognition. *Perception*, 31(1):133, 2002.
- Turk, M. & Pentland A. (1991). Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, 3, 1, March 1991, 71-86.
- Ullman, S.; Sali, E. & Vidal-Naquet, M. A. (2001). Fragment-Based Approach to Object Representation and Classification. *Proceedings of the 4th International Workshop on Visual Form*, pp 85-102. May 2001.
- Ullman, S.; Vidal-Naquet, M. & Sali E. (2002). Visual features of intermediate complexity and their use in classification, *Nature Neuroscience*, 5, 7, July 2002, 682-687.
- Young, A.W.; Hay, D.C.; McWeeny, K.H.; Flude, B.M. & Ellis, A.W. (1986). Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14, 1986, pp 737-746