

# On the Complementarity of Face Parts for Gender Recognition

Yasmina Andreu and Ramón A. Mollineda

Dept. Llenguatges i Sistemes Informàtics  
Universitat Jaume I. Castelló de la Plana, Spain.  
yandreu@uji.es, mollined@uji.es

**Abstract.** This paper evaluates the expected complementarity between the most prominent parts of the face for the gender recognition task. Given the image of a face, five important parts (right and left eyes, nose, mouth and chin) are extracted and represented as appearance-based data vectors. In addition, the full face and its internal rectangular region (excluding hair, ears and contour) are also coded. Several mixtures of classifiers based on (subsets of) these five single parts were designed using simple voting, weighted voting and other learner as combiners. Experiments using the FERET database prove that ensembles perform significantly better than plain classifiers based on single parts (as expected).

**Key words:** Gender recognition, Face image analysis, Classifier ensembles

## 1 Introduction

The face of a human presents an evident sexual dimorphism that makes gender recognition a straightforward cognitive process. According to experiments on faces where exogenous and cultural features such as hair and make-up have been removed, adults are able to correctly recognize gender in more than 96% of cases [1, 2]. In recent years, a number of papers have paid attention to the problem of automatic gender recognition from face image inspection [3–7]. Possible applications can be envisaged in access control, in demographic descriptions of a population for commercial or statistical purposes, and so forth.

Most of these papers have proposed new methodologies to carry out automatic gender recognition using descriptions of full faces [3–5]. Considering that the structural relationships between face parts contribute to discriminate between genders [1, 2], this holistic approach seems to be the most suitable choice when images of full faces are available. Recognition rates above 96% in controlled environments have been obtained in [4].

As another point of view, not based on holism, three recent papers [6–8] have studied the importance of face parts in gender recognition. Given a face image, these works extract a number of subimages containing relevant features of the face, mainly eyes, nose, mouth and chin. These features are transformed into vector representations and used in learning experiments. The differences

between these papers lie in the face parts extracted, the databases and the learning models used, and the scope of the experimental methodology; the most discriminant parts lead to gender recognition rates above 85%.

Why could it be interesting to evaluate the effectiveness of using face parts in gender recognition? Apart from its intrinsic scientific value, a practical answer could be related to gender classification in the presence of partial occlusions of the face, where the holistic representation of the full face might not be suitable. In such situations, the recognition of the gender could be supported by aggregating (a subset of) visible individual parts. When more than one part is observable, ensembles of classifiers based on the (expected) complementarity of these visible parts could model their joint contribution to the prediction of gender.

The goal of this paper is to find out the level of the (expected) complementarity of some parts of the face in discriminating gender. Several ensembles of classifiers are used to prove that the joint contribution of separate face parts has more discriminant capabilities than the plain classification of isolated parts. In particular, mixtures of classifiers based on 3 and 5 parts are designed, whose decisions are combined using simple and weighted voting and the learner SVM. The results of such ensembles are compared with those of plain classifiers based on individual parts and on holistic representations of the face. The experiments were carried out on the FERET database [9]. This paper is part of the groundwork of a project that is currently under way to investigate gender recognition with partial occlusion of the face.

## 2 Previous Related Works

In [6], the differentiation capabilities of full face, jaw, lip, nose, and eyes were evaluated by linear discriminant analysis. The most discriminant parts were the full face and jaw with recognition rates of 93.7% and 89.8%, respectively, while the poorest rates of accuracy were lower than 80% and correspond to nose and eyes. These results were obtained over a collection of expressionless Asian faces. Another closely related work is [7], where the roles of the full face, eyes and mouth in gender recognition were compared. Classification was performed by an SVM [10] with an RBF kernel, and results were obtained over an ad hoc set of 400 face images selected from three well-known databases: FERET [9], AR [11] and BioID [12]. The best accuracies were 85.5% and 81.25% for eyes and mouth, respectively. Considering that previous results depend on specific components (dataset, classifier, face part description), care should be taken when interpreting their conclusions. For example, the eyes appear to be more discriminant than the mouth in [7], while the mouth seems to be more accurate than the eyes in [6].

The work in [8] was addressed to generalize the results of the previous two works concerning the number of face parts and the diversity of the experimental design. In particular, gender recognition was performed from separate descriptions of single parts (eyes, nose, mouth and chin), and from holistic descriptions of the internal face, external face (hair, ears, contour) and the full face. Experiments were defined for two standard datasets, FERET [9] and XM2VTS [13], and several classifiers: SVM [10], {1,5,10}-NN, Quadratic Bayes Normal and

Parzen. The new results concerning the relevance of facial parts show highly correlated classifier behaviours within a database, but a strong dependence on the database.

### 3 Methodology

#### 3.1 From Face Images to Data Vectors

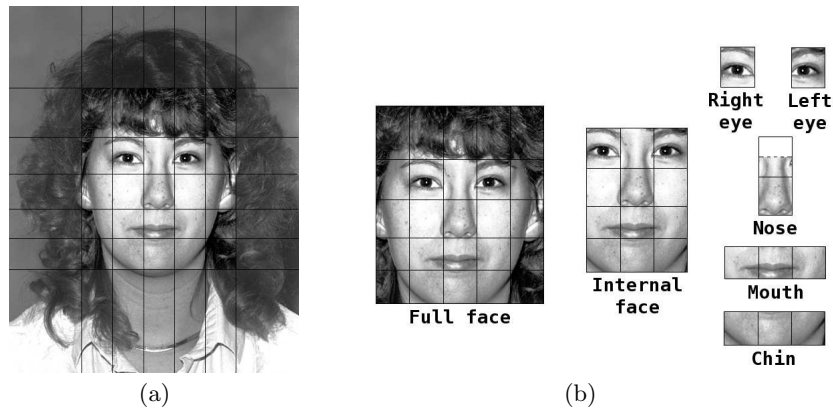
Given a grey image of a frontal pose of a face and the coordinates of the two eyes, seven subimages containing the two eyes, nose, mouth, chin, the rectangular area of the internal face and the full face are defined from expected proportions of an aesthetic face. Then a histogram equalization process is applied separately to these subimages in order to improve contrast and make lighting conditions more uniform. The subimages are represented as grey-level linear vectors, which are finally transformed using PCA to reduce dimensionality and to boost data information. Some details are given in the following paragraphs.

The process of subimage extraction is based on an empirical rule about the ideal balance of a human face sketched by Leonardo da Vinci [14, 15]. Leonardo stated that *perfect* facial harmony exists when the face can be divided vertically into three equal sections whose boundaries match the hairline, the eyebrows, the bottom of the nose and the chin, and when the face can be partitioned horizontally into five sections of identical width that approximate the width of one eye. In this paper, the horizontal division was carried out as defined by the previous rule, but vertical division was performed following a particular layout, in which some key facial features were centred in their corresponding regions (see Fig. 1(a)). The division points were automatically computed from the knowledge of eye coordinates.

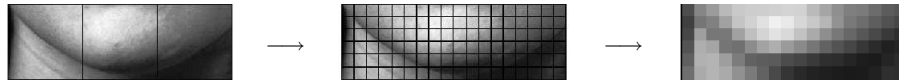
The rectangular regions of the proposed grid can fully contain eyes, but other important parts like mouth, nose and chin are bounded in only one direction. For example, the nose is bounded vertically, while the mouth is only bounded horizontally. To fully describe these features, new zones, which are referred to as subimages, are created by joining adjacent (parts of) rectangular regions (see Fig. 1(b)). The resulting subimages are then scaled down to low-resolution regions where new pixels are computed by averaging the original ones. This process was conducted by reducing each rectangular region of the grid to a new 6x6 rectangular window. This paper has not explored other sizes, which could provide even more accurate descriptions and better results. Figure 2 illustrates this process. The new reduced subimages are then represented as linear vectors.

#### 3.2 From Vectors to Gender Recognition

Given a dataset of face images with gender labels, seven new labelled datasets can be created, each composed of all the vectors that describe the same part (including the full face) of all the faces. From now on, this task can be managed as a standard pattern recognition problem. In fact, there are seven simple related tasks with the same number of instances and the same set of labels, which can be



**Fig. 1.** Images involved in the vector extraction process: (a) Image with the preprocessed face and the proposed grid; (b) Subimages, which are transformed into vectors.



**Fig. 2.** Example of scaling a chin subimage down to a low-resolution representation.

handled as individual problems or used in a combined approach where diversity would be supported by different data representation spaces.

In this paper three types of classifier ensembles are defined to evaluate the power of the joint contribution of individual face parts in gender recognition. Their results are then compared with those from the plain classifiers based on single parts and on the full face. Section 4 gives details about the learning and recognition processes.

## 4 Experiments

### 4.1 The FERET Database

The experiments are based on a well-known database of face images called FERET [9], which contain human faces acquired in a controlled environment with no restrictions as to age, race, or facial expression. There are several images of each person, which are replicated in three different sizes and organized into the same number of collections. Only faces in frontal pose without glasses were used, because their presence could strongly distort the effectiveness of eyes for gender recognition. The experiments involved 2147 medium-sized 256x384 pixel images from 834 different subjects separated into 842 female faces and 1305 male faces. From this image database, seven related vector datasets were inferred (see Sec. 3.1) containing descriptions of the left and right eyes, nose, mouth, chin, internal face and the full face. These datasets were transformed by PCA, where new attributes were selected to explain 99% of the total variance. Table 1 shows the number of appearance-based attributes that represent each facial part, and the final dimensionality after applying PCA.

**Table 1.** Number of attributes of FERET datasets before and after applying PCA.

	left eye	right eye	nose	mouth	chin	internal face	full face
before PCA	36	36	54	108	108	432	900
after PCA	20	21	29	48	36	62	101

## 4.2 The Classification Models

Several ensembles of parallel SVMs [10] were defined, whose decisions were combined by simple and weighted voting and by other SVM, respectively. The SVM with a linear polynomial kernel was chosen as the base classifier of the ensembles because of its proved effectiveness in this task [8].

With regard to the combiner used, the ensembles are referred to as  $E_{svot}$  (simple voting),  $E_{wvot}$  (weighted voting) and  $E_{svm}$  (SVM). The  $E_{svot}$  counts the number of base binary decisions for each class and chooses the most-voted one. In the case of  $E_{wvot}$  and  $E_{svm}$ , the *a posteriori* probabilities computed by the base classifiers were used. The decisions of  $E_{wvot}$  were made by regarding the greater of the two sums of probabilities of both classes over all the base classifiers. In the case of  $E_{svm}$ , a new vector space defined by the two *a posteriori* probabilities of all the base classifiers was built, the dimensionality being twice the number of classifiers. New training subsets were created from the projection in this probability space of the original training partitions.

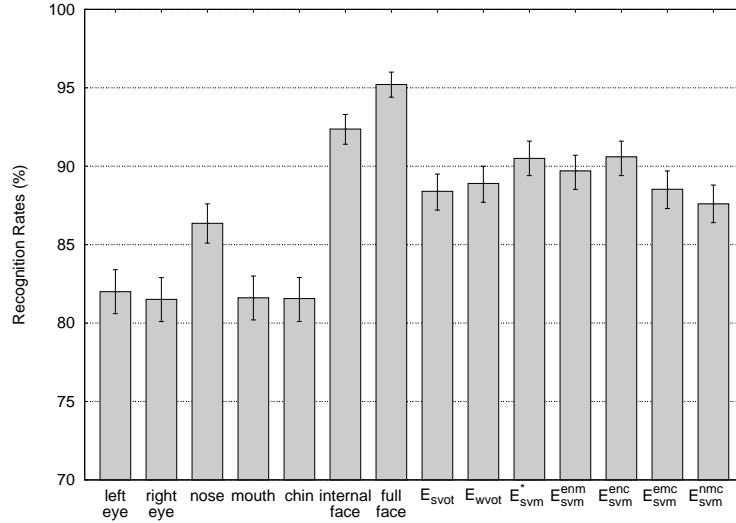
Different versions of the  $E_{svm}$  were implemented as regards the number and nature of the parts involved. The most general version combines the five individual parts and it is referred to as  $E_{svm}^*$ . The rest of the ensembles are combinations of subsets of three parts, in particular, {left eye, nose, mouth}, {left eye, nose, chin}, {left eye, mouth, chin} and {nose, mouth, chin} which are referred to as  $E_{svm}^{enm}$ ,  $E_{svm}^{enc}$ ,  $E_{svm}^{emc}$  and  $E_{svm}^{nmc}$ , respectively. These three-part ensembles simulate real situations in which only three visible parts can be used.

The plain gender classification of individual parts was also performed by the SVM with a linear polynomial kernel. Results were computed by averaging five independent runs of a 5-fold cross-validation, where all the face images of the same person were included in the same subset to avoid contamination effects between training and test partitions in each fold. The implementations of the learning algorithms used were those available in the PRTools Matlab package [16].

## 4.3 Experimental Results and Discussion

Figure 3 and Tab. 2 show the recognition performances of ensembles and of plain classifiers trained with individual descriptions of face parts and with the holistic descriptions of the full and the internal faces.

Most ensembles significantly outperform the plain gender classification of the individual parts with 95% confidence intervals for their average recognition rates. These results obtained over the FERET database show the existence of complementary information between the face parts, since the recognition rates



**Fig. 3.** Gender recognition rates [and their 95% confidence intervals] on FERET database using SVM as plain and base classifiers for single parts and ensembles.

of the ensembles were better than those achieved by the plain classifiers. The only exception is the simple classification of the nose whose interval overlaps with those of  $E_{svot}$ ,  $E_{svm}^{emc}$  and  $E_{svm}^{nmc}$ . A comparison between ensembles shows a better behaviour of  $E_{svm}^*$  with respect to the other two aggregations of the five parts ( $E_{svot}$  and  $E_{wvot}$ ).

The ensembles based on 3 parts, which use the SVM as the combiner, appear to be as discriminant as the combinations of 5 parts. Particularly, those in which the eye and the nose coincide,  $E_{svm}^{enm}$  and  $E_{svm}^{enc}$ , perform better than  $E_{svot}$  and  $E_{wvot}$ , which combine all the parts by voting. This is a very good result considering that these ensembles are meant to be useful in gender recognition under partial occlusion of faces, where combinations of no more than 2

**Table 2.** Gender recognition rates [and their 95% confidence intervals] on FERET database using SVM as plain and base classifiers for single parts and ensembles.

Plain classifiers						
left eye	right eye	nose	mouth	chin	internal face	full face
82.0	81.5	86.4	81.6	81.6	92.4	95.2
[80.6, 83.4]	[80.1, 82.9]	[85.1, 87.64]	[80.2, 83.0]	[80.1, 82.9]	[91.4, 93.3]	[94.4, 96.0]
Ensembles						
$E_{svot}$	$E_{wvot}$	$E_{svm}^*$	$E_{svm}^{enm}$	$E_{svm}^{enc}$	$E_{svm}^{emc}$	$E_{svm}^{nmc}$
88.4	88.9	90.5	89.7	90.6	88.5	87.6
[87.2, 89.5]	[87.7, 90.0]	[89.4, 91.6]	[88.5, 90.7]	[89.4, 91.6]	[87.3, 89.7]	[86.4, 88.8]

or 3 visible prominent parts appear as a very likely scenario. It is worth noting that the use of an SVM as the combiner was more effective than voting. It is possible that a classifier works as an error-correcting combiner by learning how the base classifiers make errors and how to associate their combinations with correct outputs.

Nevertheless, the ensembles were unable to achieve the rates of plain classifiers based on the holistic descriptions of internal and full faces. While the best ensemble is about 2% less accurate than the classifier based on the internal face, the difference between that ensemble and the classifier trained with the full face rises to 5%. There are two major causes that seem to explain the previous results. Firstly, the holistic representations of the internal and full faces include information about the configurational relationships of face parts, which also provide a valuable source of differences between genders. Secondly, the full face contains other prominent features of the face that have not been considered in this work, like ears, hair and face contour, which are very discriminant by themselves [8].

Finally, it should be noted that these results are subject to particular methods and data, so more experiments should help to deliver more robust conclusions.

## 5 Conclusions

This paper has evaluated the extent to which the most prominent face parts can complement each other in recognizing genders. To this end, several classifier ensembles based on the (expected) discriminant diversity of face parts were designed and tested. These combined schemes were introduced as suitable solutions to the gender recognition problem when faces are partially occluded and holistic representations are not possible. The experiments involved combinations of base classifiers trained with separate descriptions of the left and right eyes, nose, mouth and chin, whose collective decisions were made by simple and weighted voting and by a learner based on vectors of *a posteriori* probabilities. The classification results of ensembles were compared to those of plain classifiers defined from individual parts of the face and from holistic descriptions of the face.

The experiments over the FERET database proved that the joint contribution of separate parts is more effective for gender recognition than isolated parts, but less discriminant than the holistic descriptions of faces. Unlike the simple aggregation of parts, the holistic representation includes the relationships between face parts, whose usefulness at discriminating between female and male has been proved by psychological experiments [1, 2].

This work partially supports an interesting line of research consisting in gender recognition in human faces under partial occlusions. Instead of the traditional holistic description of the face, a distributed solution for identifying and combining only the visible parts of the face seems to be more suitable.

## 6 Acknowledgments

This work was partially funded by Projects CSD2007-00018 and DPI2006-15542 from the Spanish Ministry of Science and Education, and P1-1B2007-60 from the Fundació Caixa Castelló-Bancaixa.

## References

1. Bruce, V., Burton, A., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., Linney, A.: Sex discrimination: how do we tell the difference between male and female faces. *Perception* **22**(2) (1993) 131–152
2. Burton, A., Bruce, V., Dench, N.: What's the difference between men and women? evidence from facial measurement. *Perception* **22**(2) (1993) 153–176
3. Wu, J., Smith, W., Hancock, E.: Learning mixture models for gender classification based on facial surface normals. In: LNCS: Proc. of 3rd IBPRIA. Volume 4477., Girona, Spain, Springer (2007) 39–46
4. Lapedriza, A., Marín-Jiménez, M., Vitrià, J.: Gender recognition in non controlled environments. In: Proc. of 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, IEEE (2006)
5. Moghaddam, B., Yang, M.: Learning gender with support faces. *IEEE Transactions on PAMI* **24**(5) (2002) 707–711
6. Kawano, T., Kato, K., Yamamoto, K.: A comparison of the gender differentiation capability between facial parts. In: Proc. of 17th International Conference on Pattern Recognition (ICPR'06), Cambridge, U.K., IEEE (2004)
7. Buchala, S., Davey, N., Frank, R., Gale, T., Loomes, M., Kanargard, W.: Gender classification of faces images: The role of global and feature-based information. In: LNCS: Proc. of 11th Int. Conf. on Neural Information Processing. Volume 3316., Calcutta, India, Springer (2004) 763–768
8. Andreu, Y., Mollineda, R.: The role of face parts in gender recognition. In: LNCS: Proc. of Int. Conf. on Image Analysis and Recognition (ICIAR 2008). Volume 5112., Póvoa de Varzim, Portugal, Springer (2008) 945–954
9. Phillips, H., Moon, P., Rizvi, S.: The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(10) (2000)
10. Cortes, C., Vapnik, V.: Support-vector networks. *Machine Learning* **20**(3) (1995) 273–297
11. Martinez, A., Benavente, R.: The AR face database. Technical Report 24, CVC (June 1998)
12. Jesorsky, O., Kirchberg, K., Frischholz, R.: Robust face detection using the hausdorff distance. In Bigun, J., Smeraldi, F., eds.: *Audio and Video based Person Authentication: Proc. of AVBPA 2001*, Springer (2001) 90–95
13. Messer, K., Matas, J., Kittler, J., Lüttin, J., Maitre, G.: XM2VTSDB: The extended M2VTS database. In: *Audio and Video-based Biometric Person Authentication, AVBPA'99*. (1999) 72–77
14. Nainia, F.B., Mossb, J.P., Gilc, D.S.: The enigma of facial beauty: Esthetics, proportions, deformity, and controversy. *American Journal of Orthodontics and Dentofacial Orthopedics* **130**(3) (2006) 277–282
15. Oguz, O.: The proportion of the face in younger adults using the thumb rule of Leonardo da Vinci. *Surgical and Radiologic Anatomy* **18**(2) (1996) 111–114
16. Duin, R., Juszczak, P., Paclik, P., Pekalska, E., de Ridder, D., D.M. Tax, J.: *PRTools4, A Matlab Toolbox for Pattern Recognition*. Delft University of Technology, The Netherlands. 4.0 edn. (2004)