Gender Recognition in Non Controlled Environments

Àgata Lapedriza†, Manuel J. Marín-Jiménez‡ and Jordi Vitrià†

† Computer Vision Center (CVC)
Computer Science Dpt.
Universitat Autònoma de Barcelona
Bellaterra, Spain 08193
{agata, jordi}@cvc.uab.es

‡ Computer Science and Artificial Intelligence Dpt.
University of Granada
Granada, Spain 18071
mjmarin@decsai.ugr.es

Abstract

In most of the automatic face classification applications, images should be captured in natural environments, where partial occlusions or high local changes in the illumination are frequent. For this reason, face classification tasks in uncontrolled environment are still nowadays unsolved problems, given that the loss of information caused by these artifacts can easily mislead any classifier. We present in this paper a system to extract robust face features that can be applied to encode information from any zone of the face and that can be used for different face classification problems. To test this method we include the results obtained in different gender classification experiments, considering controlled and uncontrolled environments and extracting face features from internal and external face zones. The obtained rates show, on the one hand, that we can obtain significant information applying the presented feature extraction scheme and, on the other hand, that the external face zone can contribute useful information for classification purposes.

1. Introduction

Many face classification problems such as subject verification or subject recognition can take benefit from a previous successful gender classification. Since its importance, several works can be found in the literature focusing their attention on this problem. For example Cottrell et al. [1] used a two layer neural network approach, where each face image was compressed in the first layer of the network, and classified in the second layer. They obtained an accuracy of 63% using only 64 training images. Later, Brunelli et al. [2], used a set of 16 geometric features per image to train two hyper basis function networks, and achieved accuracies of 79% in a database composed of 168 training images. Also similar results were achieved by Abdi et al. [3] training a perceptron classifier using PCA-based features of the input images, achieving a performance of the 91.8%. Nevertheless, these results were far from the human capability, taking into account that some psychological studies [4] have shown that we are able to achieve accuracies close to 96%.

More recently, Moghaddam et al. [5] obtained the best performance, up to our knowledge, using a Support Vector Machine with Radial Basis Function kernels. They achieved a 96.6% recognition rate using a large face database, more concretely 1755 faces of the FERET face database. Nevertheless we want to point up that the images in the FERET face database are acquired under controlled conditions.

These previous works have motivated us to set out the gender classification problem in uncontrolled environments, where the loss of information caused by partial occlusions and high local changes in the illumination is usual. For this reason, it is important to consider as sources of information as possible.

Face features can be divided in two sets, depending on the zone they are located: internal features, composed by eyes, nose and mouth, and external features, located at head, chin, and ears. These two face zones are illustrated in figure 1. Our idea is to consider in our experiments not only the internal face zone, like most of the current face classification methods, but also take into account the information located at external face zone. Nevertheless, to aim this purpose, we need a method valid to extract external information, given that classical bottom-up approaches can not be here applied because of the high variability of these face zones.

In this work we present a system to extract information from face images that can be used to obtain features from any face zone. Moreover we have perform different gender classification experiments to test this method and given that the proposed method is valid to extract information from

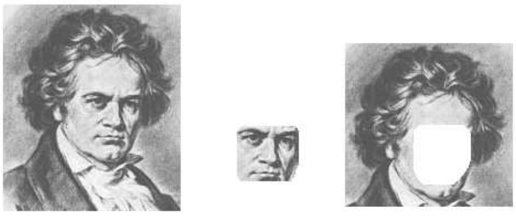

Figure 1. Full face, internal features and external features of a human face.

the whole face, we will be also able to compare the usability of internal and external features for face classification purposes.

The paper is organized as follows: section 2 describes the feature extraction method, section 3 presents our experiments and the obtained results and, finally, section 4 concludes this work.

2. Fragment-based Robust Face Feature Extraction

Our objective is to develop a method for extracting features from all the zones of a human face image, even from the chin, ears or head. Nevertheless, the external face areas are high variable and it is not possible to establish directly in these zones a natural alignment. For this reason, we propose fragment based system to aim this purpose. The general idea of the method can be divided in two steps. First, we select a set of face fragments from any face zone that will be considered as a model. After that, given an unseen face image, we weight the presence of each fragment in this new image. Proceeding like this, we obtain a positive weight for each fragment, and each weight is considered as a feature. Moreover, we obtain in this way an aligned feature vector that can be processed by any known classifier.

To establish the model we select a set of fragments $F = \{F_i\}_{i=1..N}$ obtained from face images. This selection should be made using an appropriate criterion, depending on the task we want to focus on and on the techniques that will be used to achieve the objective. In our case we wanted a high quantity of different fragments to obtain a rich and variable model. For this reason we have selected them randomly, adding a high number of elements. Once the fragments are selected, we apply to them the feature extraction method studied by Marín-Jiménez et al. in [6], which is based on responses to multi-scale filter banks (oriented first and second order Gaussian derivatives) and combined following the HMAX model proposed by Poggio et al. [7] and lately revised by Serre et al. [8]. This kind of features exhibits interesting invariance properties (illumination, translation, scale, small rotations). In this way we obtain a new representation of these fragments less sensitive to scale or short rotation changes, $R(F) = \{R(F_i)\}_{i=1..N}$.

After the construction of the model, given a new unseen image I we look for these selected fragments in the image and obtain for each fragment a value that indicates how present it is in the image. To obtain this value we proceed as follows: for each fragment $F_i \in F$ we consider all the subimages $\{J_k\}_{k=1..M}$ in I having the same size of F_i. Then, we obtain their particular representations $\{R(J_k)\}_{k=1..M}$ and compute the distances between $R(F_i)$ and each $R(J_k)$. The minimum distance indicates the contribution of the fragment F_i in the image. Also the distance criterion considered in this step should be selected depending on our objectives. We have used in this work the following criterion to compute the distance between two representations R_1 and R_2:

$$d(R_1, R_2) = \exp(-\gamma \parallel R_1 - R_2 \parallel^2) \quad (1)$$

where γ is a positive value that we can choose to modulate the exponential curve and $\parallel \cdot \parallel$ symbolizes the euclidean distance.

3. Experiments and Results

The experiments have been performed using the FRGC Database (*http://www.bee-biometrics.org/*). We have considered separately two sets of images: on the one hand images acquired under controlled conditions, having uniform grey background, and on the other hand images acquired in cluttered scenes. These sets are composed by 3440 and 1886 samples respectively. Some examples of these images can be seen in figure 2.

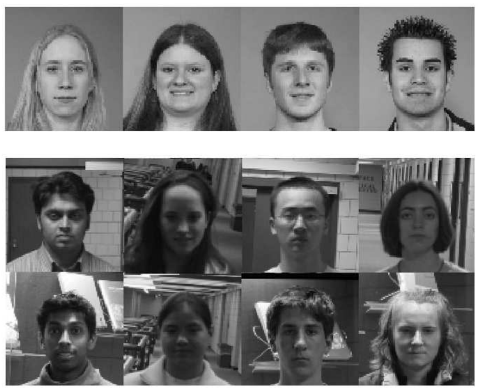

Figure 2. Examples of images included on the FRGC database acquired under controlled conditions (first row) and in uncontrolled scenes (second and third row).

All the experiments have been performed three times: first considering only the external features, second considering only the internal information and finally considering both feature sets together. With these results we are able to test the presented feature extraction method and to compare the contribution of the external and the internal face features separately. We encode the internal and the external information following in both cases the feature extraction method explained in section 2. In concrete, the filter bank selected for building the features is based on second order Gaussian derivative and Laplacian of Gaussian functions. In this way, we construct the models randomly selecting 2000 fragments from the desired zone and, after that, we separate the 90% of the samples to train the classifier and the rest of the considered images are used to perform the test.

We have used in the experiments two boosting classifiers, given that they have been proved to be effective in several classification applications. First AdaBoost [9] (with decision stumps), that is the most commonly used version of this technique, and second JointBoosting [10], a more recently development of this system characterized by the possibility of its application in multi-class case. The main idea of this new boosting procedure is finding common features that can be shared across the classes, improving this way the global performance of the classification and reducing the computational cost.

We have performed a 10-fold crossvalidation test in all the cases and we show for each experiment the mean of the rates and the corresponding confidence interval.

3.1 Controlled environments

Some examples of the fragments that compose the model of internal and external features are shown in figure 3. Here the places where the method has found some of the fragments of the model are also illustrated, separating the localization of some internal fragments, on the left hand, and the localization of some external fragments on the right hand. We observe that in this case, where we have uniform grey background, the fragments are in general correctly placed.

The results of the experiments performed using these set of images are included in table 3.1. We can see that the accuracies obtained using only external features or only internal features are quite similar, although the best result considering these sets separately is achieved using external features and classifying with JointBoosting. Nevertheless, in controlled environments the best accuracy that we have obtained is 96.77%, considering external and internal features together and classifying also with JointBoosting.

3.2 Uncontrolled environments

Figure 4 include some fragments of the models obtained in the case of uncontrolled environments. Also we include the localization of some blocks of the internal and external feature models in a new unseen image. Here we observe that the most vulnerable zone for our feature extraction method is clearly the external one, since the model is constructed with fragments that contains background pixels and the background here is not uniform or constant. Nevertheless, given a block from the external zone model, we can see that a considerable amount of pixels is from real external face zone and for this reason the method is able of placing enough fragments in the correct position.

The achieved accuracy rates in the experiments performed using the images acquired in uncontrolled environments are included in table 3.2. We can see again that the results obtained using only external or only internal features are also quite similar. And, like before, the best result considering only one of these feature sets is obtained using external features and JointBoosting classifier. Nevertheless, the best global accuracy achieved with this image set is obtained again considering both internal and external features together and classifying with JointBoosting. This accuracy rate is 91.72% and also in this case we have the lowest confidence interval.

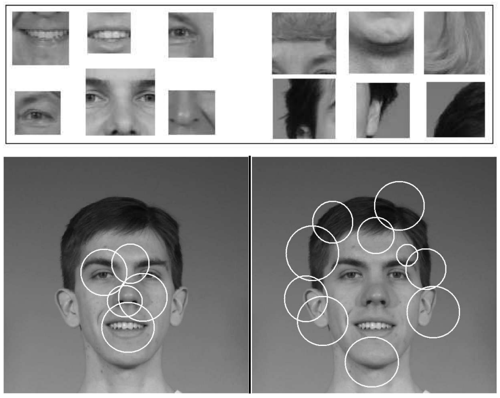

Figure 3. Samples of the model (first row) and position returned by our method for the fragments included in the model in an unseen image acquired in controlled environments.

Table 1. Gender recognition in controlled environments experiments: achieved results.

	AB	JB
External	$94.60\% \pm 0.60\%$	$96.70\% \pm 0.80\%$
Internal	$94.66\% \pm 0.76\%$	$94.70\% \pm 1.10\%$
Combination	$94.60\% \pm 0.60\%$	$96.77\% \pm 0.47\%$

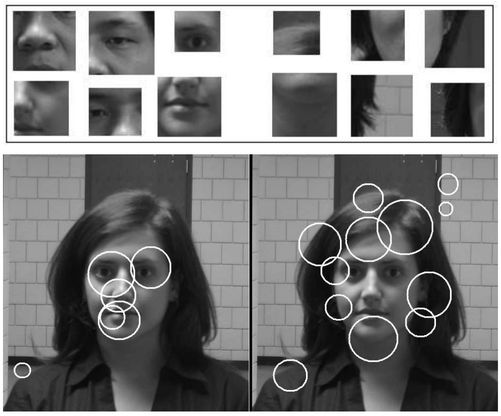

Figure 4. Samples of the model (first row) and position returned by our method for the fragments included in the model in an unseen image acquired in uncontrolled environments.

Table 2. Gender recognition in uncontrolled environments experiments: achieved results.

	AB	JB
External	$87.38\% \pm 2.46\%$	$90.61\% \pm 1.80\%$
Internal	$87.04\% \pm 3.16\%$	$89.77\% \pm 2.34\%$
Combination	$87.99\% \pm 2.20\%$	$91.72\% \pm 1.56\%$

4. Conclusions and Future Work

In this paper we have presented a system to extract internal and external information from human face images. With this method we obtain an aligned feature vector that can be processed by any known classifier.

To test this feature extraction scheme, we include the results obtained by different gender classification experiments, using two current classifiers: AdaBoost and JointBoosting. Half of these experiments have been performed with images acquired under controlled conditions and the rest with images acquired in natural environments. Moreover, each test have been made considering only internal features, only external features and both feature sets together to be able to compare the usefulness of these two features sets for gender classification purposes.

From the results obtained by our experiments we can conclude that the presented system allows to obtain information from face images useful for gender classification. For this reason, we think that it can be extended to other computer vision classification problems such as subject verification or subject recognition. Moreover, since our method is valid to extract features from any face zone, we have compared the usefulness of external against internal features and it has been shown that both sets of features play an important role in gender classification purposes. For this reason, we propose to use this external face zone information to improve the current face classification methods that consider only internal features.

Acknowledgments

This work is supported by MCYT grant TIC2003-00654, FP2000-4960 Ministerio de Ciencia y Tecnologia (Spain), grant MEC-FPU AP2003-2405 and project TIN2005-01665.

References

[1] G. Cottrell, "Empath: Face, emotion, and gender recognition using holons," *Advances in Neural Information Processing Systems*, vol. 3, p. 564:571, 1991.

[2] R. Brunelli and T. Poggio, "Hyperbf networks for gender classification," 1992, pp. 311–314.

[3] H. Abdi, D. Valentin, B. Edelman and A. O'Toole, "More about the difference between men and women: evidence from linear neural networks and the principal component approach," *Perception*, vol. 24, p. 539:562, 1995.

[4] A.M. Burton, V. Bruce and N. Dench, "What's the difference between men and women? evidence from facial measurement," *Perception*, vol. 22, p. 153:76, 1993.

[5] Moghaddam B. and Yang, "Learning gender with support faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 707–711, jan 2002.

[6] M.J. Marín-Jiménez and N. P. de la Blanca, "Empirical study of multi-scale filter banks for object categorization," in *IEEE ICPR*, August 2006.

[7] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience*, vol. 2, no. 11, pp. 1019–1025, 1999.

[8] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," in *IEEE CSC on CVPR*, June 2005.

[9] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," Dept. of Statistics. Stanford University, Tech. Rep., 1998.

[10] A. B. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing features: Efficient boosting procedures for multiclass object detection." in *CVPR (2)*, 2004, pp. 762–769.