Gender Recognition from a Partial View of the Face Using Local Feature Vectors

Yasmina Andreu, Ramón A. Mollineda, and Pedro García-Sevilla

Dpto. Lenguajes y Sistemas Informáticos
Universidad Jaume I. Castellón de la Plana, Spain
{yandreu,mollined,pgarcia}@uji.es

Abstract. This paper proposes a gender recognition scheme focused on local appearance-based features to describe the top half of the face. Due to the fact that only the top half of the face is used, this is a feasible approach in those situations where the bottom half is hidden. In the experiments, several face detection methods with different precision levels are used in order to prove the robustness of the scheme with respect to variations in the accuracy level of the face detection process.

Keywords: Gender recognition, Face image analysis, Local features.

1 Introduction

Automatic gender recognition is an emerging issue in the face analysis area. A number of fields can benefit from the application of this technique, for example, demographic data acquisition, access control and other face analysis tasks such as age, ethnic or face recognition. In the latter cases, the prior prediction of the gender can reduce the search space to approximately half.

Some recent papers have addressed this problem, most of them proposing solutions based on descriptions of full faces [1,2,3]. This holistic representation is undoubtedly the most useful one for discriminating between genders when a frontal view of the whole face is available. The reasons for this are obvious: along with all the parts of the face, this description contains the structural relationships between parts that also contribute to discriminating between genders [4].

However, in the real world faces can usually be partially covered by accessories such as scarves, hats, veils and other objects. In these situations, the use of local approaches seems to be a suitable choice.

Furthermore, the unconstrained problem of face recognition and analysis is far from solved. A general solution should deal with a wide diversity of poses, lighting conditions, hidden areas, facial expressions, image resolutions and many other factors. One of the most ambitious works is [5], where a face recognition task is carried out on test faces that are imprecisely localized, partially occluded and with no restrictions on their expressions, while using a single training sample per class. In the case of gender recognition, as far as we know, no previous efforts in similar conditions have been reported.

This paper proposes a gender classification scheme based on the combination of several appearance-based local descriptions of the top half of the face. This approach can handle situations where this part of the face is available, while the bottom half could be hidden. The method applied to locally extract features from the images is similar to that used in [6].

In our experiments, the top half of the face was detected using two methods: in the first method the exact coordinates of the eyes were used to extract an accurate subimage, whereas in the second one the top half of the face was automatically detected using the face image as the only input. Additionally, experiments in which the area that contains the face was intentionally shifted were carried out.

The experiments demonstrate that the classification method based on local features used to recognize gender from the top half of the face is robust to the accuracy level of face detection.

The experiments were defined for the FERET database [7] and a 1-NN classifier.

The rest of this paper is organized as follows. Section 2 presents the methodology used to extract local features of the faces and also the classification technique implemented. Section 3 presents the description of the image database as well as the experiments and the discussion of their results. Finally, Section 4 summarizes the conclusions and future work.

2   Methodology

2.1   Feature Extraction

Given a greyscale image of a frontal view of a face, this section explains how the features that describe the face are extracted (see Fig. 1 for a graphical explanation). As a general overview, the process detects the face in the image and extracts the top half of that face. Then several vectors of local features are obtained from the extracted area using a window that enables us to locally select the features. Finally, the features are filtered in order to reduce complexity and to discard the information that is not very relevant.

This process can be divided in the following steps:

1. **Face detection**
   Two different approaches have been used to detect the face in the image. In the first method the face is detected using the exact coordinates of the two eyes, which are available for each image in the database, while in the second one the detection is completely automatic.
   The process followed in the first approach is based on an empirical rule about the ideal balance of a human face sketched by Leonardo da Vinci [8,9]. The division points are automatically computed from the knowledge of the coordinates of the eyes and the distance between them. For the automatic method, the Viola and Jones algorithm [10] implemented in the OpenCV library [11] is used. On the left, Fig. 2 shows the areas obtained after applying
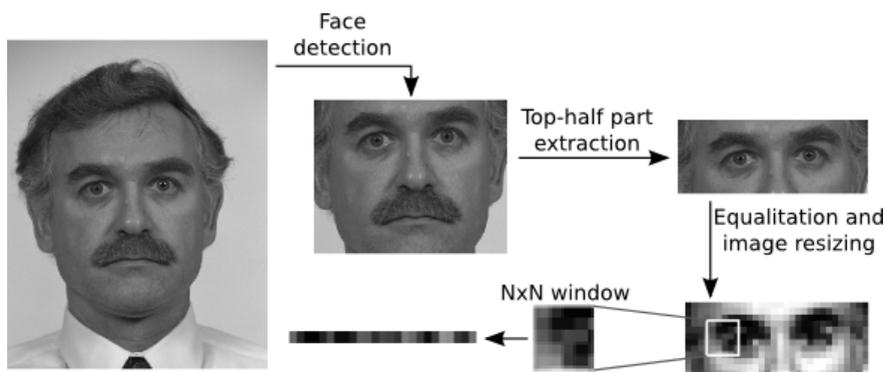
Fig. 1. Feature extraction process. Firstly, the face is detected and the top half is extracted, equalized and resized. Then the image is scanned with an $N \times N$ window to obtain a set of local feature vectors.
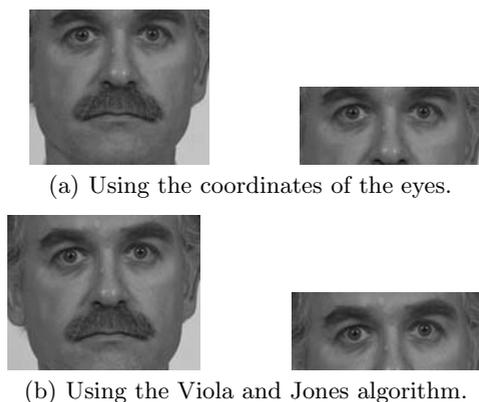

(a) Using the coordinates of the eyes.

(b) Using the Viola and Jones algorithm.

Fig. 2. Facial area and the top half of a face detected using two different methods

each face detection method to the same image and on the right, it shows the top half used in the experiments.

2. **Image pre-processing**
   Given the coordinates of the face in the image, a subimage containing the top half of that face is extracted (see right side of Fig. 2) and it is resized to a given width. A three-lobed Lanczos windowed sinc function [12] is used in the interpolation process required for the resizing step. Finally, histogram equalization is applied to improve contrast and to make lighting conditions more uniform.

3. **Local features extraction**
   The basic idea is similar to that introduced in [6]. It consists of a window of some moderate size $N \times N$ that is scanned across the image. From each window position, a vector containing all the grey level values of the pixels

within the window is built. As a result, a set of local feature vectors that describe the image are extracted.

4. **From grey levels to numeric labels**

   In order to obtain vector descriptions that are more robust to illumination changes, grey levels are substituted by their internal ranking position within each vector. In other words, each grey level is replaced by a numeric label that represents its position in the sorted list of all grey values in ascending order. Moreover, for moderate values of N, the size of the new attribute space $(N \times N)$ is smaller than the 256 possible grey levels, which can also contribute to making vector representations more general. This codification process can be detailed as follows:

   (a) Let $x = (x_1, x_2, ..., x_{N \times N})$ be a vector of grey-level components

   (b) Let G be a threshold that represents the maximum difference between two grey levels to consider them equal

   (c) Compute the ordered version of x, $x' = (x_{i_1}, x_{i_2}, ..., x_{i_{N \times N}})$, $x_{i_j} \leq x_{i_{j+1}}$, where i_j is the position of x_{i_j} in the original vector x

   (d) Build a parallel vector $r' = (r_1, r_2, ..., r_{N \times N})$ with the ranking positions of the components of x as regard to x', according to the algorithm:

   $k = 1$
   $ranking = 1$
   for $j = 1 : N \times N$ do,
      if $x_{i_j} - x_{i_k} > G$,
         $k = j$
         $ranking = ranking + 1$
      $r'_j = ranking$

   (e) Let r be the re-sorted vector r' with regard to the original positions of the components in x

   for $j = 1 : N \times N$ do,
      $r_{i_j} = r'_j$

   This process is shown graphically in Fig. 3.

5. **Reducing the size/complexity of the problem**

   The goal of this step is to reduce the complexity of the problem. In order to achieve this goal, the vectors where the number of different numeric labels is lower than a threshold T are ruled out. These vectors match with those regions with low contrast, which are assumed to not provide too much discriminant information.

   This step is optional so it was not used in experiments unless this is remarked on the experiments section.
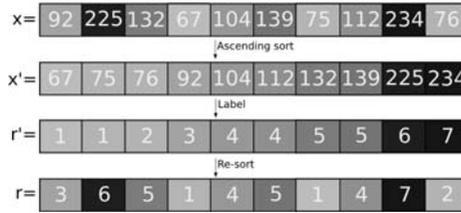
Fig. 3. Example of the transformation of a grey level vector into a ranking vector with $G = 8$. Two grey levels with a difference lower than or equal to 8 are considered similar, resulting in the same ranking position.

2.2   Classification

Applying the feature extraction process explained in Sect. 2.1 several vectors are extracted to describe the same face, so a particular classification process should be used to obtain only one class label per face from those vectors.

Firstly in the classification process, the nearest neighbour of each vector extracted from the test images is computed and its class label is assigned to the corresponding vector. Then the class assigned to each test image is the one obtained by majority voting of its vectors.

The metric used to compute the nearest neighbour is $\sum_{i=1}^{n} |v_i - w_i|$, where v and w are vectors extracted from a training image and a test image respectively.

3   Experiments

The experiments were designed to prove that the gender recognition method presented is robust to inaccuracies in the detection of top half of the face. To this end, two different face detection algorithms were used. The first one uses the exact coordinates of the eyes obtained from the face database as input information whereas the second one is completely automatic. Different combinations of these algorithms were applied to the training and test image sets leading to the following experiments:

Experiment 1 : Training set: based on the exact coordinates; test set: based on the exact coordinates.

Experiment 2 : Training set: based on the exact coordinates; test set: automatic method.

Experiment 3 : Training set: automatic method; test set: automatic method.

In addition to these experiments another one (**Experiment 4**) where the coordinates of the area that contains the face were randomly shifted in all directions between 0 to 15 pixels was performed. In this case, the method that detects the face from the exact coordinates of the eyes was used.

Two series of these four experiments were carried out. In the first series all the feature vectors extracted from the images were used and in the other series the number of feature vectors was reduced before the classification process took place (see step 5 in Sect. 2.1).

3.1   Description of the Database

The experiments are based on the well-known FERET database of face images [7]. It contains human faces acquired in a controlled environment with no restrictions such as age, race, or facial expressions. There are several images of each person, which are replicated in three sizes organized into as many collections. Only faces in a frontal pose without glasses were used, because glasses could strongly distort the effectiveness of the top half of the face in the gender recognition task. The experiments involved 2147 medium-sized images of 256×384 pixels from 834 subjects separated into 842 female faces and 1305 male faces.

3.2   Description of Experiments

The images used in the experiments were divided into two sets: training and test, consisting of 60% and 40% of the total amount of images respectively. As there are several images per subject, this division was carefully implemented in order to assign all the images of the same subject to the same set.

Each of the top half faces used was resized to obtain images of 30 pixels in width using a resizing process that maintains the aspect ratio of the original image.

The feature extraction process uses a window of size 9×9 (see step 3 in Sect. 2.1). As a result, vectors of 81 components each were obtained and the number of vectors extracted from each face was usually 110, since the top half of the faces used had the same width-height ratio in most cases.

For the process that transforms the grey level vectors into numeric label vectors, the threshold used to consider two grey levels equal was $G = 8$, so 32 different numeric labels are possible since 256 grey levels exist (see step 4 in Sect. 2.1).

For the process of reducing the complexity of the data, a threshold $T = 16$ was used (only in the second series of experiments) to rule out those local feature vectors with poor information for discriminating between genders (see step 5 in Sect. 2.1). It is worth noting that due to the fact that the threshold T was set to 16 and G set to 8, the discarded vectors are those whose number of different label values is at most 16 (which is half the total amount of labels).

Table 1. Recognition rates for both series of experiments

	Tra and Tst coordinates (Exp. 1)	Tra coordinates Tst auto detection (Exp. 2)	Tra and Tst auto detection (Exp. 3)	Shifting the area of the face (Exp. 4)
1st series	79.55%	83.16%	82.57%	79.21%
2nd series	81.07%	82.30%	88.51%	83.06%

<div style="display:flex; justify-content:space-between">
(a) Males classified as females.   (b) Females classified as males.
</div>

Fig. 4. Misclassified subjects

3.3   Discussion of Results

Table 1 shows the recognition performance for the experiments that use all the features extracted from the images (first series of experiments) and the rates from the experiments where the complexity of the problem was reduced (second series of experiments) respectively.

All the results obtained are very close to or higher than 80% despite the fact that the system does not have any restriction on how accurate the detection of the face should be.

In the first series of experiments, the best recognition rates were achieved when the face was automatically detected in the test images or in both sets. This could be explained because this automatic method provides a wider area that contains more of the face and the method that uses the coordinates of the eyes obtains a smaller region that provides less information.

As expected, the recognition rates achieved for the second series of experiments are higher than those obtained in the first series. Taking into account that in this case a lower number of vectors, which were (hopefully) the most discriminant, were used to describe the faces, it is possible that those vectors that were ruled out were the ones that provided information that confused the classifier.

As a further benchmark to assess the results of this work, work of paper [13] is considered. It also focuses on the task of gender classification over the FERET database. Its experiments were based on the full face, the internal face (which excludes hair, ears and facial contour) and the eyes, and other parts of the face. The feature extraction process was executed using the coordinates of the eyes to accurately locate the facial region of interest. Under such controlled conditions, the best recognition rates were 95.21%, 92.37% and 85.47% using the full face, the internal face and the eyes respectively. The best rate in this paper is 88.51%, which unlike the results in [13] was achieved from the completely automatic extraction of the top half of the face, so it is more suitable for use in real situations.

Fig. 4 shows some misclassified faces. It could be that these are somehow confusing even for human beings, since in these cases there is little information to enable us to guess the gender of these people.

4   Conclusions

This paper has proposed a gender recognition scheme based on local feature descriptions from the top half of the face, which makes it suitable for use in real situations where the bottom half is hidden. As the experiments demonstrate this method is robust to variations in the accuracy of face detection, since the classification results obtained from the intentional shift of the face regions produced similar results to those achieved under normal conditions.

Future lines of research may include the definition of methodologies for learning the model parameters, the comparison of the local feature vectors used in this paper with other local representations such as Local Binary Patterns and the design of more powerful classifiers to handle the gender recognition task.

Acknowledgments

References

1. Lapedriza, A., Marín-Jiménez, M., Vitrià, J.: Gender recognition in non controlled environments. In: Proc. of 18th ICPR. IEEE, Hong Kong (2006)
2. Moghaddam, B., Yang, M.: Learning gender with support faces. IEEE Trans. on PAMI 24(5), 707–711 (2002)
3. Wu, J., Smith, W., Hancock, E.: Learning mixture models for gender classification based on facial surface normals. In: Martí, J., Benedí, J.M., Mendonça, A.M., Serrat, J. (eds.) IbPRIA 2007. LNCS, vol. 4477, pp. 39–46. Springer, Heidelberg (2007)
4. Burton, A., Bruce, V., Dench, N.: What's the difference between men and women? evidence from facial measurement. Perception 22(2), 153–176 (1993)
5. Martínez, A.: Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. IEEE Trans. on PAMI 24(6), 748–763 (2002)
6. Paredes, R., Pérez-Cortes, J.C., Juan, A., Vidal, E.: Local representations and a direct voting scheme for face recognition. In: PRIS 2001: Proceedings of the 1st International Workshop on Pattern Recognition in Information Systems, pp. 71–79. ICEIS Press (2001)
7. Phillips, H., Moon, P., Rizvi, S.: The FERET evaluation methodology for face recognition algorithms. IEEE Trans. on PAMI 22(10) (2000)
8. Nainia, F.B., Mossb, J.P., Gillc, D.S.: The enigma of facial beauty: Esthetics, proportions, deformity, and controversy. American Journal of Orthodontics and Dentofacial Orthopedics 130(3), 277–282 (2006)
9. Oguz, O.: The proportion of the face in younger adults using the thumb rule of Leonardo da Vinci. Surgical and Radiologic Anatomy 18(2), 111–114 (1996)
10. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vision 57(2), 137–154 (2004)
11. Bradski, G.R., Kaehler, A.: Learning OpenCV. O'Reilly (2008)
12. Turkowski, K.: Filters for common resampling tasks, 147–165 (1990)
13. Andreu, Y., Mollineda, R.A.: The role of face parts in gender recognition. In: Campilho, A., Kamel, M.S. (eds.) ICIAR 2008. LNCS, vol. 5112, pp. 945–954. Springer, Heidelberg (2008)