Gender Recognition from Face Images
Using a Geometric Descriptor

Marcos Vinicius Mussel Cirne*, Helio Pedrini*

*Institute of Computing, University of Campinas (UNICAMP)
Campinas, SP, 13083-852, Brazil
marcosvcirne@gmail.com, helio@ic.unicamp.br

Abstract—Gender recognition from face images is a challenging problem with applications in various knowledge domains, such as biometrics, security and surveillance, human-computer interaction, among others. In this work, we propose and evaluate a novel method for gender recognition based on a geometric descriptor constructed from a pre-defined face shape model. The proposed approach, tested on four different face datasets, achieved superior results for most cases when compared to other methods based on geometric descriptors.

Index Terms—Gender recognition, facial images, geometric descriptor, fiducial points, face detection

I. INTRODUCTION

Automatic gender recognition [1], [2], [3], [4] from face images has become an active research field in the context of image analysis and machine learning due to its large number of applications, such as human-computer interaction, surveillance systems, demographic studies, customized advertising, biometrics, among others.

Although gender recognition is usually trivial when performed by human beings, it is a challenging task for machines. Illumination change, pose, occlusion, noise, age and ethnicity are some examples of conditions that may affect the recognition process.

Gender classification can be defined as a two-class problem, where a given face image is assigned to one of the two classes, male or female. Typically, a gender recognition system is composed of the following steps: (i) pre-processing, (ii) feature extraction and (iii) classification. In the pre-processing step, techniques for face detection [5], illumination normalization [6], noise filtering [7] can be applied to prepare the datasets. Facial features are then extracted from the pre-processed images through global or local characteristics. Finally, classifiers are trained with the extracted facial features and used to test image samples. Quantitative metrics, such as accuracy or error rate, are used to assess the classifier performance.

In this work, we propose and evaluate a method for automatic gender recognition from face images based on a geometric descriptor constructed over a pre-defined face shape model. The method was evaluated on four different face datasets (FEI [8], AR Face [9], FERET [10] and Adience [11]) by using distinct cross-validation parameters. Superior results were achieved with our approach when compared against other methods that also employ geometric descriptors.

This paper is organized as follows. Section II briefly reviews some relevant concepts and techniques related to the topic under investigation. Section III describes the proposed gender recognition method. Section IV presents and analyzes the experimental results obtained with our method, as well as a comparison against other methods available in the literature. Section V concludes the paper with some final remarks and directions for future work.

II. RELATED WORK

The problem of gender recognition has been investigated over the last decades. This section briefly reviews some relevant related work available in the literature.

Golomb et al. [12] employed a neural network, denoted as SexNet, to classify gender from 90 human face images sampled at 30 × 30 pixels. Gutta et al. [13] employed radial basis function and decision trees for gender classification, where normalized images of 64 × 72 pixels, manually segmented, from FERET dataset [14] were used in their experiments.


Baluja and Rowley [18] developed an Adaboost classifier for identifying gender on images from FERET dataset with resolution 20 × 20 pixels. Li et al. [19], [20] proposed a gender recognition method based on motion extracted from human silhouettes segmented from a number of action components.

Alexandre [21] proposed a fusion scheme based on shape and texture features from multiple resolutions of normalized images (resized to 20×20, 36×36 and 128×128 pixels). Bekios-Calfa et al. [1] discussed the use of linear techniques, such as Linear Discriminant Analysis [22], for gender recognition.

Perez et al. [23] explored intensity, texture and shape features combined with mutual information. Results were evaluated on the FERET dataset. Dago-Casas et al. [24] conducted experiments with cross-database benchmarks for gender classification under unconstrained conditions.
Shan [25] developed a gender recognition approach based on a combination of texture features using local binary patterns (LBP) with an AdaBoost classifier. Experiments were conducted on the Labeled Faces in the Wild (LFW) benchmark [26].

Toews and Arbel [27] described a framework for detecting, localizing and classifying visual traits of objects from a viewpoint-invariant model derived from local scale-invariant features. A Bayesian classifier was employed to identify the visual traits. FERET dataset was used to evaluate the gender classification method.

Mansanet et al. [28] proposed a local deep neural network, denoted as local-DNN, which integrates local features and deep architectures. Experiments were conducted on the Labeled Faces in the Wild (LFW) [26] and the Gallagher’s dataset [29].

Fellous [30] investigated the use of 24 normalized horizontal and vertical distances calculated from fiducial points extracted from a set of 109 images. The model was trained on FERET dataset and other images acquired in his laboratory in order to predict the gender of various facial expressions.

Gupta [31] presented an approach to detecting gender of people through frontal facial images based on data mining and Delaunay triangulation techniques. Various classifiers, such as functional trees, random forests, naïve Bayes, AdaBoost and J48, were used to recognize a gender as male or female on the FEI Face dataset [8]. Patel et al. [32] developed a facial gender recognition method based on a compass local binary pattern descriptor, which was evaluated on CUFS and CUFSF datasets.

Levi and Hassner [33] developed an automatic gender classification using a convolutional neural network (CNN) to learn representations from the data and improve the performance of the classification task. Experimental results were evaluated on the Adience benchmark.

For further details on gender classification approaches, the reader can refer to [34], [35].

III. PROPOSED METHOD

A general scheme of our proposed gender recognition method is illustrated in Figure 1. Given a dataset containing images of faces of different men and women, an algorithm for detecting face fiducial points is applied to each image according to a pre-defined face shape model. In our work, the algorithm and the model provided by the Dlib toolkit [36] were used in this process. The face detector included in this toolkit is an implementation of the algorithm developed by Kazemi and Sullivan [37], which uses an ensemble of regression trees to quickly estimate facial landmarks from a sparse subset of pixel intensities.

The localization and indexing of such fiducial points is shown in Figure 2. Based on this model, the algorithm searches for a face on each image and, if it succeeds, extracts the exact localization of the pixels that correspond to each point of the model. The set of points can then be used to create an annotation file so that the face detection algorithm does not need to be executed again for new tests on the same dataset.

\[
\binom{68}{2} = 2278 \text{ distances.}
\]
After computing the descriptors for all images in the dataset, they are split into two different sets: one used for training and validation of a classifier and other for testing the prediction accuracy of the trained classifier.

For the training procedure, the \( k \)-fold cross validation was used. In this validation method, the entire dataset is first split into \( k \) parts (folds) of equal size. Then, one of these parts is left apart for the tests and the \( k - 1 \) remaining ones are used as the training set for the classifier. The test set is used to evaluate the accuracy of the trained model. This entire process is then repeated for the other folds such that each fold is used once as the test set, resulting in a total of \( k \) iterations. The final accuracy is then calculated from the average of the accuracy rates obtained in each iteration of the cross validation process.

A very important procedure in this method lies on the proper choice of the type of classifier and its tuning parameters used to maximize the prediction accuracy of the trained model. Since there are only two classes for the gender recognition problem (man and woman) and the classes of each image in the datasets are already known a priori, the Support Vector Machine (SVM) classifier [38] with a linear kernel was used in the classification process. Among the various possibilities of kernel functions [39], such as radial-basis function (RBF), polynomial and sigmoid, the linear kernel demonstrated to produce the highest accuracy.

The advantage of this approach is the robustness of the geometric descriptor, which produces very reasonable results regardless of color information or image size. Furthermore, the approach can be easily readapted to work with any face dataset and any image classifier. As a drawback, the proposed descriptor is relatively large when compared to other geometric descriptors, which slows the training process.

IV. EXPERIMENTAL RESULTS

The proposed method was implemented using Scikit-Learn [40], an open source Python library that contains several tools for machine learning, image processing, data mining and data analysis.

The result analysis is done by comparing our method against others that also employ geometric descriptors for gender recognition. The selected approaches for such analysis are: 24 Distances of Fellous [30] and Gupta Triangulation [31], which were detailed in Section II. The descriptors of both approaches were obtained from the points detected by Dlib’s face detection algorithm, however, only 19 out of the 24 distances could be extracted for the Fellous descriptor due to the absence of fiducial points on the forehead in the shape model.

The tests were conducted on four different face datasets: FEI [8], AR Face [9], FERET [10] and Adience [11]. The following subsections describe both of them, along with the respective results for each of the aforementioned approaches.

A. FEI Face Dataset

The FEI Face dataset [8] is provided by the Artificial Intelligence Laboratory of FEI university, located in Brazil. The original dataset contains 14 colored images of size 640 \( \times \) 480 pixels for each of the 200 individuals (100 men and 100 women) that voluntarily made part of this collection, with a total of 2800 images.

All pictures were taken against a white background in an upright frontal position and with different rotation angles. For this test, only a subset containing frontal images was taken. Here, each individual has exactly two frontal images, where one shows a neutral face expression and the other with a smiling face expression. Later, these images were cropped to a size of 360 \( \times \) 260 pixels. Therefore, this subset has a total of 400 images, with 200 men and 200 women.

Moreover, three different versions of this subset are provided by the authors:

- **Manually Aligned (MA):** images that were manually registered according to a common template in which the centers of the eyes and the nose of an individual are taken as a reference.
- **Spatially Normalized (SN):** grayscale images of size 300 \( \times \) 250 pixels where only the heads of each individual are framed in the respective images.
- **Spatially Normalized, Cropped and Equalized (SNCE):** grayscale images of size 193 \( \times \) 162 pixels where only the areas corresponding to each individual’s face are framed in the respective images. A histogram equalization process is also applied to each image.

Figure 3 shows some examples of images for each subset. In the case of SNCE dataset, the face detection algorithm was not able to detect a face on some images. Since the belonging images have a very restricted space for finding a face, the absence of extra information around the face makes the detection task even harder. Considering the model shown in Figure 2, the manual crops made for this dataset might have eliminated some of the points around the face (1-17), or even in at least one of the eyebrows (18-22 and 23-27), leading to all those misdetections. Hence, this dataset was reduced to a total of 280 images, with 66 men and 74 women.

B. AR Face Dataset

The AR Face dataset [9] contains over 4000 colored images of size 768 \( \times \) 576 from 126 individuals (70 men and 56 women). For each individual, images of 13 different facial expressions were taken, where some of them have two versions of each expression (26 images).

In this work, a subset of these images was created by taking only the images of neutral face expressions. As done for the FEI dataset, two versions of images for each individual were left in this subset. Thus, the final subset contains a total of 240 images from 120 individuals (66 men and 54 women). No transformations on the original images were made. Some examples of this dataset are shown in Figure 4.

C. FERET Dataset

The Face Recognition Technology (FERET) dataset [10] is a collection of 14126 images of size 768 \( \times \) 512 from 1199 individuals, encompassing several facial expressions. The
numbers of images for each individual vary, but every individual has at least two images with a neutral face expression.

Hence, for the gender recognition problem, all of the neutral face expressions were separated into “Male” and “Female” subsets according to annotations provided by the dataset. After generating the respective image annotations, the final sizes for each subset were, respectively, 1709 and 1001 images, resulting in 2710 images. Figure 5 illustrates some examples from these subsets.

D. Adience Dataset

The Adience dataset [11] was created by the Open University of Israel (OUI) to facilitate the study of both age and gender classification problems. The collection contains over 26580 images of 2284 subjects from 8 different age groups, from newborns to old-aged people. All photos were taken with several variations in appearance, posing, lighting, background and facial expressions. Besides the original images, a special version containing cropped and aligned face images is also available. In addition, the dataset provides information for a 5-fold cross-validation procedure, listing the images that make part of each split.

For this work, a subset containing only frontal face images (i.e., the ones within a ±5° range of yaw angle, according to the ground-truth information) was used. After running the face detection algorithm and generating the image annotations, the subset was reduced to a total of 11400 images. Examples of images from this subset can be seen in Figure 6.

E. Results for All Datasets

Table I shows the accuracies for all described methods and datasets. Three different cross-validation configurations were also tested: 5-fold, 10-fold and the leave-one-out cross validation (LOOCV), which is a particular case of the k-fold cross-validation where the number of folds is equal to the number of samples in the dataset. The only exception was the Adience dataset, where only the 5-fold variation was tested, since the split configuration is already provided.

From the table, it can be observed that our approach is
TABLE I

FINAL ACCURACY RATES OBTAINED FROM EACH METHOD FOR EACH DATASET, WITH DIFFERENT CROSS-VALIDATION SETTINGS.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FEI – MA</td>
<td>5-fold CV</td>
<td>91.25</td>
<td>80.25</td>
<td>59.75</td>
</tr>
<tr>
<td></td>
<td>10-fold CV</td>
<td>92.50</td>
<td>83.75</td>
<td>66.75</td>
</tr>
<tr>
<td></td>
<td>LOOCV</td>
<td>93.00</td>
<td>85.75</td>
<td>71.75</td>
</tr>
<tr>
<td>FEI – SN</td>
<td>5-fold CV</td>
<td>86.00</td>
<td>78.00</td>
<td>53.50</td>
</tr>
<tr>
<td></td>
<td>10-fold CV</td>
<td>86.25</td>
<td>79.75</td>
<td>60.00</td>
</tr>
<tr>
<td></td>
<td>LOOCV</td>
<td>87.75</td>
<td>82.00</td>
<td>66.25</td>
</tr>
<tr>
<td>FEI – SNCE</td>
<td>5-fold CV</td>
<td>79.29</td>
<td>81.79</td>
<td>29.29</td>
</tr>
<tr>
<td></td>
<td>10-fold CV</td>
<td>82.14</td>
<td>81.43</td>
<td>48.57</td>
</tr>
<tr>
<td></td>
<td>LOOCV</td>
<td>83.93</td>
<td>84.29</td>
<td>61.79</td>
</tr>
<tr>
<td>AR Face</td>
<td>5-fold CV</td>
<td>92.08</td>
<td>87.08</td>
<td>75.83</td>
</tr>
<tr>
<td></td>
<td>10-fold CV</td>
<td>96.67</td>
<td>90.42</td>
<td>78.75</td>
</tr>
<tr>
<td></td>
<td>LOOCV</td>
<td>97.50</td>
<td>89.58</td>
<td>82.50</td>
</tr>
<tr>
<td>FERET</td>
<td>5-fold CV</td>
<td>86.57</td>
<td>72.18</td>
<td>44.02</td>
</tr>
<tr>
<td></td>
<td>10-fold CV</td>
<td>89.19</td>
<td>77.45</td>
<td>50.48</td>
</tr>
<tr>
<td></td>
<td>LOOCV</td>
<td>90.34</td>
<td>81.40</td>
<td>67.31</td>
</tr>
<tr>
<td>Adience – Original 5-fold CV</td>
<td>72.50</td>
<td>65.24</td>
<td>55.38</td>
<td></td>
</tr>
<tr>
<td>Adience – Aligned 5-fold CV</td>
<td>75.14</td>
<td>67.35</td>
<td>57.57</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. Examples of frontal face images from the Adience dataset [11].

superior to the other methods for most of the cases. The exceptions were the 5-fold and the LOOCV for the SNCE dataset, where the Fellous distances method achieved the best results.

Despite the good results, our method still produces a very large descriptor size. Furthermore, the fact that the Fellous method did better in a few cases shows that, among all pairs of distances, a subset of distances which works at least as well as the original set may exist.

Actually, it is possible to apply specific algorithms to produce a compact descriptor with the most relevant features, especially regarding feature selection [41] and dimensionality reduction [42]. However, finding an optimal number of relevant distances is an exhaustive task. Some of these algorithms can be found in the Scikit-Learn library [40], but even after trying those algorithms with several numbers of features to select, the method produced worse results than the ones listed in Table I.

V. CONCLUSIONS AND FUTURE WORK

This paper described a novel approach to gender recognition from face images and proposed a geometric descriptor based on a pre-defined face shape model. The approach was tested on four face datasets with different cross-validation parameters, achieving the best results for most cases when compared against other methods that employed geometric descriptors as well.

Besides the need to create a more compact descriptor, as discussed in Section IV, another possibility of improving the classification results lies on the choice of the shape model used by the face detection algorithm. An in-depth study about the anthropometric differences between men’s and women’s faces may lead to the development of a more accurate model (possibly with less fiducial points) and, therefore, a more robust geometric descriptor can be created, even if it is strictly composed of distances between every pair of points.
Some future directions of this work include: use of different classifiers and supervised learning methods (along with a fine adjustment of their tuning parameters), addition of new face datasets, investigation of different face shape models and study of algorithms for choosing the most representative geometric features.

ACKNOWLEDGMENTS

The authors are thankful to FAPESP (grant #2014/12236-1) and CNPq (grant #305169/2015-7) for their financial support.

REFERENCES


