Pattern Recognition Letters 000 (2015) 1–9

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

On using periocular biometric for gender classification in the wild ☆

Modesto Castrillón-Santana*, Javier Lorenzo-Navarro, Enrique Ramón-Balmaseda

SIANI, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, 35017, Spain

A R T I C L E I N F O

Article history:
Available online xxx

Keywords:
Gender recognition
Periocular
Facial analysis
Soft biometrics

A B S T R A C T

Gender information may serve to automatically modulate interaction to the user needs, among other applications. Within the Computer Vision community, gender classification (GC) has mainly been accomplished with the facial pattern. Periocular biometrics has recently attracted researchers attention with successful results in the context of identity recognition. But, there is a lack of experimental evaluation of the periocular pattern for GC *in the wild*. The aim of this paper is to study the performance of this specific facial area in the currently most challenging large dataset for the problem. As expected, the achieved results are slightly worse, roughly 8 percentage points lower, than those obtained by state-of-the-art facial GC, but they suggest the validity of the periocular area particularly in difficult scenarios where the whole face is not visible, or has been altered. A final experiment combines in a multi-scale approach features extracted from the periocular, face and head and shoulders areas, fusing them in a two stage ensemble of classifiers. The accuracy reported beats any previous results on the difficult *The Images of Groups* dataset, reaching 92.46%, with a GC error reduction of almost 20% compared to the best face based GC results in the literature.

1. Introduction

Humans exhibit an extraordinary ability to extract information from facial images that serves for multiple purposes in daily human interaction. The reliability of the human system has attracted the attention of psychologists for decades. More recently, Computer Vision researchers are interested in developing automatic systems with similar skills. In this work, we specifically focus on gender classification (GC).

Different biometric traits have been applied to the GC problem as gait, hair or clothes, as described in [6,7,17,22,42]. However, most automatic GC systems have been and are being designed to determine the human gender from the facial pattern, as evidenced in the recent NIST survey by Ngan and Grother [30]. The face is indeed a valid cue for humans, but a deeper analysis of the human system highlights the main importance of the ocular and mouth areas, as suggested by Gosselin and Schyns [20]. In this paper, we analyze the significance of the ocular area for this particular problem, but in the wild.

The recent survey on ocular biometrics by Nigam et al. [31] reviews the results achieved making use of three different modalities: iris, periocular and retina. Among them, we refer below to the first two, the external ones. The human iris is certainly a powerful biometric trait as proved in identity recognition. However, to be reliable, iris based systems require user cooperation and some specific acquisition conditions related to the pattern resolution. A less restricted alternative in terms of user cooperation, is the periocular area, i.e. the eye and its surrounding area. There is an increasing interest in the still relatively unexplored field of periocular biometrics, with recent promising and successful results in identity recognition.

The mentioned survey by Nigam et al. [31] also describes the advantages of modalities fusion. Indeed, the fusion of facial and periocular information improves identity recognition performance, proving the benefits in surveillance scenarios, where noise and non-cooperation excludes the use of other ocular cues.

Our hypothesis is that the same behavior will be observed in GC even in the wild. Thus, we first study automatic GC based on only the periocular region. This is done analyzing the performance of different local descriptors achieved in the currently hardest dataset. Once compared to face based approaches we evaluate the fusion of both modalities: periocular and face.

The contributions of the work are three-fold: (1) an exhaustive evaluation of local descriptors in the currently most challenging dataset for the GC problem, proving that the periocular is a valid cue for large populations; (2) different descriptors provide complementary information, that serves to setup a score level fusion of selected descriptors to improve the overall periocular based GC performance; and (3) the confirmation that the multi-scale combination of selected features extracted from the periocular, face, and head and shoulders areas steps forward a more accurate gender recognizer.

☆ This paper has been recommended for acceptance by Maria Frucci (guest editor) and Maria De Marsico (managing guest editor).

* Corresponding author. Tel.: +34 928 458 743; fax: +34 928 458 711.
E-mail address: mcastrillon@iusiani.ulpgc.es, modesto.castrillon@ulpgc.es (M. Castrillón-Santana).

http://dx.doi.org/10.1016/j.patrec.2015.09.014
0167-8655/© 2015 Elsevier B.V. All rights reserved.

Fig. 1. Sample images respectively of *The images of Groups* [18], *The Labeled Faces in the Wild* [21] and MORPH [36] datasets. Their respective original resolutions are 391 × 293, 249 × 249 and 200 × 240 pixels.

Table 1
Facial based GC state-of-the-art accuracies in large datasets. [1] Inter ocular distance > 20, [2] 22, 778 aut. detected faces, [3] 7443 of the total 13, 233 images, [4] BEFIT protocol, [5] balanced subset with 14, 244 of the total 55, 134 images.

Reference	Dataset	Accuracy (%)
[8]	GROUPS[1]	89.8
[12]	GROUPS[2]	90.4
[9]	GROUPS	90.8
[38]	LFW[3]	94.8
[40]	LFW[3]	98.0
[15]	LFW[4]	97.2
[23]	LFW	96.9
[35]	MORPH[5]	97.1

2. State of the art

2.1. Facial GC

The 2015 NIST evaluation by Ngan and Grother [30] remarks the differences between constrained and unconstrained or in the wild GC. The best evaluated system was able to reach an accuracy up to 96.5% in a constrained dataset containing around one million samples. However, those accuracies dropped significantly for some particular in the wild datasets. Two remarkable smaller datasets were analyzed: *The Labeled Faces in the Wild* (LFW) by Huang et al. [21], and *The images of Groups* (GROUPS) by Gallagher and Chen [18]; reporting a similar accuracy for LFW, while for GROUPS the best commercial solution dropped to an accuracy of 90.4%. This effect is also confirmed by the research community. Table 1 summarizes recent results on three large datasets, adding MORPH, by Ricanek and Tesafaye [36], to the mentioned GROUPS and LFW. A sample of images of each dataset is presented in Fig. 1. Observing Table 1, GROUPS achieves an accuracy that hardly reaches 90%. Compared to the other datasets, there is an evident gap between the accuracy achieved for GROUPS and the rest, being almost 7 points lower.

GROUPS contains 28, 000 labeled faces, but a detailed analysis of its characteristics reveals some dataset difficulties: (1) contrary to the other two datasets most individuals present a single image in the collection; (2) the mean inter-eye distance, 25 pixels, is respectively the half and the quarter than those present in LFW and MORPH faces; and (3) the subject pose and the illumination conditions are much less controlled.

2.2. Ocular GC

Compared to the face, that has indeed large tradition among the facial analysis community, the ocular region has certainly received less attention in the GC problem. This is however not the case of identity recognition based on the periocular area, that aims at exploiting useful non-iris features of that area as suggested by Ambika et al. [3] and Santos and Proenca [37]. In this sense, there are already interesting results with challenging datasets as reported by Mahalingam and Ricanek [28]. Compared to the face trait, there is a significant accuracy in identity recognition using remarkable less facial information, as argued by Bakshi et al. [4]. Different authors claim that ocular based systems are able to cope with challenging scenarios where the whole face is not available, e.g. partially occluded, or even when available, it has been altered with plastic surgery, as evidenced by Jillela and Ross [24], Merkow et al. [29] and Park et al. [32].

Centering on the GC problem, iris may be an alternative to facial pattern, as pointed out by Tapia et al. [41]. However, these systems require higher resolution images and more user cooperation, limiting the application scenarios.

There are however some results of periocular GC, that we briefly summarize here. The validity of the ocular area for GC was already suggested in video processing, as described in [11]. More recently similar conclusions have been achieved with image datasets. In [29] an experiment was carried out with a collection of less than 1000 faces reaching a GC accuracy of 85% with local binary patterns (LBP) features.

A slightly larger dataset is considered in [27], where 4232 images of 404 subjects were analyzed in terms of gender and ethnicity, reaching for gender up to 97% of accuracy with local descriptors, comparable results to using the facial region. Certainly, the dataset contains high quality images (ocular area of 250 × 250 pixels).

A slightly harder scenario is tackled in [25] where the FERET dataset, see [33], is used with a periocular region of 74 × 88 pixels. However, the final dataset evaluated contains just 200 images reaching an accuracy around 90%. Certainly this dataset has already achieved very high GC rates with facial features (close to 100%), see [2,5].

Even if the reported results are promising, we conclude that the literature on periocular GC is limited, as the existing results were obtained with reduced datasets, and/or high resolution images. We therefore argue the interest of the analysis in the wild, where we compare the periocular GC performance with state-of-the-art facial GC systems on the currently hardest dataset, i.e. GROUPS.

3. System overview

The proposed system works with periocular images. This area is extracted after normalizing the face in terms of scale and rotation. Given the rough eye location annotations, the normalized facial image is obtained automatically after rotating, scaling and cropping the original images to force the ground truth eye locations to the specific eye positions (16, 17) and (42, 17), where the inter-eye distance is of