Pattern Recognition Letters 31 (2010) 1422–1427

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Gender recognition: A multiscale decision fusion approach

Luís A. Alexandre *

Department of Informatics, Univ. Beira Interior, IT – Instituto de Telecomunicações, R. Marquês d'Ávila e Bolama, 6201-001 Covilhã, Portugal

ARTICLE INFO

Article history:
Received 8 June 2009
Received in revised form 16 November 2009
Available online 23 February 2010
Communicated by H.H.S. Ip

Keywords:
Gender recognition
Decision fusion
Histogram of oriented gradients
Local binary patterns

ABSTRACT

Gender recognition from face images has many applications and is thus an important research topic. This paper presents an approach to gender recognition based on shape, texture and plain intensity features gathered at different scales. We also propose a new dataset for gender evaluation based on images from the UND database. This allows for precise comparison of different algorithms over the same data. The experiments showed that information from different scales, even if just from a single feature, is more important than having information from different features at a single scale. The presented approach is quite competitive with above 90% accuracy in both evaluated datasets.

1. Introduction

Gender recognition from face images has received much attention recently (Baluja and Rowley, 2007; Mäkinen and Raisamo, 2008a,b; Mayo and Zhang, 2008; Lu et al., 2008). This is justified by its many applications: improving search engine retrieval accuracy, demographic data collection, human–computer interfaces (adjusting the software behavior with respect to the user gender), etc. Moreover, in a biometric recognition framework, gender classification can help by requiring a search of only half the subject database.

An issue, up until recently, has been the difficulty in objective comparison between different gender recognition approaches since much of the published work was evaluated in non-replicable datasets.

In (Mäkinen and Raisamo, 2008a) an effort was made towards improving the standards in the field by making available the details of one of the datasets used. This allowed Mayo and Zhang (2008) to compare against the same dataset and we also do this in this paper. We further propose another dataset using the UND database (Flynn et al., 2003) and we make available the information regarding the images and the crop sizes used so that this set can also be used for objective evaluation of gender recognition algorithms.

Gutta et al. (1998) presents one of the first gender recognition proposals that used a large dataset. They worked with FERET images and used an ensemble of RBF networks and C4.5 decision trees obtaining a 96% accuracy on a dataset of 3006 images of 1009 subjects using cross validation. The same person could appear in both training and test sets which meant that the system could be learning to recognize faces instead of gender. Images were manually segmented and normalized to 64×72 pixels.

Moghaddam and Yang (2002) also used the FERET database. They used an automatic face-processing system which compensated for translation, scale and slight rotations. They subsampled images to 21×12 pixels. The presented accuracy was quite high (96.62%) but the experiments were done using several images from the same subjects thus having the same problem as the work of Gutta et al. (1998). Evaluation used a 5-fold cross validation scheme.

In (Mäkinen and Raisamo, 2008a), the authors main goal was to see how did face alignment influence the accuracy of gender classification methods. They tested 5 different alignment approaches and no alignment. They conclude that the best results were obtained without automatic alignment. But better results could be obtained with manual alignment. They also tested three image sizes: 24×24, 36×36 and 48×48. The best results (87.1% accuracy) were obtained with 36×36 pixel size.

Augmenting the training set can improve generalization as Mayo and Zhang (2008) found. They showed that by introducing rotated and translated versions of the original images, the gender recognition accuracy could be increased: they improved the results of Mäkinen and Raisamo (2008a) in the FERET dataset from 87.1% to 92.50%.

An AdaBoost based approach 50 times faster than using SVMs, was applied by Baluja and Rowley (2007) to images of the FERET database. The best accuracy reported was 94.40% on 20×20 pixel images, with normalization similar to the one in (Moghaddam and Yang, 2002).

* Tel.: +351 275319891; fax: +351 275319899.
E-mail address: lfbaa@di.ubi.pt

0167-8655/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2010.02.010

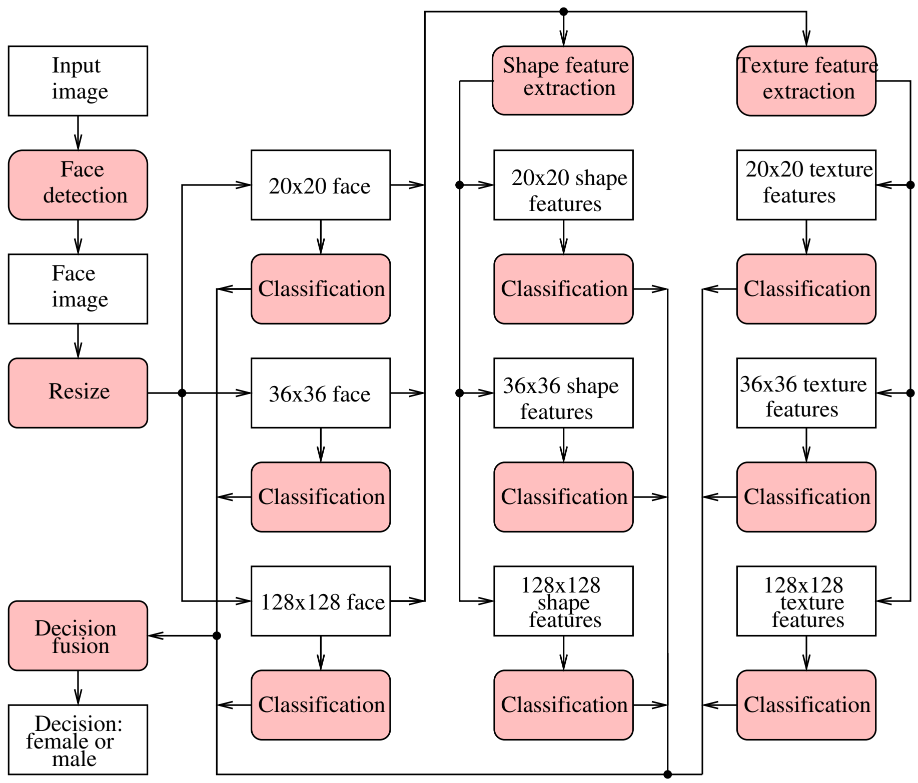

Fig. 1. The proposed multiscale decision fusion approach. White boxes represent data and gray boxes represent operations.

In (Zafeiriou et al., 2008), a variant of SVM inspired on the optimization of Fisher's discriminant ratio is applied to gender recognition using only luminance information and automatic aligned images. The size of face segmented images is 85×156. Images (total of 2360) are from the XM2VTS database. Fivefold cross validation was used to compared the proposed method with standard SVM and complete kernel Fisher discriminant. The proposed method presented better results with error rates of 2.86%.

In (Xu et al., 2008), a hybrid method using local features (10 features extracted with Active Appearence Model) and global features (extracted using AdaBoost with Haar-like features) is proposed. The authors show that a better accuracy can be obtained by fusing these features before classification. Overall accuracy using 5-fold cross validation on 1000 images is 92.38%.

In (Lu and Shi, 2009), a method fusing three face regions (left eye, upper face region and nose) is proposed. Experiments on FERET images show that fusion outperforms individual features and using CAS-PEAL images it was alo shown that the fused results improve on the whole face approach.

In this paper we present an approach that is capable of improving accuracy on the FERET dataset defined in (Mäkinen and Raisamo, 2008a) and we propose a new dataset for objective evaluation of gender recognition accuracy based on images from the UND Collection B database.

The rest of the paper is organized as follows: the next section contains a description of the proposed multiscale approach including a description of the features used; Section 3 describes the proposed dataset; the following section presents the experiments and the final section contains the conclusions.

2. A multiscale approach

In this paper we propose the use of a multiscale approach for gender recognition. The idea is to extract features at different image resolutions, obtain a classification based on these features and fuse the decisions obtained.

Fig. 1 contains a sketch of the proposed approach. Note that the boxes labeled 'Classification' represent several classifiers, each receiving their input from only one of the feature types (intensity, shape or texture) collected at a particular scale. We will investigate in the experiments section how different combinations of the decisions obtained from these types of features and scales perform on this problem.

The rationale behind this comes from the supposition that the gender discrimination accuracy can be improved by fusing decisions based on features from different image resolutions. Our experiments reported below backed this assumption.

Image description is done traditionally with three types of descriptors: color, shape and texture. Since color is not essential for gender determination from face images (humans can perform this task quite well on grayscale images), the images used in this paper are grayscale.

2.1. Shape features

The shape features used are histograms of edge directions, much like the histograms of oriented gradients in (Dalal and Triggs, 2005). The main differences are that, in our case, the input images are grayscale and no normalization to the histograms is performed.

The process is now described. First vertical and horizontal edge maps are found using the following masks: $[-1, 0, 1]$ and $[-1, 0, 1]^T$. Consider v and h to be the vertical and horizontal edge values at a pixel, respectively, obtained by convolution of an edge detector filter with the original image. Then, the edge map direction is found using

$$\theta = tan^{-1}\left(\frac{v}{h}\right) \quad (1)$$

and the edge magnitude (or intensity) is obtained with