Automatic Gender Recognition Using Fusion of Facial Strips

Ping-Han Lee
Department of Computer Science and Information Engineering, National Taiwan University

Jui-Yu Hung
Graduate Institute of Networking and Multimedia, National Taiwan University

Yi-Ping Hung
Graduate Institute of Networking and Multimedia, National Taiwan University

Abstract

We propose a fully automatic system that detects and normalizes faces in images and recognizes their genders. To boost the recognition accuracy, we correct the in-plane and out-of-plane rotations of faces, and align faces based on estimated eye positions. To perform gender recognition, a face is first decomposed into several horizontal and vertical strips. Then, a regression function for each strip gives an estimation of the likelihood the strip sample belongs to a specific gender. The likelihoods from all strips are concatenated to form a new feature, based on which a gender classifier gives the final decision. The proposed approach achieved an accuracy of 88.1% in recognizing genders of faces in images collected from the World-Wide Web. For faces in the FERET dataset, our system achieved an accuracy of 98.8%, outperforming all the six state-of-the-art algorithms compared in this paper.

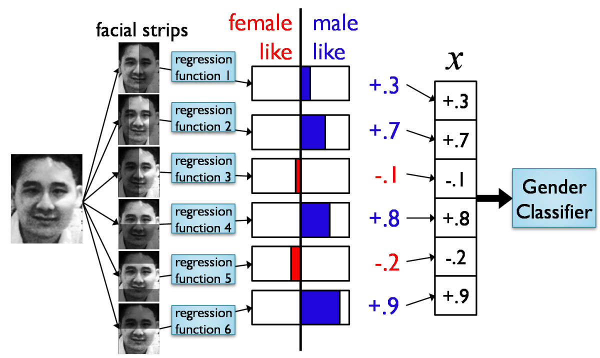

Figure 1. Gender recognition using fusion of facial strips. A face is first decomposed into several horizontal and vertical strips. A regression function for each strip gives the likelihood that the strip patch belongs to a specific gender. The likelihoods from all strips are concatenated to form a new feature (x), based on which a gender classifier carries out the final decision.

1 Introduction

Recently, gender recognition using facial images has attracted a great deal of attention. [7] used Support Vector Machines (SVMs) for appearance-based gender classification. In [4], a gender classifier based on Fuzzy SVM (FSVM) is developed to improve the generalization ability. Besides SVM-based approaches, Viola et al. proposed an Adaboost-based gender classification scheme [11]. The three algorithms used whole faces as features.

To circumvent the problem caused by pose, expression, or illumination variations, some have proposed approaches based on *local facial regions*. [1] extracted a normalized feature vector formed by matching N local regions of the face against some fixed set of M face images using the FaceIt® algorithm. [8] presented a hierarchical feature fusion model constructed by an evolutionary learning algorithm. The locations, widths and heights of local regions were automatically determined during learning. [5] defined seven local facial regions, which are inner face, upper and lower faces, left eye, nose, mouth, and the whole face itself. Three most significant regions were selected and the decision-level fusion strategy was applied to yield the final gender classification results.

In this paper, we propose an automatic gender recognition scheme based on local facial regions. The local facial regions used in our work are different from all the aforementioned approaches. Instead of either selecting local regions automatically or manually, we extract vertical and horizontal *facial strips* systematically. Unlike the previous region-based approaches which fuse local regions in *decision-level* or *score-level*, we measure the likelihood of each local region belonging to a certain gender and concatenate these likelihoods to form new features. Figure 1 illustrates the proposed scheme. In addition, a face preprocessing is employed which cor-

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.285

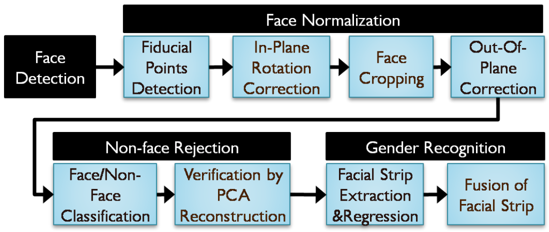

Figure 2. System overview.

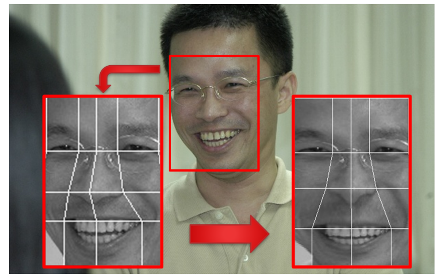

Figure 3. Correcting the in-plane and the out-of-plane rotations given a face detected in an image.

rects the in-plane and out-of-plane rotations of faces, and aligns faces according to the estimated eye positions. This scheme boosts the final gender recognition by 6% in our experimental study.

Figure 2 shows the overview of the proposed system. The system can be divided into four major modules: the *Face Detection* module, the *Face Normalization* module, the *Non-Face Rejection* module and the *Gender Recognition* module. The **Face Detection** module detects faces in images and it applies the standard *Viola-Jones face detector* [10]. In the following, Section 2 describes the Face Normalization and Non-Face Rejection modules. Section 3 introduces the proposed gender recognition algorithm using facial strips. A comprehensive experimental study is given in Section 4, followed by the conclusion and the future work in Section 5.

2 Face Preprocessing

2.1 Face Normalization

Viola-Jones face detector gives rough positions and regions of human faces. The proposed normalization module gives more accurate face positions by aligning their eyes. Given a face detected by the Viola-Jones detector, we first detect salient facial feature points, or *fiducial points*, using a latest extension of the Active Shape Model [6] (ASM), which used two-dimensional landmark templates and stacked two ASMs in series. [6] gives 68 fiducial points, based on which we normalize a face as the following:

1. Calculate the eye centers, and rotate the face to make the line passing through eye centers horizontal. This step corrects the *in-plane rotation*.
2. Crop the face from the image in the way that the eye positions are the same for all the cropped faces.
3. Warp the fiducial points such that points on the left half of the face and those on the right half of the face are *symmetric*. This step corrects the *out-of-plane rotation*.

Figure 3 illustrates an example of normalizing and cropping a face in a photo.

2.2 Non-Face Rejection

There exists some non-face image patches that the Viola-Jones face detector wrongly detects them as a face and the Active Shape Model wrongly locates some fiducial points on them. To reject these non-faces, two schemes are applied: (1) we first employ a Support Vector Machine [3] to classify *face candidates* detected by the Viola-Jones detector into faces and non-faces; (2) for those classified as faces, we reconstruct them using *eigen-components* of human faces and calculate the difference between the original images and the reconstructed ones. We reject those with large reconstruction errors.

3 Gender Recognition Using Facial Strips

In general, using multiple local facial regions yield better accuracies than using the whole faces alone in facial recognition. Local facial region approaches typically involve two steps: (1) the definition of facial regions, and the (2) fusion of facial regions. Most existing algorithms selected a bunch of local regions, either automatically or manually. Each region has its own gender classifier. Given a face, region-specific classifiers make their own classifications, and these classification results are *fused* to give an overall decision. Two typical fusion strategies were applied in most of the previous works: the *decision-level fusion* applies the *majority rule*; the *score-level fusion* sums up some distance measures given by region-specific classifiers. The gender with a smaller overall distance is the result.

We have carried out an extensive experimental study on region-based gender recognition and had the following observations: (1) Human-perceivable facial regions are not necessary good in discriminating genders;

set	λ_x	λ_y	t_x	t_y	# strips
A	W/2	H/2	W/4	H/4	6
A_2	W/2	H/2	W/8	H/8	10
A_3	W/2	H/2	W/16	H/16	18
B	W/4	H/4	W/8	H/8	14
C	W/8	H/8	W/16	H/16	30

Table 1. Definitions of facial strip sets used in this study.

(2) Smaller regions contains less information, and they are sensitive to the misalignment of faces; (3) All facial regions contribute in facial gender recognition. Using only partial regions results in a lower accuracy; (4) Some facial regions are more effective in discriminating genders than others, and should be assigned larger weights. However, weighting facial regions in both decision-level and score-level does not improve the accuracy. These observations lead to the development of a new region-based gender recognition approach with novel regions extraction and fusion scheme. The proposed scheme is illustrated in Figure 1. In the following, Section 3.1 and 3.2 describe the region extraction and fusion scheme respectively.

3.1 Facial Strip

A **facial strip** is a strip of face, and it can be a *horizontal strip* or a *vertical strip*. A facial strip can be written as $\{x, y, w, h\}$, where $\{x, y\}$ is the upper left corner of this strip, $\{w, h\}$ is its width and height. Let the width and the height of the face be W and H, then a horizontal and a vertical strip can be written as $\{1, y, W, \lambda_y\}$ and $\{x, 1, \lambda_x, H\}$, respectively, where λ_x and λ_y can be regarded as the *bandwidth* of the facial strip. To extract horizontal strips on a face, one can start with $\{1, 1, W, \lambda_y\}$, sliding it downward on the face and crop a new strip every t_y pixels. Using this scheme, a **uniform facial strip set** can be written as $\{\lambda_y, t_y, \lambda_x, t_x\}$. Table 1 gives examples of some uniform facial strip sets. The six facial strips of the set A in Table 1 are shown in Figure 1.

3.2 Fusion of Facial Strips

For each facial strip, we construct a *regression function* $s : x \rightarrow [-1, 1]$ to minimize squared error, $Err_{x,y\sim D}(s(x) - y)^2$ given a set of training data D, where x is the feature vector and y is either '-1' or '1', standing for 'male' and 'female' respectively. To recognize the gender of a given facial image, each regression function gives its own estimation, and estimations of all the regression functions are concatenated to form a new feature vector. Based on this feature vector, an overall classifier is trained to classify it into a male or a female (see Figure 1).

normalize face	image size	subspace learning	feature dim.	accuracy (%)
no	24x24	NONE	576	79.0 [11]
	20x25		500	78.0
yes				84.0
		PCA	130	85.0 [7]
		LDA	1	82.0

Table 2. Comparison between whole-face based approaches. The 'normalize face' indicates faces were pre-processed by the procedures described in Section 2.1.

The features of the proposed approach are: (1) Facial regions are defined systematically instead of manually assigned; (2) No hard decision is made by any facial region; (3) All the facial regions contribute in the final decision, and (4) their weights are learned implicitly by the overall gender classifier.

4 Experimental Results

We collected 2229 images from the World-Wide Web, and our system detected more than 10,000 human faces in these images. Most of faces have age between 20 and 45, while some belong to children and elders. 3,000 images composed of equal amount of *yellow male*, *yellow female*, *white male*, *white female*, were randomly selected to form *training set*. Another 3,000 images with the same make up were selected to form the *testing set*. Table 4 gives the results of whole-face approaches. We can see that via normalizing faces, the recognition rate improves from 78% to 84%. Applying the Principal Component Analysis (PCA) yields a slightly better rate of 85%, while applying the Linear Discriminant Analysis (LDA) does not help.

Table 3 gives the results of region-based approaches using different facial regions and fusion schemes. For the proposed approach, the regression function for each strip and the overall gender classifier are constructed using the ε-SVR and C-SVC respectively provided by the libsvm [3]. Some conclusions can be drew from this table: (1) Facial strips (set A, A_2 and A_3) yields better results than manually defined facial regions (set M); (2) Score-level fusion is better than decision-level fusion. The proposed fusion scheme outperforms the two strategies. (3) Sets using smaller facial strips (set B and C) yield lower accuracy than the one using larger facial strips (set A); (4) Comparing sets A, A_2 and A_3,

set	#regions	decision	score	proposed
M	7	85.0 [5]	85.3	86.7
A	6	85.8	87.4	87.8
A_2	10	86.2	87.6	**88.1**
A_3	18	86.8	87.7	88.0
B	14	85.4	86.7	86.6
B'	10	85.0	86.0	86.3
C	30	85.5	85.9	86.5
C'	22	85.5	86.1	86.3

Table 3. Comparison between region-based approaches using different facial regions and fusion schemes. Facial regions are cropped from facial images of size 80x100. Sets A, A_2, A_3, B, C are defined in Table 1. The 7 facial regions in set M is similar to those used in [5]. Set $B'(C')$ is the subset of $B(C)$, and it contains 10(22) best strips in terms of classification accuracy. The 'decision' and 'score' stands for decision-level and score-level fusion scheme respectively. The 'proposed' applies the strip-specific regression functions.

algorithm	M|F	accuracy
Moghaddam 2002 [7]	1044|711	96.6
Tivive 2006 [9]	1150|612	96.5
Baluja 2007 [2]	1495|914	94.4
Leng 2008 [4]	160|140	98.1
proposed	1158|615	**98.8**

Table 4. Comparison between the proposed approach and four state-of-the-art algorithms on the FERET dataset. 'M' and 'F' stands for the number of the male and female images respectively.

we can see sets using the same size of facial strips but smaller *step sizes* (t_x and t_y) yield slightly higher accuracies, but this benefit saturates when the step size is too small compared with the size of the facial strip. (5) The results for B' and C' are slightly worse than B and C respectively. It suggests that using the set of strips covering the whole face area is a better strategy.

Table 4 compares the proposed approach with the state-of-the-art algorithms on the FERET dataset. In this table, [9] applied a convolutional neural network; [2] is an Adaboost-based approach. Although the numbers of test images were different for algorithms in Table 4, these algorithms all used the *fa* partition in the FERET dataset. Our algorithm also used the *fa* partition. The proposed algorithm achieves the best accuracy.

5 Conclusion and Future Work

In this paper, we propose an automatic system that detects and recognizes faces in images. The contributions of this paper are twofold: (1) we propose a face pre-processing scheme that corrects the in-plane and out-of-plane rotations of faces, and aligns faces based on the estimated eye positions. This scheme boosts the gender recognition accuracy by 6%; (2) we propose the facial strip and a novel fusion scheme for gender recognition. We have demonstrates experimentally that facial strips are better than manually defined facial regions, and the proposed fusion scheme outperforms two commonly used schemes. The proposed approach is validated on both images from World-Wide Web and the FERET dataset. It outperforms all six state-of-art algorithms evaluated in this study. Our future work will apply the facial strips to face recognition.

Acknowledgement

This work was supported in part by the Ministry of Economic Affairs, Taiwan, under Grant 98-EC-17-A-02-S1-032 and by the National Science Council, Taiwan, under grant NSC 98-2221-E-002-127-MY3.

References

[1] C. B. Abdelkader and P. Griffin. A local region-based approach to gender classi.cation from face images. In *CVPRW*, pages 52–52, June 2005.

[2] S. Baluja and H. A. Rowley. Boosting sex identification performance. *IJCV*, 71(1):111–119, 2007.

[3] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[4] X. Leng and Y. Wang. Improving generalization for gender classification. In *ICIP*, pages 1656–1659, Oct. 2008.

[5] L. Lu and P. Shi. A novel fusion-based method for expression-invariant gender classification. In *ICASSP*, pages 1065–1068, April 2009.

[6] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. In *ECCV*, pages 504–513, Berlin, Heidelberg, 2008. Springer-Verlag.

[7] B. Moghaddam and M.-H. Yang. Learning gender with support faces. *PAMI*, 24(5):707–711, 2002.

[8] F. Scalzo, G. Bebis, M. Nicolescu, L. Loss, and A. Tavakkoli. Feature fusion hierarchies for gender classification. In *ICPR*, pages 1–4, Dec. 2008.

[9] F. H. C. Tivive and A. Bouzerdoum. A shunting inhibitory convolutional neural network for gender classification. In *ICPR*, volume 4, pages 421–424, 0-0 2006.

[10] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, 2004.

[11] P. Viola, G. Shakhnarovich, G. Shakhnarovich, P. A. Viola, B. Moghaddam, and B. Moghaddam. A unified learning framework for real time face detection and classification. In *FGR*, 2002.