Gender Recognition Based on Multi-model Information Fusion

Guo Shiang Lin[1] Min Kuan Chang[2] Yu Jui Chang[2]
[1]Dept. of Computer Science and Information Engineering, Da Yeh University, Chang-Hua
khlin@mail.dyu.edu.tw
[2]Dept. of Electrical Engineering, National Chung Hsing University, Tai-Chung
minkuanc@dragon.nchu.edu.tw and changyr@wmc.ee.nchu.edu.tw

***Abstract*—In the paper, we proposed a gender recognition scheme based on multi-model information fusion. The proposed gender recognition scheme is composed of four parts: face detection and rectification, eye detection, feature extraction, and gender classifier. To evaluate the proposed scheme, a large number of images containing different-size faces are captured by using low-cost webcam. Experimental results show that our proposed scheme can detect facial regions as well as eyes well. In addition, the accuracy of our gender recognition scheme is more than 95%. These results demonstrate that our proposed scheme can achieve not only face and eye detection but also gender recognition.**

I. INTRODUCTION

So far, gender recognition has been interesting in many application fields such as smart human identification, human computer interface, facial expression recognition, emotion repression, and age estimation. For example, a gender recognition system can provide customers' gender information and adaptively play suitable advertisements. There are currently several kinds of gender recognition methods. The gender recognition methods can be classified into four classes: gait-based [9], face-based [10],[13],14], hand-based [11], and voice-based [12]. This means that there is not a unique or generic solution to the gender recognition problem.

As we know, there is a variety of information in human faces and the information shown in faces can indicate social interactions among people. Each person can process a face in some ways to not only discern who it is but also categorize demographic characteristics, e.g., gender, age, and ethnicity. This means that the face is an important biometric feature of human beings. This is why we aim at developing a face-based gender recognition scheme.

There are some face-based gender recognition methods [10],[13],14]. For example, Hadid et al. [14] extracts features from whole face by LBP (local binary pattern) operator. They first scaled face image into 3 difference resolutions: 20×20, 40×40 and 60×60 pixels. Then, they extracted features based on original LBP operator and VLBP (volume local binary pattern) operator in local region of sizes from 10×10 to 20×20 pixels. However, most of gender recognition methods process images with controlled lighting condition and non-textured background. In real environments, e.g., surveillance videos, the resolution of images captured from mobile phone and surveillance system may not be high. It is expected that the information in faces obtained from low-resolution images may not be easily analyzed for gender recognition.

As mentioned above, it motives us to develop a face-based gender recognition scheme to deal with low-resolution uncontrolled images. The remainder of this paper is organized as follows. Section II describes the proposed gender recognition system. Section III shows experimental results. Section IV draws some conclusions.

II. PROPOSED GENDER RECOGNITION

Figure 1 illustrates the block diagram of our proposed gender recognition scheme, which is composed of face detection and rectification, eye detection, feature extraction, and gender classifier.

The first step of gender recognition is face detection. Since each face may not be upright in images, it is necessary to rectify each detected face for improving the accuracy of gender recognition. In many applications such as face recognition and emotion expression, eyes are one of the most important parts in face images. To improve the accuracy of face detection and gender recognition, we develop an eye detection algorithm. To effectively achieve gender recognition, we select some useful features combined with a classifier. To effectively fuse these features, a gender classifier based on multi-model information fusion is developed.

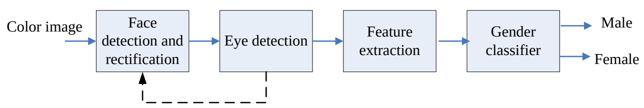


Fig. 1 Block diagram of proposed gender recognition scheme

A. *Face detection and rectification*

Since face detection can provide a rough output, most gender recognition methods think of the detected patches as desired facial regions. However, the performance of gender recognition is limited to the accuracy of face detection and rectification.

On one hand, some faces in images may not be upright, features may not be extracted from important parts of faces. On the other hand, since the light may not be uniform in images, this situation makes feature extraction difficult. Therefore, face rectification and lighting adjustment are necessary in the proposed scheme before feature extraction.

To detect facial regions, a face detector using Haar-like features trained by Adaboost algorithm [1] is adopted. To remove some errors of face detection, the information of eyes is also used. We will introduce eye detection later.

To make faces upright, a symmetric property of a face is used to rectify a rotated face image. Since SSIM (structure similarity index) [3] can measure the structural similarity of two images, we use SSIM as a measurement to evaluate the structural similarity of a face image. It is expected that SSIM should be large when a frontal face image is upright. Then, for a face region R_i, its rotation angle θ_{R_i} can be obtained as

$$\theta_{R_i} = \arg\max_{\theta}\left(S_{R_i}\right), \theta \in \left\{-30^{\circ},\ldots,30^{\circ}\right\}, \tag{1}$$

where S_{R_i} is the value of SSIM for R_i. After obtaining θ_{R_i}, we can rotate the face image to be upright.

To reduce the impact of lighting conditions on gender recognition, a lighting adjustment algorithm [2] is used. In [2], since the perception of human beings is nonlinear for the light, the pixel values should be adjusted by using a nonlinear function. This means that pixel values in a dark region increase and pixel values in a bright region decrease.

After face rectification and lighting adjustment, each processed face region is also normalized into a rectangle one with 100×100 pixels.

B. Eye detection

After rectifying faces, we develop a feature-based eye detection algorithm. In the feature-based eye detection algorithm, texture features based on LBP [4] is calculated as inputs of a SVM (support vector machine) classifier.

In Section II.A, normalized face images are obtained after face detection, face rectification, and lighting adjustment. For a normalized face image, the steps of eye detection are described as follows:

(1) Compute the LBP features at one pixel of a search area.
(2) Determine whether this pixel is within an eye by using a trained SVM classifier. If yes, an array A^E adds by one, i.e., $A^E(x,y) = A^E(x,y)+1$.
(3) Repeat Steps (1) and (2) until each pixel in the search area is tested.
(4) Find the locations of the peaks in A^E as the locations of eyes.

After detecting eyes, the information of detected eyes is used to eliminate some errors of face detection. As we know, the probability that there exists a face is high when there are two detected eyes in a face region.

C. Feature extraction

After obtaining the locations of eyes and internal face, we adopt three kinds of features for gender recognition. In addition, according to [6], internal face and eyes are the important part of a face. The features extracted from internal face and eyes should be useful for gender recognition. In addition, the hair information is also a useful feature to distinguish women from men. Therefore, the three kinds of features are adopted in the paper.

To obtain the internal face for feature extraction, we need to exclude or eliminate hairs from a facial region as possible. According to the locations of eyes, a central zone of a face region can be obtained and then a dominant color is first measured from the central zone. Based on the measured dominant color, an internal face can be obtained. Figure 2 illustrates an example of internal face.

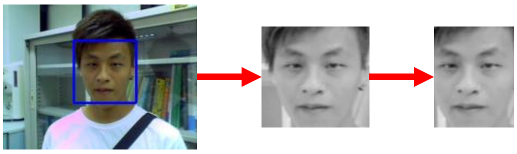


Fig. 2 An illustration of internal face

The adopted features in the proposed scheme are summarized as follows. First, we extract texture features from eyes by using local binary pattern (LBP) [4]. Second, we extract texture features from internal face by using Law's mask [5]. In the paper, 8 Law's masks are adopted.

The last feature is computed from hairs. Since the internal face can be found, the hair information can be extracted from two sides beside the internal face. The hair information beside the internal face for male persons is often different from that for female persons.

D. Gender classifier

Since the three kinds of features are extracted from different parts of a facial image, features in each class have their particular properties. Then it is necessary to fuse these kinds of features to make the gender classifier better. In addition, gender recognition can be considered as a binary classification problem. Therefore, a multiple-sensor fusion method [7],[8] based on minimizing a Bayesian risk can be used to fuse the three kinds of features here. Figure 3 illustrate the proposed gender classifier based on multi-mode information fusion.

Consider a binary hypothesis testing problem with two hypothesis H_0 and H_1. Their priori probabilities are defined as $P(H_0) = P_0$ and $P(H_1) = P_1$, where $P_0 + P_1 = 1$. Assume that observations at individual sensors are statistically independent. The conditional probabilities are defined as $P(x_i|H_j)$, i=1, …, N_S, j=0,1, where N_S represents the number of sensors. The local decision of each sensor is expressed as

$$c_i = \begin{cases} -1 & \text{if } H_0 \text{ is declared} \\ +1 & \text{if } H_1 \text{ is declared} \end{cases}. \tag{2}$$

The probabilities of false alarm (FA) ($P(x_i|H_0) < P(x_i|H_1)$, $x_i \in H_0$) and missing rate (MR) ($P(x_i|H_0) > P(x_i|H_1)$, $x_i \in H_1$) for each sensor are denoted by

P_{F_i} and P_{M_i}, respectively. P_{F_i} and P_{M_i} can be estimated from training data. After processing the observations locally, the local decisions c_i's are transmitted to the data fusion center. Then, based on the Bayesian criterion, the fused decision C can be expressed as

$$C = f(c_1,\ldots,c_n) = \begin{cases} -1 & \text{if } \omega_0 + \sum_{i=1}^{N_S}(\omega_i \cdot c_i) > 0, \\ +1 & \text{otherwise} \end{cases} \tag{3}$$

where $\omega_0 = \log P_1 / P_0$

and
$$\omega_i = \begin{cases} \log \dfrac{1-P_{M_i}}{P_{F_i}} & \text{if } c_i = +1 \\ \log \dfrac{1-P_{F_i}}{P_{M_i}} & \text{if } c_i = -1 \end{cases}. \tag{4}$$

According to Eq. (3), the fused result is a weighted sum of local decisions. As we can see in Eq. (4), the weights of sensors, ω_i's, are calculated from P_{F_i} and P_{M_i}. This means that ω_i is related to the discriminating power of the sensor.

As shown in Fig. 3, different classifiers are used for features in different classes. Features from eyes and internal face are combined with SVM classifiers and the hair information is combined with a Bayesian classifier. The final decision can be obtained by fusing the results of three individual classifiers.

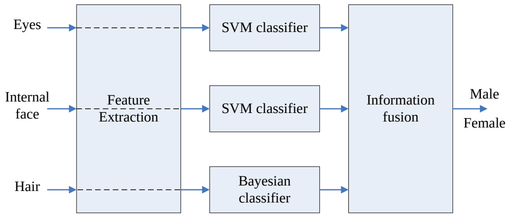


Fig. 3 An illustration of gender classifier based on multi-mode information fusion

III. EXPERIMENTAL RESULTS

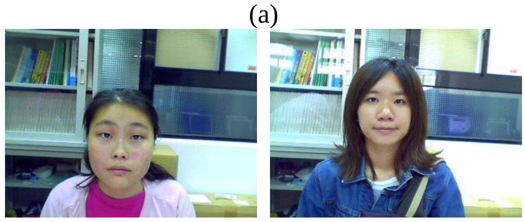
To evaluate the performance of our proposed scheme, we use a low-cost webcam to obtain an image database where the size of each image is 320×240 pixels for testing. In this database, there are 510 frontal images where there are 270 and 240 images for male and female, respectively. In these images, the sizes of faces are from 65×65 to 177×177 pixels. In addition, there are several expressions in these images and the background is not simple and fixed in these test images. Figure 4 shows some test images. As shown in Fig. 4, the light is not fixed and the background is not non-textured.

A. Face detection and rectification

We first evaluate the performance of face detection and rectification in the proposed scheme. Table I lists the result of face detection. As shown in Table I, compared with [1], our proposed scheme can eliminate many false alarms and only few missing rates increase. The result shows that our scheme can provide a better performance of face detection compared with [1].

Figure 5 shows two examples of face rectification. As we can see in Fig. 5, two rotated images can be rectified to be upright. To evaluate our proposed scheme, 510 face images are tested and only 6 face images fail. The result shows that the proposed scheme can rectify a rotated image well. It is expected that the impact of face rotation on gender recognition can be reduced effectively.

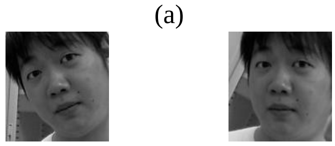
(a)



(b)

Fig. 4 Some test images: (a) male, (b) female

(a)



(b)

Fig. 5 Examples of face rectification: (a) 10 degree rotated, (b) -12 degree rotated.

Table I. Results of face detection

	Total faces	# of MR	# of FA
[1]	510	4	64
Proposed	510	10	41

B. Eye detection

Figure 6 shows the result of eye detection and each black rectangle window indicates a detected eye. As we can see in Fig. 6, each eye can be localized well by the proposed eye detector. In addition, to evaluate the accuracy of eye detection, a metric is defined as follows:

$$\eta = \frac{1}{N_E}\sum_{i=1}^{N_E} U\left(\frac{\Omega_{E_i} \wedge \tilde{\Omega}_{E_i}}{\Omega_{E_i}} - 0.5\right), \tag{5}$$

where $U(x)$ is unit step function ($U(x)=1$ for $x>0$ and $U(x)=0$ for $x \le 0$); N_E is the number of eyes; Ω_{E_i} and $\tilde{\Omega}_{E_i}$ denote the true and detected areas of eyes, respectively; $\wedge$

represents the logical AND operator. For all of detected faces in test images, the value of η for our eye detector is equal to 1. Therefore, the results demonstrate that the proposed eye detection function well.

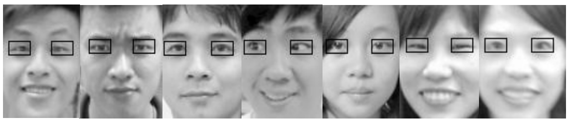


Fig. 6 Examples of eye detection

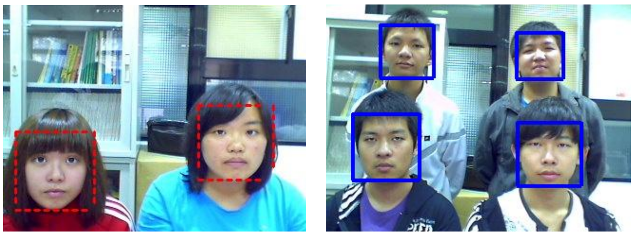
C. *Gender classifier*

As shown in Fig. 3, three kinds of features are extracted as inputs of three individual classifiers. Each classifier is learned by a trained image set. Based on multi-model information fusion, we can obtain the final result. From Table II, the accuracy is 95.2% for all detected faces. The result demonstrates that our proposed scheme can function well for low-resolution images.

Table II. The accuracy of gender recognition.

	Male	Female	Accuracy
Proposed	248/266	228/234	95.20%

We also capture some images with multiple persons. Figure 7 shows the result of gender recognition for our proposed scheme when there are several persons in an image. As we can see in Fig. 7, male and female persons can be correctly identified.



(a) (b)

Fig. 7 Results of gender recognition for multiple-person images

IV. Conclusions

In the paper, we proposed a gender recognition scheme based on multi-model information fusion. The proposed gender recognition scheme is composed of four parts: face detection and rectification, eye detection, feature extraction, and gender classifier. The face detection and rectification part is used to detect and rectify faces for improving the performance of gender recognition. The eye detection part can not only localize the eyes for feature extraction but also improve the performance of face detection. Three kinds of features from eyes, internal face, and hairs are adopted and combined with a gender classifier based on multi-model information fusion.

To evaluate the proposed scheme, a large number of images containing different-size faces are captured by using low-cost webcam. Experimental results show that our proposed scheme can detect facial regions as well as eyes well. In addition, the accuracy of the proposed gender recognition scheme is more than 95%. These results demonstrate that our proposed scheme can achieve not only face and eye detection but also gender recognition.

References

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, vol.1, pp. 511–518 , 2001.

[2] C. Y. Yu, Y.-C. Ouyang, C. M. Wang and C. I Chang, "Adaptive inverse hyperbolic tangent algorithm for dynamic contrast adjustment in displaying scenes," *EURASIP Journal on Advanced in Signal Processing*, vol.2010, pp.1-20, 2010.

[3] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions Image Processing*, vol.13, pp. 600–612, 2004.

[4] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol.24, pp. 971–987, 2002.

[5] K. I. Laws, "Rapid texture identification," *In SPIE, Image Processing for Missile Guidance*, vol. 238, pp.376-380, 1980.

[6] Y. Andreu and R. A. Mollineda, "The role of face parts in gender recognition," *LNCS*, vol.5112, pp. 945–954, 2008.

[7] Z. Chair and P.K. Varshney, "Optimal Data Fusion in Multiple Sensor Detection Systems," *IEEE Transactions Aerospace and Electronic Systems*, vol. AES-22, pp. 98-101, 1986.

[8] W. N. Lie and C. K. Su, "News Video Classification Based on Multi-modal Information Fusion," in *Proc. IEEE Int'l Conf. on Image Processing*, ICIP-2005, Genova, Italy.

[9] Y. Shiqi, T. Tieniu, H. Kaiqi, J. Kui and W. Xinyu, "A Study on Gait-Based Gender Classification," *IEEE Transactions Image Processing*, vol.18, pp. 1905-1910, 2009.

[10] Duan-Yu Chen and Kuan-Yi Lin, "Robust gender recognition for uncontrolled environment of real-life images," *IEEE Transactions Consumer Electronics*, vol.56, pp.1586–1592, 2010.

[11] G. Amayeh, G. Bebis, and M. Nicolescu, "Gender classification from hand shape," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, vol.1, pp. 1-7, 2008.

[12] G. Tzanetakis, "Audio-based gender identification using bootstrapping," in *Proc. IEEE Pacific Rim Conference on Communications, Computers and signal Processing*, pp. 432-433, 2005.

[13] S. Aji, T. Jayanthi and M.R. Kaimal, "Gender identification in face images using KPCA," in *Proc. IEEE Conference on Nature & Biologically Inspired Computing*, pp. 1414-1418, 2009.

[14] A. Hadid and M. Pietikainen, "Combining motion and appearance for gender classification from video sequences," in *Proc. IEEE Int'l Conf. on Pattern Recognition*, pp. 1-4, 2008.