JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 35, 105-123 (2019)
DOI: 10.6688/JISE.201901_35(1).0006

Gender Classification with Jointing Multiple Models for Occlusion Images

CHIAO-WEN KAO[1], HUI-HUI CHEN[2,*], BOR-JIUNN HWANG[2],
YU-JU HUANG[1] AND KUO-CHIN FAN[1]
[1]*Department of Computer Science and Information Engineering*
National Central University
Taoyuan, 320 Taiwan
[2]*Department of Computer Communication and Engineering*
Ming Chuan University
Taoyuan, 333 Taiwan
E-mail: {chiaowenk; louise121307}@gmail.com;
{huichen; bjhwang}@mail.mcu.edu.tw; kcfan@csie.ncu.edu.tw

A facilitated and effective gender recognition approach is desirable for various applications such as for intelligent surveillance systems, human-computer interactions, and consumer behavior analysis. Since the human face conveys clear sexual dimorphism, the use of facial features seems an intuitive way to recognize gender. This paper proposes an efficient gender classification method using multiple classifiers to overcome the occlusion problem. The experiment is tested via 5-fold cross validation on the FERET and AR databases to evaluate the performance. The results show the proposed approach achieves higher accuracy than previous methods.

Keywords: gender classification, component based, multiple classifiers, occlusion image, SVM

1. INTRODUCTION

Gender is an essential personal attribute in human beings. It is also an important kind of data in sociology to understand demographic structure. In recent years, gender classification has been widely used in the Human Computer Interface (HCI), intelligent robots, surveillance, differentiated marketing and so on. For surveillance, an automatic door security system with gender classification can provide a much securer place for single sex schools or lavatories. The system will automatically send out an “alert” to notify the security guard when the system detects an abnormal situation. For human-computer interactions, an anthropomorphic robot applying gender classification can automatically respond to different genders.

In addition, gender classification is widely used for effectiveness measurement in advertising. Digital signage can vary advertising content according to the specific gender ratios in public places, such as stations, bus shelters, shopping malls and sidewalks. Based on consumer behavior analysis, adaptive content can attract potential customers while demonstrating custom advertisements accordingly. Therefore, an automatic and effective gender classification system brings more convenience to life.

Gender classification is a binary problem. It is easy for people to differentiate between male and female through the human face because the face provides much observ-

Received September 13, 2017; revised November 27, 2017; accepted January 15, 2018.
Communicated by Chung-Lin Huang.

able information such as identity, gender, age, ethnicity, and emotion. Even when the facial image is incomplete, it remains possible for people to distinguish gender. However, it is really a challenging task in the computer vision field when the deficient/occluded image is captured in an uncontrolled environment. Hence, similar to most pattern recognition problems, how to extract useful features and determine pattern classifiers are the two primary issues for gender classification.

In feature extraction, whether static or video images, the feature extraction methodologies are generally either geometric based, appearance based or template based [12]. Geometric based feature extraction pertains to the measurements of the relationships between noticeable facial components, such as the eyebrows, eyes, nose, mouth and other significant facial elements. In [3], the forty fiducial points of the face are extracted manually, and then twenty-four horizontal and vertical distances are calculated to describe the facial image features. The characteristics of this approach result in a large number of features to be used by the classifier. The method of appearance based [4-7] refers to some operations or transformations performed on image pixels to extract the features which can be used to train a classifier directly. However, this kind of approach requires good quality images to extract the features correctly.

The method of the template based matches the features from input images to previous modeling templates. This approach works when the query and the model image have the same scale, orientation and illumination properties. This approach is only effective in a controlled environment. Regardless of the features extracted through any of the abovementioned methods, numerous classification approaches, such as a Neural Network, Support Vector Machine (SVM), Bayes classifier, Decision Tree, Fuzzy, Distance based classifier have been proposed.

Although the full facial image can supply complete information for a classifier, a component based facial image can provide more robust features when the occluded facial image is captured. Hence, to reduce the influence caused by an incomplete image and varying conditions, such as low resolution, variety offset angles, occlusion, blur or underexposed/overexposed pictures, this paper proposes using multiple classifiers based on facial components to overcome the occluded image problem. The experimental results show component based facial gender classification based on multiple classifiers is more efficient and accurate when fusing several distinctive components.

The remainder of this paper is organized as the followings. Section 2 presents the theoretical background of the used approaches. Section 3 presents the proposed methodology for gender classification to solve the problem of an occluded facial image. The details and results of the experimental scenarios are explained in section 4. Finally, conclusions and discussion are given in section 5.

2. RELATED WORK

Although numerous approaches have evolved for gender classification, the search for a highly accurate and efficient method continues. Mozaffari, Behravan and Akbari proposed a Geometrical Distance Features (GDF) added to other appearance-based features 4. After calculating the best-fit ellipse of the face, the two values from the fitted ellipse are used for GDF to represent the face. These two values are the ratio of the major axes to the minor axes and the RMS between best-fit ellipse and face contour, respectively.

In [4-7], they used the Local Binary Pattern (LBP) for gender classification. The original LBP is a simple texture operator which labels the pixels of an image by setting the threshold of the neighborhood of each pixel and considers the result as a binary number. To improve classification accuracy, Xia [8] proposed the Local Gabor Binary Mapping Pattern (LGBMP) method combining LBP features with Gabor filters and histogram mapping.

As mentioned above, most studies consider the full facial image where the accuracy is reduced as the partial facial component is occluded. Therefore, to obtain higher performance and robustness, some studies extract the facial components by avoiding occlusion of angle poses and illumination. Lee and Hung [9] decomposed a face into several horizontal and vertical strips after the Active Shape Model (ASM) located the facial feature points and then calculated a likelihood of each strip by regression function. The likelihoods from all strips are concatenated to form a new feature vector inputted to the overall gender classifier (ε-SVR and C-SVC). Hu [9] selected five regions on the facial image in terms of Eyes, Lips, Nose, Chin and whole face. Then, the Maximum Response-8 (MR-8) filter set was used to extract the features of each region. Following feature extraction, these feature vectors were passed to several classifiers, such as SVM, Naive Bayes, and Logistic regression for classification. Li, Lian and Lu [7] not only considered facial components but also external information such as hair and clothing. For each component, they extract LBP features and train an SVM with a probabilistic output. Patel *et al.* [11] presented the multi-quantized local binary patterns for facial gender classification. In order to improve the accuracy, the sign and magnitude are adopted to execute multi-level vector quantization of gray level difference with SVM classifier. The four public available databases (FERET, PAL, CASIA and FEI) are adopted. The accuracy of the approach is about 95% but the number of used images is unknown. Juan [11] predicted the gender by using the binary iris code with SVM classifier. They also observe that the features from subset of iris region can get better accuracy than using whole iris region. This approach yielded an accuracy of 89% for gender classification. Andreu *et al.* [12] presented a study of gender classification techniques using non-distorted and distorted faces, such as neutral, expressive and partially occluded faces. A comprehensive comparison of two representation approaches (global and local), three types of features (grey levels, PCA and LBP) and three classifiers (1-NN, PCA + LDA and SVM) has been provided by means of three statistical tests applied to two performance measures (CCR and D-prime).

As already mentioned, the image must be positive and also display a clean frontal facial image to improve accuracy. In a realistic scenario, partial occlusion images may occur. These images may occur when subjects wear sunglasses or scarves, or when their faces are covered by masks, cellphones, hair or other substances. To reduce the classification performance affected by the occluded facial appearance, this paper proposes a component based facial gender classification approach from multiple classifiers (CFMC).

3. PROPOSED METHOD

The CFMC consists of two stages, including the testing stage and training stage. The system architecture is shown in Fig. 1. In the training stage, the main task is using the collected training images to train the Gender Classifier Model (GCM) and Occlusion Classifier Model (OCM). To accomplish this aim, the segmentations of the facial components are preprocessed, and the feature of each facial component is extracted. Then the gender and occlusion classifiers are trained, respectively. After completing the training stage, to obtain the feature vectors, the test image executes the segmentations and extractions of the facial components, the same as in the two procedures used in the training stage. The predicted outputs from the Gender classifier and the Occlusion classifier are used to estimate the values of gender confidence (*GC*) and the occlusion degree for each component. Finally, these two variables are considered in the dynamic decision fusion procedure to judge the gender of the inputted test image.

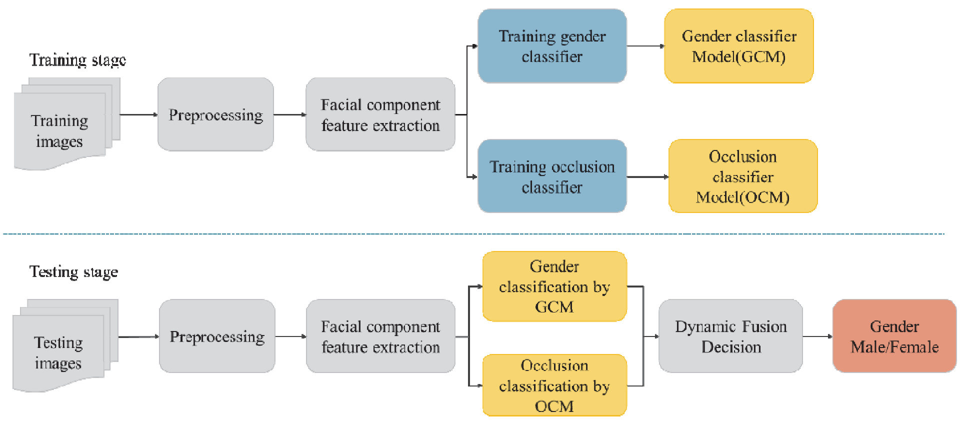


Fig. 1. System architecture.

3.1 Preprocessing Procedure to Segment the Facial Components

The purpose of the preprocessing procedure is to acquire a facial image and normalized facial components to crop the clear facial components for feature extraction. Hence, the Stacked Trimmed Active Shape Model (STASM) algorithm [14] is utilized to extract the facial landmarks automatically. The STASM is extended from the traditional ASM algorithm [15]. It increases the number of landmarks using two-dimensional models, adding noise during the training stage, reducing the number of eigenvectors of the shape model, decreasing the covariance matrices and adopting the double ASM searching stacked model to improve the accuracy of the labeled facial points. The STASM decreases computational complexity and thus increases robustness.

In this paper, the 76 facial landmarks model is adopted to label the facial features. The rotation calibration procedure is executed to decrease the impact of the facial image rotation after labeling the feature positions. To measure an angle (θ) for rotation calibration, the central point landmark between the left and right pupil is used as the datum mark, as shown in Fig. 2.

The calculated angle θ is between the line containing the datum marks and the positive x-axis. After angle θ is obtained, set the central point (C_x, C_y) of the facial image to be the origin, then the other points can be transferred by Eq. (1).

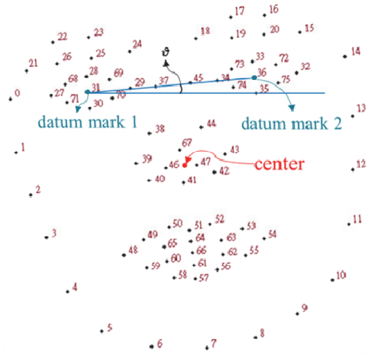


Fig. 2. The facial landmarks and variables for rotation calibration.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} C_x \\ C_y \end{bmatrix} \tag{1}$$

where (x, y) is the point in the facial image and (x', y') is the point following the calibration process.

Following the rotation calibration procedure, the facial image is segmented into four components in terms of the left or right eye area (comprising the eyebrow and eye), nose and mouth, which are normalized to the standard size. Table 1 shows the normalized size of four components, respectively.

Table 1. The size of each component.

Component	Image size (pixel)
Left Eye Area	30 × 30
Right Eye Area	30 × 30
Nose	40 × 20
Mouth	60 × 20

The illumination source of the shooting environment will influence the brightness of the facial images. Hence, histogram equalization is used to overcome the illumination problem and strengthen the credibility of the feature, as shown in Fig. 3.

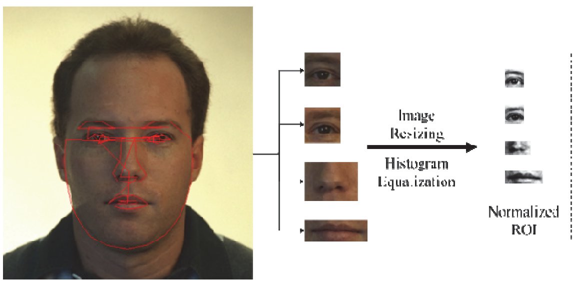
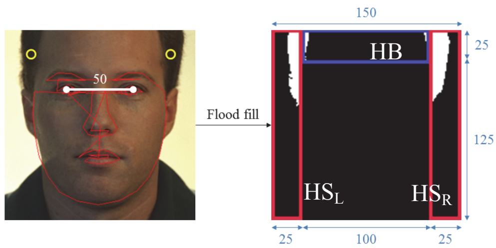


Fig. 3. Facial components before and after normalization.

Besides the facial features, hairstyle is also an important feature of people in distinguishing gender. It can offer useful information because the length and style can significantly differ between males and females. However, in this paper, to avoid mistaking a man with long hair for a woman, and vice versa, the hair feature is dynamically adopted when the gender probability confidence value cannot precisely identify the gender.

The procedure of obtaining hair information occurs in three stages: decide the region of hair, extract the hair feature and discriminate the hair style. To reduce the computation time and improve the estimation accuracy of the hair style, searching the exact region of the hair is necessary. First, with the fixed distance between binocular pupils at 50 pixels. Next, the facial image is relocated with the positions of the right pupil and left pupil point, as (50, 50) and (100, 50), and the facial image is cropped into a size of 150× 150 pixels. Last, referring to Fig. 4 (b), the search of the hair region is the upside down U-shaped at the outer part of the middle 100×125 pixels facial region.



(a) (b)

Fig. 4. The graphic of the hair searching boundary.

3.2 Feature Extraction

In this paper, multiple classifiers are utilized to classify the gender and occlusion degree of each facial component. Fig. 5 shows the relationship between the different feature extractors and classifiers. Before training the classifier, the valid and positive correlation features can improve the accuracy of the classification and reduce the influence of noise. The fine features can be optimized to improve the classification efficiency; therefore, the Principal Component Analysis (PCA) feature descriptor [16], Histogram of Oriented Gradients (HOG) descriptor [17] and Flood Fill Algorithm [17, 18] are adopted to extract the individual features in this paper.

The operation of PCA can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. It is done by using the first few principal components to represent the most informative features to describe the image instead of considering the original high dimensional feature space. Hence, PCA is suitably used to extract the feature vectors of the cropped facial images.

Besides, another feature descriptor HOG is used to classify the occlusion degree. HOG is computing the histogram of gradient orientation in the localized point of an image. The gradients of an image are useful because there is large inconsistency around the edges and corners (the region of abrupt intensity changes). The edge and the corner pack in more information about the object shape than the flat regions do. It is therefore better to describe the facial components according to such characteristics of edges and corners.

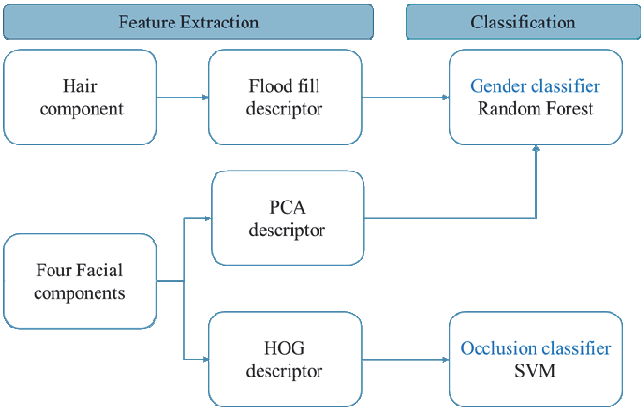


Fig. 5. The relationship between the different feature extractors and classifiers.

3.2.1 Hairstyle discrimination

After labeling the hair region, the Flood Fill Algorithm is adopted to extract the hair feature in this paper with a regression model that uses the color seed to search and fill the connected components until the difference in color degree of the neighborhood is larger than a threshold. The color seed is adaptively selected according to each facial image because the hair color is varied. To obtain the representative color seed, the two points, each on either side of the forehead, are selected to be color seeds and the start points, as shown by the yellow circles in Fig. 5 (a). Going over the assigned start point, color and the color difference tolerance calculated, the algorithm searches the hair region for all pixels that are connected by two start points and by a path of the seed color, and it adds up the number of pixels matching the criteria. According to the layout in Fig. 5, HS_L and HS_R, represent the left/right hair ratio, which can be defined as Eq. (2):

$$HS_i = \frac{\sum floodfill(x, y)}{M \times N} \quad i \in R, L \tag{2}$$

where *floodfill* (x, y) means the pixel (x, y) whether similar to the color seed. M and N represent the width and height of the area. In the experiment, a width of 25 pixels and height of 150 pixels are assigned. The sum of the HS_L and the HS_R to HS present the hair volume and calculates the hair ratio in the bang region (H_B). Therefore, the hair feature vector can be presented as two-dimensions (HS, H_B).

3.2.2 Training the classifiers

After gathering the feature set, the structure of the learned function and corresponding learning algorithm need to be determined by a training set for the supervised classifier. As previously mentioned, the Random Forest algorithm and Support Vector Machine are used to predict the gender of cropped facial components and whether the component is an occlusion.

First, the gender classifier, Random Forest algorithm [19] is a useful classifier based on multiple decision trees to deal with numerous input variables. In order to obtain the regression tree, the training algorithm for the Random Forest applies Bootstrap Aggregation to the calculation. After training, predictions for unseen samples can be made by averaging the predictions from all individual regression trees or by taking the majority vote with the decision trees. This paper adopts the first statistical method.

According to the above procedure, in this paper, the feature set, including the feature of facial components extracted by PCA and the hair feature extracted by Flood Fill Algorithm, is used to train the decision tree individually. To classify a new component from an input vector, the input vector is put down to each of the trees in the forest. Each tree gives a classification result, with a value of 1 being assigned to the male group and 0 being assigned to the female group.

So, when the input is feature x, it is assumed t_x^1 is the quantity of classification results in the male group, the prediction from the t decision trees. And t_x^0 is the quantity of predicted results for the female group. Thus, the gender confidence (GC) value for feature x can be computed as Eq. (3):

$$GC(x) = \frac{t_x^{\ 1}}{t} = 1 - \frac{t_x^{\ 0}}{t} \tag{3}$$

where the value of GC ranges from 0 to 1.

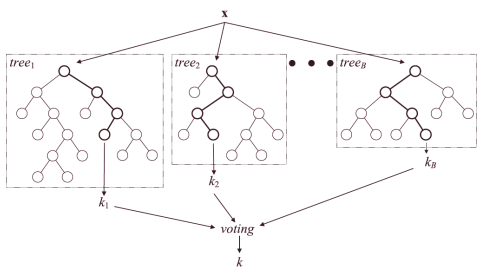


Fig. 6. The basic structure of the random forest.

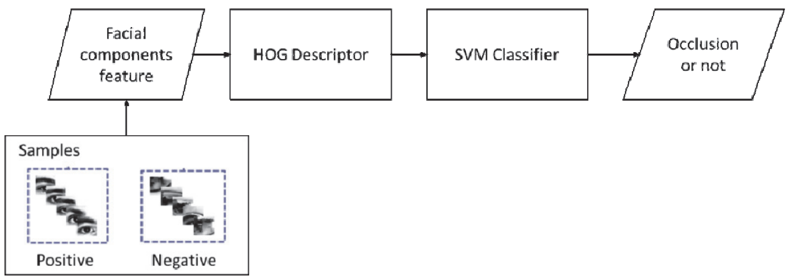


Fig. 7. The occlusion classification procedures of the facial feature.

To prevent the accuracy being influenced by the occlusion problem, the SVM classifier [20] is used to classify whether the component is either the occlusion or not, as shown in Fig. 7.

The SVM is also a supervised learning model. Given a set of training samples, each sample is labeled into two categories: either the occlusion or the available (non-occlusion). An SVM model is a representation of the samples as points in space. These points are mapped to be divided by a clear gap (termed the margin), so the samples can be classified into separate categories. The margin can be as wide as possible. Then, new samples are mapped into the distinguished space and can be categorized based on which side of the gap they fall on. Fig. 8 shows the maximum margin hyperplane and margins for an SVM trained with the samples into separate categories. The sample on the margin is termed the support vector.

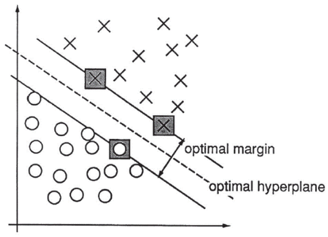


Fig. 8. The SVM graphic trained with samples into separate categories.

As already mentioned, the HOG feature extracted from the cropped facial components is input to train each SVM model to predict the occlusions. Each model gives a classification result of 0 for the occlusion group or 1 for the available group. In the final dynamic fusion decision procedure, only the component of the available group is selected. When the input is feature x, the output $S(x)$ can be presented as Eq. (4)

$$S(x) = \begin{cases} 0, & \text{if geaturex belongs occlusion category} \\ 1, & \text{otherwise} \end{cases}. \tag{4}$$

3.3 Dynamic Fusion Decision

The main purpose of this section is using the dynamic fusion decision to integrate the output predictions. The $f(GC(x), S(x))$ feature vectors are considered in the dynamic fusion decision procedure. First, the confidence values (CV) are computed with facial components, as Eqs. (5) and (6):

$$CV = \sum_{i=1}^{4} GC(b_i)S(b_i) \tag{5}$$

$$N = \sum_{i=1}^{i=4} S(b_i) \tag{6}$$

where b_i denotes the feature of one of the facial components, i ranges from one to four and represents the component of Left eye, Right eye, nose and mouth in sequence. $GC(b_i)$ denotes the gender confidence value of b_i. $S(b_i)$ represents whether b_i is selected. N is the number of selected components. Finally, the mean of gender confidence value (CV) of the input image is defined as Eq. (7):

$$\overline{CV} = \frac{CV}{N}, N \neq 0 \tag{7}$$

where the value of $\overline{CV}$ is in the range [0-1]. The value gets close to 1, meaning a higher probability of a male. Conversely, a value near 0 represents a higher probability of a female. Nevertheless, the hair feature, $GC(b_5)$, is considered while CV is in an ambiguous range. Otherwise, if all the facial components are classified as the occluded category, the holistic components including hair feature are considered. Hence, Eq. (8) is used to

$$R = \begin{cases} (\overline{CV} + GC(b_5))/2, & \text{if } 0.45 < \overline{CV} < 0.55 \\ GC(b_5), & \text{if } N = 0 \\ \overline{CV}, & \text{otherwise} \end{cases} \tag{8}$$

where R represents the ultimate gender confidence value. Then, the subject of the input image is classified as male if the value R is greater than 0.5; otherwise, the subject is classified as female.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The experiments were performed with two publicly available sets: Facial Recognition Technology and AR databases. The FERET database [21] contains 994 individuals of 591 males and 403 females used in the experiments. The image resolution is 512×768 pixels. AR database [22] consists of different facial images of 76 males and 60 females (total 136 persons), with details in Table 2. In this paper, the neutral facial images (frontal facial images without any expression) and faces covered with sunglasses or scarves from the AR database are used. Sample images are shown as in Fig. 9. The experimental accuracy rates in diverse scenarios for these two databases are illustrated in Table 3.

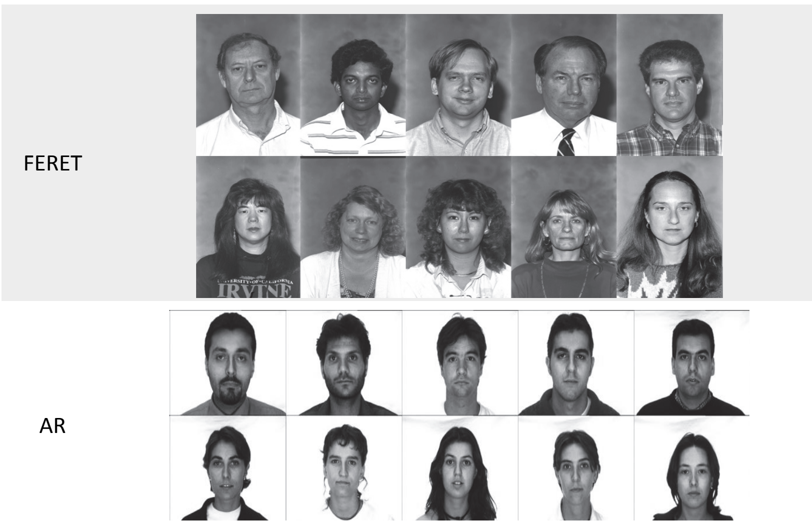
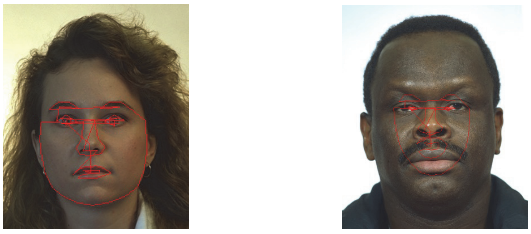


Fig. 9. A part of the FERET and AR databases.

Table 2. The total images of each face database.

Database	Total	Normal	Glasses	Sunglasses	Scarf
FERET	591 / 403	528 / 392	63 / 11	–	–
AR	76 / 60	44 / 52	32 / 8	76 / 59	45 / 51
Total – all situations					
Normal – not wearing glasses					
Glasses – wearing glasses					
Sunglasses – wearing sunglasses					
Scarf – wearing a scarf					

To obtain the facial landmarks and extract the normalized facial components, the face detection algorithm, STASM, is used in this paper. This step requires at least one facial component to completely obtain the successful executions in the following procedures. Fig. 10 (a) shows a successful example and Fig. 10 (b) shows a failed example. The result shows the success rate can reach almost 90%. Table 3 shows the accuracy rates from STASM on images in respective databases in different scenarios.



(a) A success example. (b) A failure example.
Fig. 10. The results of labeling landmarks.

Table 3. The accuracy rates of each database in different scenarios.

Database	Type	Accuracy
FERET	Normal	93.15 %
	Glasses	95.95 %
AR	Normal	100.0 %
	Glasses	97.50 %
	Sunglasses	90.37 %
	Scarf	90.63 %

4.1 Accuracy on Single Classifier

This experiment is designated to measure the gender classification performance of each facial component, in terms of the eye, nose, mouth and hair region. As for sampling reliability, the number of male and female images is the same in this experiment, as shown in Table 4. The 5-fold cross-validations are used to obtain the average accuracy of the gender classification for each component. Table 5 presents each performance where only one component of a single classifier with the FERET and AR databases is utilized. The first column shows the selected component, and the second and third columns show the average accuracy of each component with the FERET and AR databases.

The results show the Left eye component and nose component achieve the best classification performance, 82.93% in the FERET database and 83.67% in the AR database, respectively. It can be also found the eye area is a forceful feature for gender classification. The hair feature is not a significant feature since the background can easily affect it and there is no certain kind of hairstyle for different genders. Therefore, the dynamic fusion decision procedure is necessary to prevent the influence on the prediction accuracy through an unrepresentative feature.

Table 4. The number of the images without glasses in this experiment.

Database	Train (M/F)	Test (M/F)	Total
FERET	299 / 299	75 / 75	748
AR	35 / 35	9 / 9	88

Table 5. The accuracy results of each component with a single classifier.

Components \ Database	FERET	AR
Left eye	**82.93 %**	82.67 %
Right eye	80.71 %	83.11 %
Nose	79.14 %	**83.67 %**
Mouth	77.75 %	82.61 %
Hair	74.41 %	79.00 %

4.2 Accuracy of Multiple Classifiers with Dynamic Fusion Decision

To confirm whether adaptively selecting multiple components is an acceptable plan, the accuracy rates for each individual in every conceivable combination with different databases are presented in Tables 6-8. The meanings of each abbreviation for each component are explained as follows: Left Eye (LE), Right Eye (RE), Nose (N), Mouth (M), Hair (H).

Table 6. Classification accuracy using combination sets of two components.

Database	Accuracy of Combination sets (%)									
Component set:	LE+RE	LE+N	LE+M	RE+N	RE+M	N+M	LE+H	RE+H	N+H	M+H
FERET	85.47	87.92	**88**	86.99	87.41	82.81	79.19	78.85	79.61	78.18
AR	85.61	90.44	**91.22**	90.28	90.61	88.06	82.78	83.56	83.50	82.17

Table 7. Classification accuracy using combination sets of three components.

Database	Accuracy of Combination sets (%)									
Component set:	LE+RE +N	LE+RE +M	LE+N +M	RE+N +M	LE+H +RE	LE+M +H	L +N +H	RE+M +H	RE+N +H	N+M +H
FERET	89.45	**89.65**	88.84	88.76	82.77	82.52	84.47	83.77	82.22	83.03
AR	90.83	92.78	92.39	**93.00**	85.89	86.11	87.39	85.89	87.33	87.56

Table 8. Classification accuracy using combination sets of four and five components.

Database	Accuracy of Combination sets (%)					
Component set	LE+RE+N+M	LE+RE+N+H	LE+RE+M+H	LE+N+M+H	RE+N+M+H	LE+RE+N+M+H
FERET	**90.90**	87.48	86.17	87.15	86.53	89.66
AR	**92.83**	89.89	91.00	91.33	91.78	92.89

In general, the fusion decision combines the components that can provide better classification performance. The results in Tables 6-8 also show that the classification accuracy is highly relevant with the number of combined components. The results as shown in Table 6, combined the two components of the highest accuracy are not as high as expected. For example, the accuracy for the combination of the Left Eye and Right Eye is not the highest in the FERET database. In addition, the accuracy decreases when the hair feature is used to replace any feature in the FERET database, but the accuracy increases in the AR database if the hair feature replaces any feature, as shown in Table 8. As mentioned, these phenomenon exhibit that the each facial image has the own significantly feature. There is no one specific combination set can suitable for different facial image. Hence, it is reasonable to use the dynamic fusion decision method to select the combined components and it has the opportunity to get the higher accuracy.

Table 9 shows the accuracy of the proposed method. The accuracy of the FERET database is 90.28% and that of AR is 93.94% respectively. Dynamically speaking, choosing the classified component can provide higher accuracy and robustness. Some successful examples of the classification are shown in Fig. 11. The meanings of each different color are illustrated as follows: the blue box represents a male, the red box represents a female, the component, which is not selected, is shown as a black box, the label (H) means the confidence value of the hair feature and the number below the box means the confidence values of each selected component. A number close to 1 represents the component feature are more like male features. Conversely, a number close to 0 represents female features.

Table 9. The accuracy results of an adaptive fusion decision.

Dynamic fusion decision	Component selected	FERET	AR
without	LE+RE+N+M	90.9	92.83
	LE+RE+N+M+H	89.66	92.89
with	Proposed method (without hairstyle)	88.53	92.72
	Proposed method (with hairstyle)	90.28	93.94

The failure examples and missing detection of hair are shown in Figs. 12 and 13. Based on Fig. 12 (c), the male has delicate facial features. On the contrary, the female has the more neutral facial feature in the Figs. 12 (e) and (f). Besides, the hair component is adopted when when the gender cannot be obviously determined by using the four facial components, as shown in Figs. 12 (a), (b) and (d). However, the hair component is missing detection due to the background is influenced by the shadow or light, as shown in Fig. 13 (a) or (b).

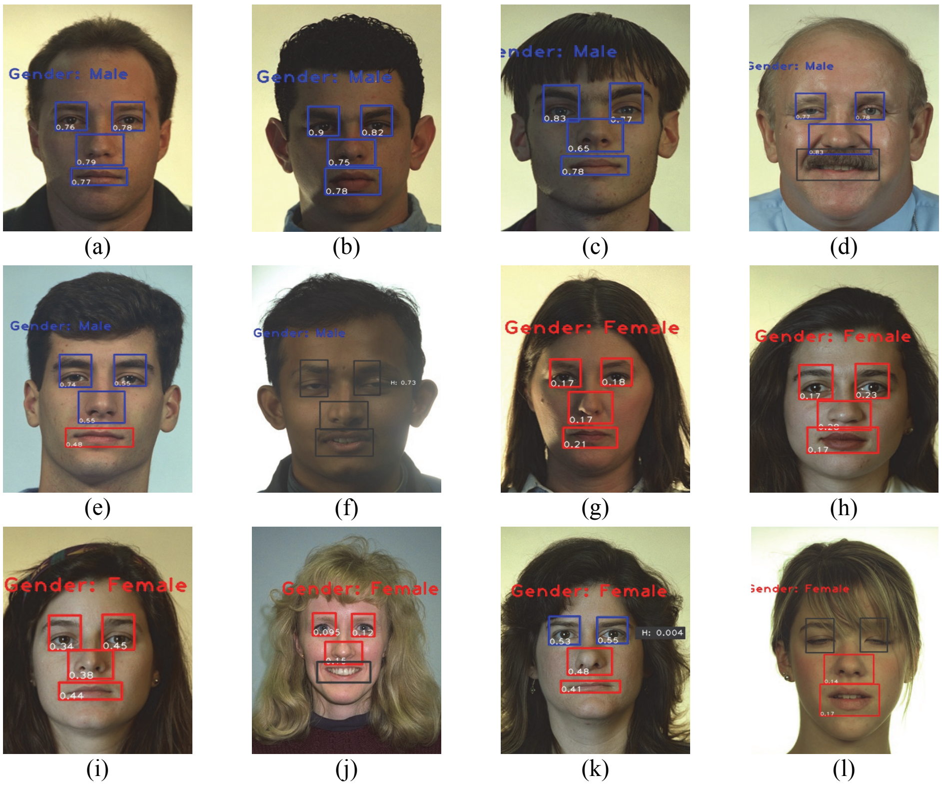


Fig. 11. Successful examples on the FERET database.

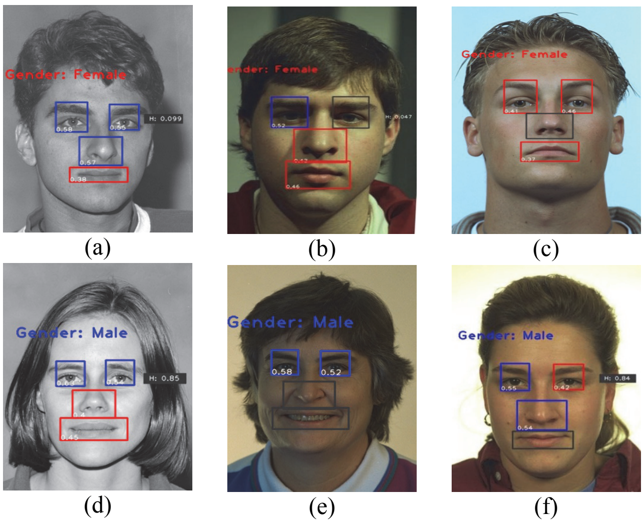
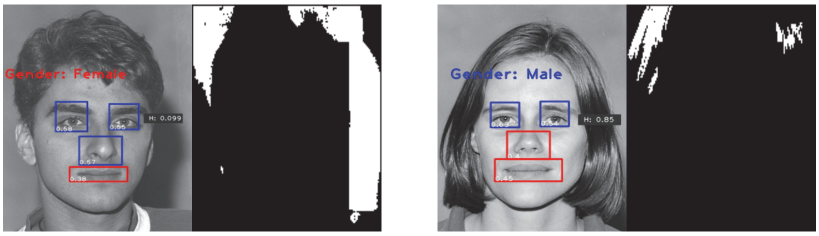


Fig. 12. Failure examples.



(a) (b)

Fig. 13. Examples of missing detection on hair.

4.3 Robustness of Adaptive Fusion Decision

The aim of this experiment was to investigate the robustness of the proposed method. The AR database used in this experiment consisted of the following characteristics: the training set had 44 images for each gender and these images are general facial images; the test set with two types: 51 images with sunglasses on for each gender and 43 images with scarves on for each gender. Not a single individual shows more than once in the dataset to prevent the classifier learning from gender features rather than recognizing features. The sample of the dataset is shown in Fig. 14. The images in this test set are more complicated than in the training set because there is less information in the occluded images for gender classification.

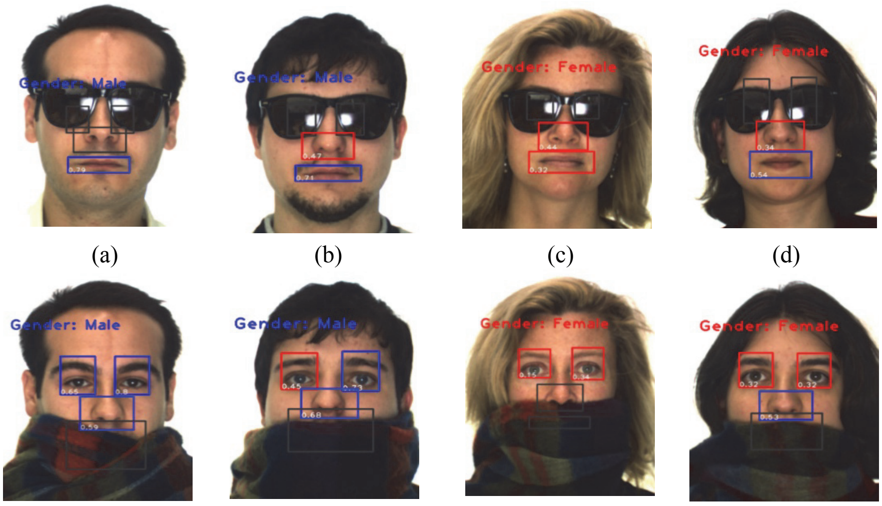
Fig. 14. Robustness test samples.

The accuracy of classifying the occluded image in the proposed method is, respectively, 88.24% with sunglasses on 94.19% with scarves on. According to Table 5, eye areas show the highest classification rate. Hence, it is evident that if eye areas are occluded, the recognition rate decreases significantly. However, the proposed method can provide a certain degree of identification accuracy.

Fig. 15 shows some examples classified as the occluded images, the type (I) shows subjects wearing sunglasses and also illustrates features of the eye areas that are not adopted when occluded. The type (II) presents the scenario when wearing scarves, the subject's nose and mouth areas may be occluded. Therefore, the non-occluded components are considered for gender classification. The proposed method can overcome the occlusion problem and is effective and robust in an uncontrolled environment.

Type (I)

Type (II)

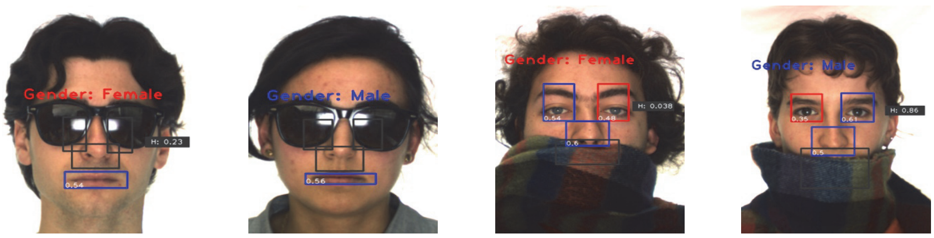


(a) (b) (c) (d)

(a) (b) (c) (d)

Fig. 15. Successful examples on the AR occlusion database.

Fig. 16 shows some failure examples. Sample (a) the hair and sunglasses are attached and influence the hairstyle detection; samples (b) and (d) show the neutral style type; sample (c) makes the hair around the forehead volumized. The incorrect judgment is often the result of too many occluded components and dressing in a neutral style.



(a) (b) (c) (d)

Fig. 16. Failure examples on the AR occlusion database.

5. CONCLUSIONS

This paper investigates gender classification on facial images in unconstrained conditions, a challenging but relatively understudied problem. It proposes a dynamic fusion decision for classifying gender using facial components. This method uses three kinds of feature extractors, PCA, Flood Fill Algorithm and HOG, to describe the feature of facial components and hair feature. In addition, another two types of classifiers, Random Forest and SVM, are used to classify the selected feature belonging to males or females.

Experimental evaluations of the proposed method evidently illustrate the effectiveness and robustness in the occlusion scenarios. Using the AR and FERET databases with neutral frontal facial images, the proposed method respectively achieved accuracy rates of 93.94% and 90.28%, and an average accuracy of 90.96% is obtained in the occlusion scenarios. The experimental results demonstrate the proposed method has excellent accuracy under various conditions. Future work includes applying this method to other databases and overcoming the wrongly located problem of the facial components due to the facial image being rotated or mostly masked.

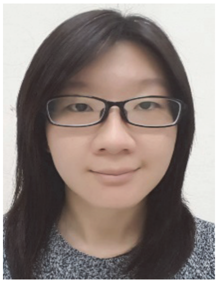
ACKNOWLEDGEMENTS

This work was supported by the Ministry of Science and Technology (MOST), Taiwan, under Contract of MOST 105-2221-E-130-016-.

REFERENCES

1. Y. Chen, S. Jiang, and A. Abraham, "Face recognition using DCT and hybrid flexible neural tree," in *Proceedings of IEEE International Conference on Neural Networks and Brain*, Vol. 3, 2005, pp. 1459-1463.
2. K. Mna, Q. Sa, and N. Riaz, "Gender classification with decision trees," *International Journal of Signal Processing*, *Image Processing and Pattern Recognition*, Vol. 6, 2013, pp. 165-176.
3. J.-M. Fellous, "Gender discrimination and prediction on the basis of facial metric information," *Vision Research*, Vol. 37, 1997, pp. 1961-1973.
4. S. Mozaffari, H. Behravan, and R. Akbari, "Gender classification using single frontal image per person: combination of appearance and geometric based features," in *Proceedings of International Conference on Pattern Recognition*, 2010, pp. 1192-1195.
5. H.-C. Lian and B.-L. Lu, "Multi-view gender classification using local binary patterns and support vector machines," in *Proceedings of International Symposium on Neural Networks*, 2006, pp. 202-209.
6. C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognition Letters*, Vol. 33, 2012, pp. 431-437.
7. B. Li, X.-C. Lian, and B.-L. Lu, "Gender classification by combining clothing, hair and facial component classifiers," *Neurocomputing*, Vol. 76, 2012, pp. 18-27.
8. B. Xia, H. Sun, and B.-L. Lu, "Multi-view gender classification based on local gabor binary mapping pattern and support vector machines," *International Joint Conference on Neural Networks*, 2008, pp. 3388-3395.
9. P.-H. Lee, J.-Y. Hung, and Y.-P. Hung, "Automatic gender recognition using fusion of facial strips," in *Proceedings of International Conference on Pattern Recognition*, 2010, pp. 1140-1143.
10. S. Y. D. Hu, B. Jou, A. Jaech, and M. Savvides, "Fusion of region-based representations for gender identification," in *Proceedings of International Joint Conference on Biometrics*, 2011, pp. 1-7.
11. B. Patel, R. P. Maheshwari, and R. Balasubramanian, "Multi-quantized local binary patterns for facial gender classification," *Computers and Electrical Engineering*, Vol. 54, 2016, pp. 271-284.
12. J. E. Tapia, C. A. Perez, and K. W. Bowyer, "Gender classification from the same

iris code used for recognition," *IEEE Transactions on Information Forensics and Security*, Vol. 11, 2016, pp. 1760-1770.
13. Y. Andreu, P. Garcia-Sevilla, and R. A. Mollineda, "Face gender classification: A statistical study where neutral and distorted faces are combined for training and testing purposes," *Image and Vision Computing*, Vol. 32, 2014, pp. 27-36.
14. S. Milborrow and F. Nicolls, "Locating facial features with an extended active shape model," in *Proceedings of European Conference on Computer Vision*, 2008, pp. 504-513.
15. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models – their training and application," *Computer Vision and Image Understanding*, Vol. 61, 1995, pp. 38-59.
16. I. T. Jolliffe, *Principal Component Analysis*, John Wiley and Sons, Ltd., 2002.
17. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Computer Vision and Pattern Recognition*, Vol. 1, 2005, pp. 886-893.
18. A. Glassner, *Graphics Gems I*, Morgan Kaufmann Publishers Inc., SF, 1999.
19. A. Liaw and M. Wiener, "Classification and regression by random forest," *R News*, Vol. 2, 2002, pp. 18-22.
20. M. A. Hearst, "Support vector machines," *IEEE Intelligent Systems and their Applications*, Vol. 13, 1998, pp. 18-28.
21. P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, Vol. 16, 1998, pp. 295-306.
22. A. Martinez and R. Benavente, "The AR face database," *CVC Technologies Report*, No. 24, 1998.



Chiao-Wen Kao (高巧汶) received the B.S. and M.S. degrees in Computer Communication and Engineering from Ming-Chuan University, Taiwan. She is currently pursuing the Ph.D. degree in the National Central University, Chung-Li, Taiwan. Her research interests include pattern recognition, image processing and Interactive media.

Hui-Hui Chen (陳惠惠) received her Ed.D. degree in Educational Instructional Technology from Texas Tech University in 2005. She is currently an Assistant Professor in the Department of Computer and Communication Engineering at Ming-Chuan University, Taiwan. Her research interests include e-learning, human-computer interaction, and integrating advanced technologies into teaching and learning.

Bor-Jiunn Hwang (黃博俊) received the B.S. degree in Electronic Engineering from Chung-Yuan Christian University, Chungli, Taiwan, in 1986, the M.S. degree in Information and Electronic Engineering in 1992 and the Ph.D. degree in Electrical Engineering in 2001 all from National Central University, Chungli, Taiwan. He is currently a professor in the Department of Computer and Communication Engineering at Ming-Chuan University, Taoyuan, Taiwan. His research interests include performance evaluation of multimedia mobile network, communication protocols and Interactive media.

Yu-Ju Huang (黃郁如) received the M.S. degree in Electrical Engineering from the National Central University. Her research interests include image processing and pattern recognition.

Kuo-Chin Fan (范國清) received the B.S. degree in Electrical Engineering from the National Tsing Hua University, Hsinchu, in 1981, and the M.S. and Ph.D. degrees from the University of Florida (UF), Gainesville, in 1985 and 1989, respectively. In 1983, he joined the Electronic Research and Service Organization (ERSO), Taiwan, as a Computer Engineer. From 1984 to 1989, he was a Research Assistant with the Center for Information Research, UF. In 1989, he joined the Institute of Computer Science and Information Engineering, National Central University, Chungli, Taiwan, where he became a Professor in 1994. Currently, he is the Dean of the Electrical Engineering and Computer Science College, National Central University. His current research interests include image analysis, optical character recognition, and document analysis.