Signal Processing: *Image Communication* 36 (2015) 179–189

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

A face detection and location method based on Feature Binding

Jing Jin [a], Bin Xu [a], Xiaoliang Liu [a], Yuanqing Wang [a,*], Liqun Cao [b], Lei Han [b], Biye Zhou [b], Minggao Li [b]

[a] *Nanjing University, School of Electronic Science and Engineering, No. 163, Xianlin Dadao, Nanjing 210046, China*
[b] *Navy General Hospital, Beijing 100088, China*

ARTICLE INFO

Article history:
Received 23 December 2014
Received in revised form
13 May 2015
Accepted 17 June 2015
Available online 7 July 2015

Keywords:
Feature Binding
Face detection
Face location
Pattern recognition
Human–machine interaction

ABSTRACT

A face detection and location method based on Feature Binding (FB) is proposed in this paper. The features used for face detection and location are classified and bound into groups. The information of each group is extracted separately during face detection. Through the combination with the constraint relationship, the precise location of the face in the image could be identified by confidence coefficients of all groups. Experimental results show that this proposed method can improve the accuracy rate obviously and has good detection effect on obscured faces. Besides, FB can be good to adapt to varieties of features.

1. Introduction

With the development of image processing and pattern recognition, the demands for intelligent processing have been growing. As a hotspot in the field of pattern recognition, face detection has been widely used in applications, such as authentication, attendance system and electronic passport. Besides, it has also been applied in a new naked-eye auto-stereoscopic display which requires accurate human face and eye locations [1]. Therefore, the accuracy of face detection and location is particularly important.

The primary task of face detection is to identify whether there is a certain face in a given image or image sequence, while that of face location is to calculate the face details including the position, size, quantity and spatial distribution. However, it is challenging to detect a face from an image with complex background because of the varying characteristics in faces such as scales, positions, orientations, and postures, as well as different facial expressions and light conditions.

With the increasing knowledge of human face, different kinds of face detection algorithms have been proposed. Conventional methods of face detection could be classified into four categories: a knowledge-based method, a feature invariant method, a template matching method and a statistical-based method. Knowledge-based method is a top-down method. It encodes human faces by regular database, which is formed by typical faces, and develops a series of criteria according to the relationships among the facial features. Faces can be detected when these criteria are met in the test region. Yang proposed a mosaic diagram in 1994 [2]. Afterwards, modified mosaic diagram [3] and a classification method [4] are proposed respectively. However, these methods strongly rely on the priori knowledge and have low recognition rate for changing faces. Different from the knowledge-based method, the feature invariant method is a down–up method, which includes a morphological method [5], a random labeled graph method [6] and a wavelet decomposition method [7]. The objective of these methods is to find invariable structural features under the variation of postures, perspectives or illumination conditions. Hence, the face could be detected and located according to these features.

* Corresponding author. Tel.: +86 13913977866.
E-mail address: yqwang@nju.edu.cn (Y. Wang).

http://dx.doi.org/10.1016/j.image.2015.06.010
0923-5965/

Nonetheless, these features can be seriously damaged due to illumination, noise or shadows. It would be difficult to carry out these methods when the boundaries of the features are weakened or the edges of the shadow are intensified. Template matching method is a method which firstly stores several standard face modes to describe the whole face and facial features respectively, and then calculates the relationship between the input image and stored modes [8]. Haar feature [9] is one of the most common used feature which uses gray and gradient information of image. The classic deformable templates are Active Shape Models (ASM) [10,11] and Active Appearance Models (AAM) [12]. Considering the deflection of face, Active Performance Model (APM) which is based on appearance [13] and Smooth Statistical Shape Model (SSSM) [14] are developed. Nevertheless, template matching methods involve a large amount of calculation and have low recognition rate for the faces with scale and rotation transformations. Taking full advantage of statistical analysis and machine learning, statistical-based method obtains statistical features of positive and negative samples from training images so that the classifiers could be built to detect faces. Statistical-based method is related to many kinds of methods, such as subspace method [15], Neural Network (NN) [16,17], Support Vector Machine (SVM) [18], Hidden Markov Model (HMM) [19], and a Boosting method [20]. Based on these methods, face detection systems could get good performance in terms of accuracy and detection speed, as well as good effect in circumstances with multi-posture faces. However, there are many requirements for training samples. For example, the accuracy of these methods depends on each training feature. At the same time, the long training time is also an influence factor.

Due to the disadvantages of the methods mentioned above, it could be an imperative requirement to find a "biological visual" recognition method. In the theory of visual perception, there is a "binding problem" that concerns the way in which people select and integrate the separate features of objects in the correct combinations [21]. Therefore, the concept of Feature Binding (FB) in the field of pattern recognition is proposed by combining with the theory of visual sensing. It is a method based on feature subspace. The features on a human face are grouped according to their locations, and the cluster of all the features in each group is regarded as a "feature set". During face detection, each feature set is detected respectively at first. The presence of a certain human face could be confirmed once the amount of constrained feature sets have reached the threshold, and the face can be located according to the locations of these feature sets. The results show that the accuracy of face location algorithm with FB could be improved more obviously compared with the traditional method. Meanwhile, the detection process can be accelerated and protection from the external interference better with this method. Besides, it has also shown good detection results for partial obscured faces in practical applications, because it only needs to reach the thresholds of countable feature sets rather than all of them. In addition, FB is an adaptive and robust method which can be applied to a variety of features rather than a particular one. Moreover, in circumstances with many kinds of features mentioned above, such as Haar feature and ASM feature, FB can be utilized to obtain better detection results.

The paper is organized as follows. In Section 2, the concept of Feature Binding is introduced. Section 3 describes a case used for Haar feature in which the advantage of the proposed approach would be demonstrated. Experiments and another case which is used for ASM are shown in Section 4, and Section 5 provides some concluding remarks.

2. Concept of Feature Binding

2.1. Feature Binding in biological vision

In biological perception theory, "feature" means an irreducible attribute of an object such as its color, orientation, form, and motion, and "binding" is the dynamic linkage of multiple features leading to the perception of a given object as a coherent and unified whole [22].

Recent findings have suggested that a primate brain codes perceived events in a distributed fashion, which are integrated into object files – episodic bindings of object-related information. Hommel put forward that the brain addresses these problems by creating multi-layered networks of bindings-"event files". These bindings produce systematic but often surprising and counter-intuitive interactions between perception and action planning [23] as well as their impairments.

Researchers in neurophysiology field have discovered that vision cells of visual pathway at all levels generally have receptive field property [24]. According to the property of reception field, cells on the visual cortex can be divided into simple cells, complex cells and hypercomplex cells. Simple cells, which are also called as direction selectivity cells, are suitable for the detection of contrasty straight edge. Complex cells have larger receptive field than simple cells. Besides, they have certain directions and shift invariance properties which are good for invariant features detection. There are some requirements for the length of the strip stimulation of hypercomplex cells, so that the optimal stimulus which can cause a strong reaction is the endpoint or the inflection point with a certain directivity. Hubel and Wiesel put forward the famous Receptive Field Level hypothesis through the research on visual cortex cells, which assumes that the receptive field of senior neuron is converged orderly by many lower neurons [25,26].

2.2. Feature Binding in pattern recognition

In the theory of computer vision, it is a common method to use local area as the processing element. Local areas have been divided artificially and regularly in previous algorithms, which lead to dispersive information and inconspicuous regional characteristics. In order to get a relatively large difference between each two local areas and obtain similar features in each area at the same time, the image shall be divided according to its own feature attributes. In the field of pattern recognition, this problem can be solved through feature subspace. With this method, test image is projected to varied feature spaces and the optimal features in these spaces best capable of distinguishing samples are selected. Besides, these features could constitute a feature subspace with certain rules [27,28].

From the perspective of information theory, the image can be divided into redundant part and mutational part according to its composition. People are often more sensitive to the mutational part than the redundant part during the observation of the target. The reason is that human visual system could inhibit the characteristic response which appears frequently and keep sensitive to unconventional mutational features. In addition, these redundant and mutational features correspond to the background and object of the image respectively.

In face detection, the features in the redundant regions of a human face only play a supporting role and represent a general trend, while the features in the mutational regions, such as eyes, nose and mouth, can convey facial information better and play a very important role. Therefore, the mutational part has more application value in practical applications and redundancy just for the sake of accelerating detection and obtaining mutation.

According to the theories mentioned above, a concept of Feature Binding (FB) in the field of pattern recognition is proposed. It is a method based on the feature subspace. The principle of FB is shown in Fig. 1. The detected image is calculated and mapped to feature subspaces according to certain rules, and quantified features corresponding to each subspace can be obtained. Some of the features are mapped from the entire or most part of the detected image, and the others are mapped from a local region or even an individual pixel. The most representative and significant areas of the image are called “hot region”, and each feature fully contained in one hot region is called a “hotspot”. For each hot region, all hotspots in it are picked out and bound into a group, which is called as “feature set”. Particularly for face detection, human eyes are the most important areas of a face, as a result, both the left eye and right eye are selected as two hot regions and both clusters of all features of them are bound for two feature sets separately. Similarly, nose and mouth could be selected as hot regions and features of these regions could be bound respectively. These bound groups are calculated in advance and the results are important basis in the process of object detection. Besides, each feature has a value, and the whole value of a feature set is calculated by all feature values contained in this group. Meanwhile, an appropriate threshold is selected as the criterion for this feature set. In the detection process, the current detected area could be identified to be the corresponding hot region when the whole value exceeds the threshold. The identified area calls as a “feature window”. Eventually, the detection and location result of the object is

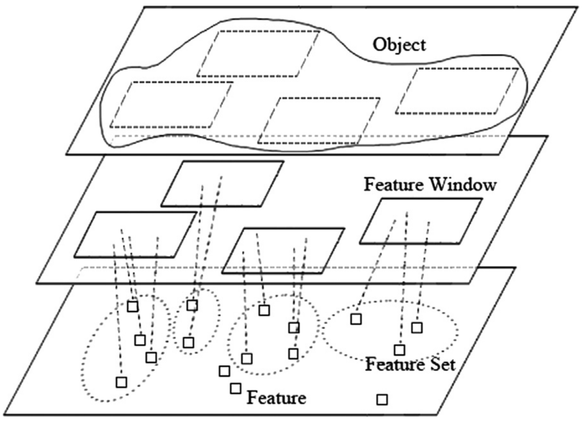

Fig. 1. Principle of Feature Binding.

determined by the feature windows and corresponding weights of each feature set. Due to the difference in the importance, the threshold and weight are also different for each group. In this algorithm, groups of two eyes have higher weight than those of nose and mouth.

In addition, FB is a wide practical method not limited to any case with specific features. For circumstances with many kinds of features, FB could be applied, and could achieve good detection effect as long as these features can be divided into several groups according to certain rules. In Section 3, one application will be described in detail. Another application example will be shown in Section 4.

3. Feature Binding used for Haar feature

3.1. Haar feature and traditional Adaboost algorithm

A human face has many obvious and regular characteristics, such as symmetry of face, edge curves and gradient directions of sense organs. Classification based on Haar feature can balance detection accuracy and speed and has a great success in object detection [29]. Constituted of two or more black and white rectangles, each Haar feature represents the local gray variance. Its feature value is defined as the difference in the sum of pixels which are covered by the white rectangle and the black one. There are five classical types of Haar features, as shown in Fig. 2. The variation of horizontal and vertical gray can be represented by the features shown in Fig. 2(a)–(d) respectively. Haar feature shown in Fig. 2(e) and its mirror feature express the variation of gray at the angle of $\pi/4$ and the angle of $3\pi/4$. For the same type, Haar features are also different in different locations and scales. The large scale feature represents the global brightness and contrast, whereas the small scale one represents the local boundaries and details. Taking the 24×24 rectangular window for example, the total number of the above five kinds of Haar features could reach more than 160,000. A large number of features can lay a good foundation for the description of the image texture information and structure weak classifiers.

Because of the good expression of texture and gray information, Adaboost machine learning algorithm based on Haar feature is adopted to detect and locate human face [30]. In this system, there are $T = 790$ Haar features selected after training. For a detected sample x, one Haar feature is an optimal weak classifier denoted by $f_t(x)$ in the t round of training. Eventually, a strong classifier $H(x)$ is constituted of these weak classifiers

$$H(x) = sign\left(\sum_{t=1}^{T} h_t(x) - thr\right), \tag{1}$$

Therein, $h_t(x)$ is confidence coefficient of f_t, and thr represents the threshold. The calculated region is identified as a face when$H(x) \geq 0$, and as a non-face otherwise.

3.2. Adaboost algorithm using Feature Binding

In a sense, a human face can be regarded as a stable geometry. Wherever it is in the image and whether there

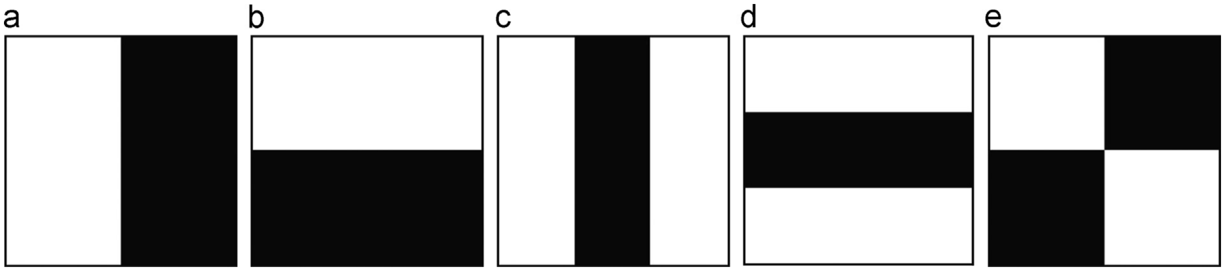

Fig. 2. Five classical types of Haar features.

is an obstruction, the face can keep a relatively inherent shape. Although a face is obscured partly, the shape of the rest part of the face still has not changed. However, in traditional Adaboost algorithm, the whole face is regarded as an entirety and the relative positions of each feature subsets are fixed. The positions of eyes, nose, mouth and others are on the basis of the standard training face. They can only change together rather than separately. Once the face is obscured, causing some weak classifiers to make the wrong judgment, the face may not be detected correctly. Therefore, Adaboost algorithm is improved by using Feature Binding. The improved algorithm can detect local features better and has good flexibility of the relationship between the locations of feature subsets.

The model of the improved algorithm is similar to a Bayesian model, as shown in Fig. 3. There are three layers in this model. The first layer is basic feature layer which is composed of five types of Haar features mentioned above. The middle layer is Feature Binding layer which has some feature sets. Each feature set consists of many Haar features with different positions and scales. The third layer is feature window decision layer. All the features in one feature set determine whether there is a corresponding feature window and where the window is.

In order to obtain the feature sets in the second layer, each Haar feature is classified according to its position on the standard face firstly. All the features with different types, sizes and locations collectively reflect the textural feature of different regions and the variation of gradient of different scales. These Haar features can be divided into two categories: global Haar feature and local Haar feature. A global feature covers many facial organs with larger size on the face, showing the variation of the gray and texture of the whole face. However, a local feature belongs to one organ and has relatively smaller size, representing changes of details. Some Haar features on the standard face are shown in Fig.4. Fig.4(a)–(d) are local features completely fall within the region of left eye, right eye, nose and mouth in sequence. Fig.4(e) and (f) are local features in other areas and Fig.4(g) and (h) are global features which cover more than one region.

In order to choose features which meet the position requirements above, selection rules are set up. The standard face is divided into five parts, which are left eye region, right eye region, nose region, mouth area and the rest region, as shown in Fig. 5.

For one region, taking the nose region for example, the coordinate of the upper left point is (r_1, s_1), and the bottom right point is (r_2, s_2). Coordinate of each pixel in this region is $p_{m,n}$, where $m \in [r_1, r_2]$, $n \in [s_1, s_2]$. For each Haar feature,

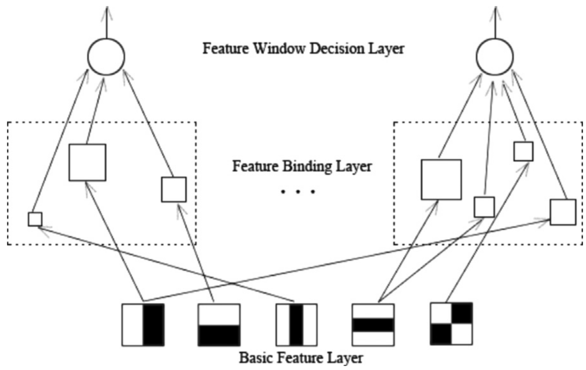

Fig. 3. The model of Feature Binding applied to Adaboost.

the upper left point is $p_1(m_1, n_1)$ and the bottom right one is $p_2(m_2, n_2)$. If

$$\begin{cases} m_1, m_2 \in [r_1, r_2] \\ n_1, n_2 \in [s_1, s_2] \end{cases}, \tag{2}$$

This Haar feature is considered to belong to this region totally and called a “hotspot” of this region. All these features constitute a “feature set”.

As left and right eyes, nose and mouth are the most distinguishable and important hot regions of human face, all the local features which are located in the left eye region are selected and bound together to be an entirety which is named as left eye set. In the same way, right eye set, nose set and mouth set can be established respectively. The left eye set, for example, is detected as a whole region at first when Adaboost is utilized for face detection. Given that the number of local features included in left eye set is N_{LE}, for three kinds of basic Haar features which are shown in Fig. 2(a), (b) and (e), the feature value of each one is

$$v_{\mathrm{LE}t}(x) = \sum\nolimits_{(k_1,l_1) \in Z_w} i(k_1, l_1) - \sum\nolimits_{(k_2,l_2) \in Z_b} i(k_2, l_2). \tag{3}$$

For another two kinds as shown in Fig. 1(c) and (d), the feature value is

$$v_{\mathrm{LE}t}(x) = \sum\nolimits_{(k_1,l_1) \in Z_w} i(k_1, l_1) - 2 \times \sum\nolimits_{(k_2,l_2) \in Z_b} i(k_2, l_2). \tag{4}$$

wherein Z_w and Z_b are areas covered by the white and black rectangles respectively. $i(k, l)$ is the gray value of the pixel (k, l) in the detected image.

Each feature value is normalized to $v_{\mathrm{LE}t}(x) \in [0, 1]$ and mapped to feature subspace which is equally divided into

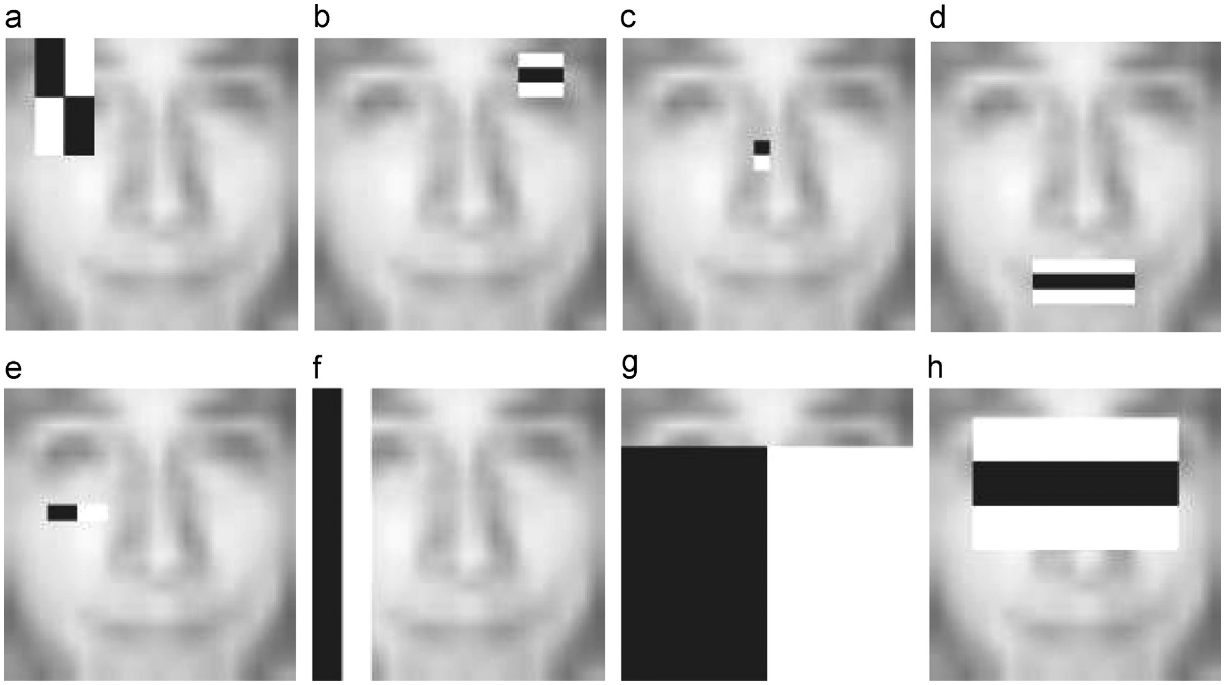

Fig. 4. Haar features of standard face.

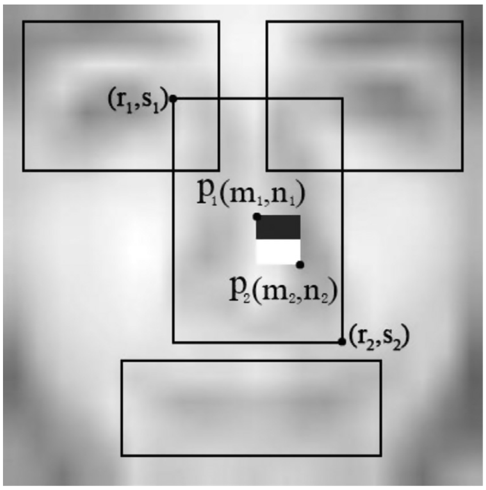

Fig. 5. Divided parts of standard face.

50 parts with each division denoted as follows:

$$bin_j = [(j-1)/50, j/50],\ j = 1, 2, \ldots, 50. \tag{5}$$

The confidence coefficient $h_{\mathrm{LE}t}(x)$ for each feature can be obtained through the following equation:

$$h_{LEt}(x) = \frac{1}{2}\ln\left(\frac{\overline{W}^j_{+1} + \varepsilon}{\overline{W}^j_{-1} + \varepsilon}\right). \tag{6}$$

wherein, $\overline{W}^j_{+1}$ and $\overline{W}^j_{-1}$ are the positive and negative probability respectively, which can be found in lookup table by training. ε is a tiny positive constant in case that the denominator is zero.

All bound features in this set constitute a strong classifier, and its confidence coefficient is

$$H_{\mathrm{LE}}(x) = sign\left(\sum_{t=1}^{N_{\mathrm{LE}}} h_{\mathrm{LE}t}(x) - thr_{\mathrm{LE}}\right). \tag{7}$$

wherein, thr_{LE} is threshold of left eye set. After Adaboost training, $thr_{\mathrm{LE}} = 1.2$ in our system. When $H_{\mathrm{LE}}(x) \geq 0$, the present detected area is identified to be a left eye and the present position is marked as the "left eye window". With the same method, threshold of right eye set $thr_{\mathrm{RE}} = 1.4$, threshold of nose set $thr_N = 0.8$ and of mouth set $thr_M = 1.1$ are obtained after training. $H_{\mathrm{RE}}(x)$, $H_N(x)$ and $H_M(x)$ which are confidence coefficients of strong classifiers corresponding to right eye set, nose set and mouth set respectively can be obtained. Then, right eye window, nose window and mouth window can also be located.

In the third layer, whether there is a complete face composed of these feature windows shall be determined. Confidence coefficient of the face is the weighted sum of all strong classifiers

$$H_{Face}(x) = sign([\omega_{\mathrm{LE}} H_{\mathrm{LE}}(x) + \omega_{\mathrm{RE}} H_{\mathrm{RE}}(x) + \omega_N H_N(x) + \omega_M H_M(x)] - thr_{Face}). \tag{8}$$

wherein thr_{Face} is the threshold, andω_{LE}, ω_{RE}, $\omega_N$andω_M are the weights of each feature window. The areas of two eyes are considered to have minimum differences and be the most difficult to be obscured among the areas of the whole face under the influence of various factors such as mask and hat. Meanwhile, there are a large number of features in these areas after training because the eye regions contain a lot of identifiable texture and gradient information. In addition, accurate positioning of two eyes is the most important step in the non-auxiliary stereo display system. For these reasons, the largest weight coefficients for left eye window and right eye window are set up. Considering that the diversity of mouth is larger than nose, the weight of mouth window should be less than nose window. Assume $\omega_{\mathrm{LE}} = \omega_{\mathrm{RE}}$ because the two feature sets have the same importance in detection. Weight of each feature set is tested in corresponding range, as shown in Fig. 6. Therefore, in our system, $\omega_{\mathrm{LE}} = \omega_{\mathrm{RE}} = 0.3$, $\omega_N = 0.15$ and $\omega_M = 0.25$. If the weighted sum value of all voting values of strong classifier is larger than thr_{Face}, a certain face is identified and the location of the face window is decided according to the positions of each feature window. In our system, $thr_{Face} = 1.2$ is determined by Adaboost training.

Although there are 790 Haar features obtained through training, a total of 215 features which meet the position requirements are selected for FB. A large portion of the rest features cover two or more feature sets, as shown in Fig.4 (g) and (h). In order to avoid the situations such as the left eye window on the right side of the right eye window, the rest features are used as constraints for the relative position of each feature window. Some false detections, such as the pattern which is similar to one single set in the background image, can also be eliminated under these constraint conditions and the total voting value.

Face detection system based on FB is shown in Fig. 7. Different from the traditional structure of serial classifiers, all the parallel classifiers process the image severally.

Given that the false negative rate (FNR) is α and the false positive rate (FPR) is β for any strong classifier after an appropriate threshold is set, wherein, $0 < \alpha, \beta < 1$. The total FNR of traditional Adaboost is

$$\alpha_A = \alpha + (1-\alpha)\alpha + \ldots + (1-\alpha)^{n-1}\alpha = 1-(1-\alpha)^n. \tag{9}$$

The total FPR is

$$\beta_A = \beta^n. \tag{10}$$

wherein n is the number of strong classifiers.

For the algorithm with FB as shown in Fig. 5, only when two or more strong classifiers have false negative detection, the face is likely to be determined as a non-face. The FNR of the proposed algorithm is

$$\alpha_{\mathrm{FB}} = C_4^3\alpha^3(1-\alpha) + \alpha^4 = 4\alpha^3 - 3\alpha^4. \tag{11}$$

Similarly, a non-face image may be determined as a face when two or more strong classifiers have false positive detection. The FPR of the proposed algorithm is

$$\beta_{\mathrm{FB}} = \beta^2(1-\beta)^2 + C_4^3\beta^3(1-\beta) + \beta^4 = \beta^2(1+4\beta-4\beta^2). \tag{12}$$

It is shown that $\alpha_{\mathrm{FB}} < < \alpha_T$, which means the FNR of the proposed algorithm is much less than that of the traditional algorithm. Although $\beta_{\mathrm{FB}} > \beta_T$, the FPR of the proposed algorithm can be reduced into the tolerance range. In a word, the detection rate of the algorithm with FB is improved obviously.

The face detection result is shown in Fig.8. The black face window is the location result of traditional algorithm and the white one is detected by the algorithm with FB. The white dotted windows are four feature windows. It can be seen that Feature Binding could have more accurate detection and location result.

Moreover, traditional Adaboost algorithm does not have priori assumption for the feature areas in both training and testing, which may cause some troubles when detecting a face which is obscured partially. For one classifier, FNR and FPR is a pair of contradiction. If it is necessary to reduce FPR, the threshold of this classifier

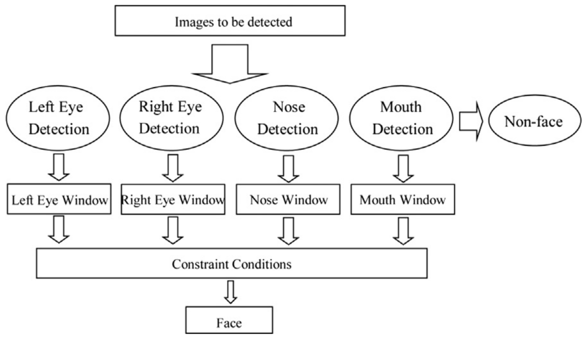

Fig. 7. Face detection system with Feature Binding.

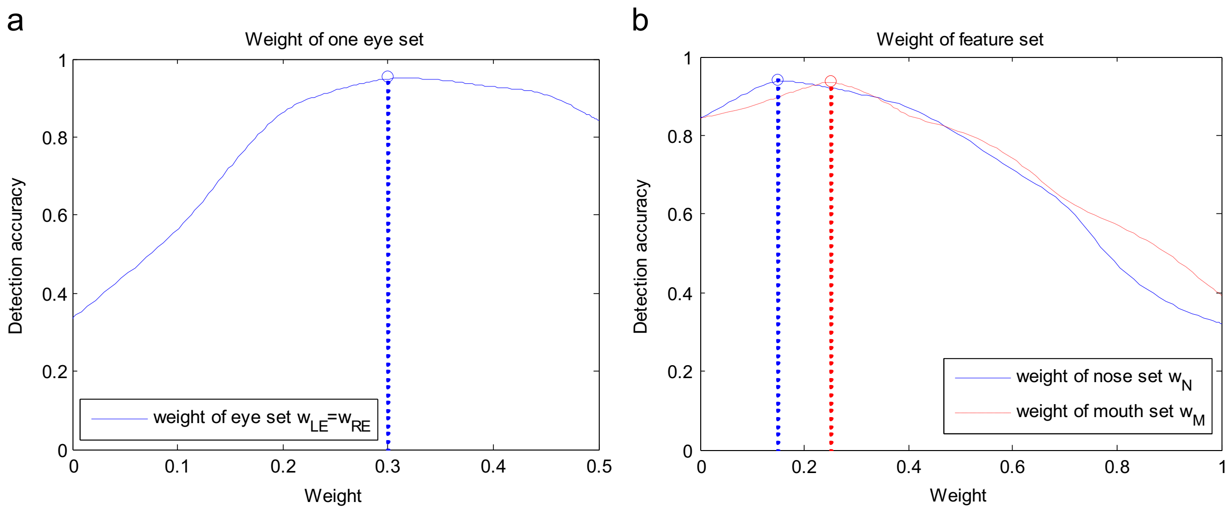

Fig. 6. Weight of each feature set.

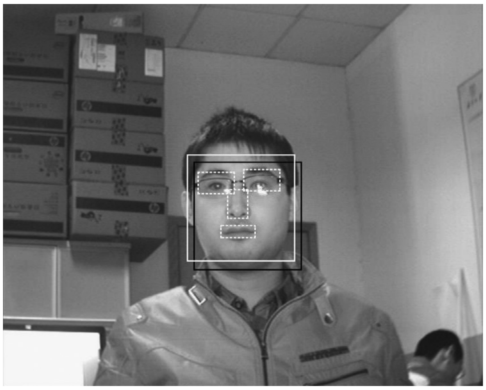

Fig. 8. Face detection result of Feature Binding.

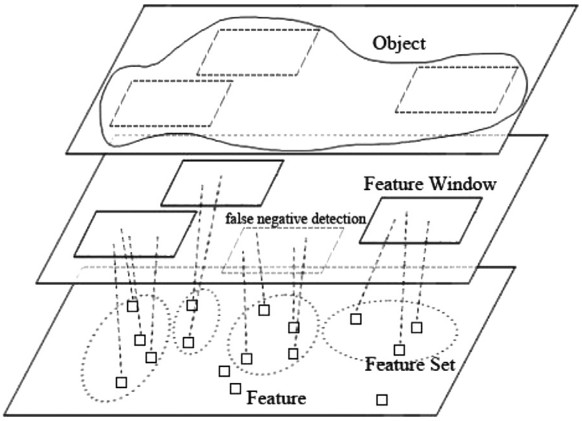

Fig. 9. Obscured object detection by Feature Binding.

needs to be increased. In this case, an obscured face could not pass through one strong classifier when major weak classifiers are on the obscured part and the detected image will be determined as a non-face. If it is necessary to get a good tolerance on partial obscured face, the threshold for each classifier shall be reduced to decrease FNR which may increase FPR. However, the proposed FB method could balance the two aspects effectively. As shown in Fig. 9, four feature sets detect the image respectively. Although the vote value(s) of one or two sets is/are negative, the other sets could also vote to positive. Especially, the face can be detected successfully when eyes are not covered in the obscured area because of the higher voting weight of left and right eye sets.

In conclusion, there is one result of obscured face detection with traditional and improved algorithms, as shown in Fig. 10. It is an image of Caltech datasets [31]. The black window represents the result of traditional Adaboost algorithm. Because the face is blocked by user's hand, traditional algorithm cannot detect the face correctly. The white window is the result of FB method and white dotted windows are feature windows.

4. Experiments

For images of Caltech datasets, face detection results by traditional Adaboost algorithm and FB method are shown

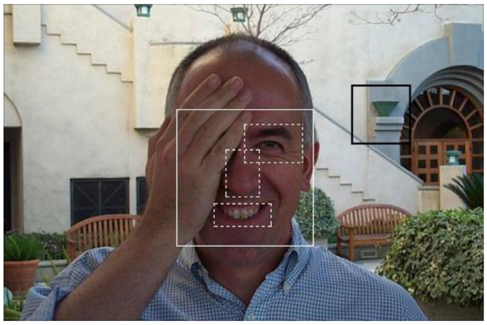

Fig. 10. Obscured face detection with traditional Adaboost algorithm and FB method.

in Fig. 11, which are represented of the black and white windows respectively.

For images of FDDB dataset [32], face detection results by FB method are shown in Fig. 12, which are represented of the white windows.

For one image sequence which contains 121 frames captured in our laboratory, face detection results by traditional Adaboost algorithm and FB method are shown in Fig. 13, which are represented of the black and white rectangular windows respectively. In order to compare the two detection methods, there is no tracking algorithm involved in our system. Every image of the sequence is detected independently and has no influence on the next frame.

Fig. 14 is the face location comparison of traditional Adaboost algorithm, FB method and the actual position of the above sequence. Fig. 14(a) and (b) are the abscissa and vertical comparison, (c) and (d) are partial details of (a) and (b) respectively. It is shown that the FB method improves the detection result obviously. Because the face location of FB is determined by locations of all feature windows, it could be more accurate than traditional method which detects the entire face. The scale of face window is also more accurate.

For the images captured in our laboratory which have different facial expressions, varying lighting conditions and different occluded, FB has very good effect on all of them. As shown in Fig. 15, white windows are face locations detected by FB.

It is worth noting that FB does not adapt to one particular kind of feature. In theory, it could be used for any feature which could be classified according to certain rules. We also use FB for Active Shape Model (ASM) [10]. For one image sequence, the results of face detection separately by traditional ASM algorithm and FB method are shown in Fig. 16, which are represented of the black rectangular window and the white one respectively. It can be seen that the detection and location results of FB are better than traditional ASM algorithm and the scale of face window is more accurate. In order to compare the two detection methods, there is no tracking algorithm involved in our system. Every image of the sequence is detected independently and has no influence on the next frame.

Mean Squared Error (MSE) is used to evaluate the performance.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2. \tag{13}$$

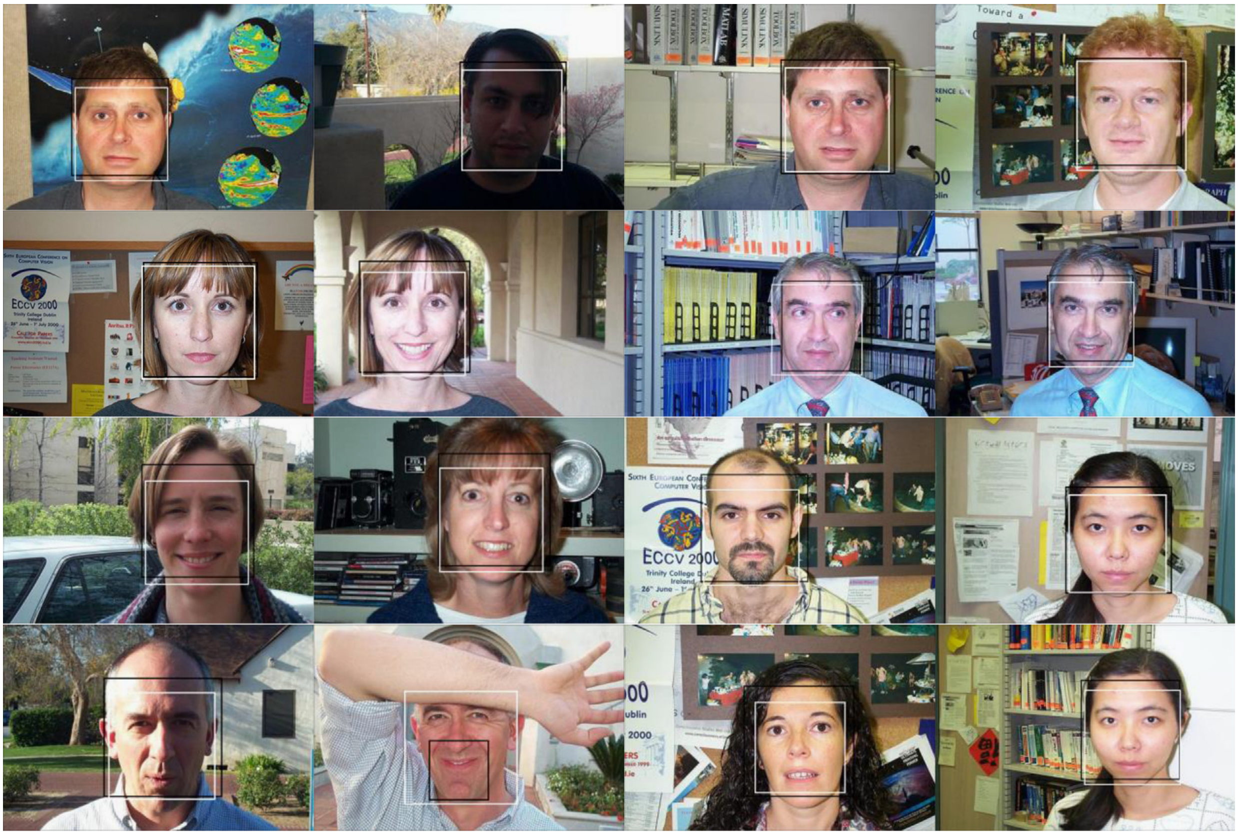

Fig. 11. Face detection results by traditional Adaboost algorithm and FB method.

Fig. 12. Face detection results by FB method on FDDB dataset.

wherein n is the length of the image sequence, $\hat{Y}_i$ is the detection result and Y_i is the actual value. The smaller the MSE is, the better the performance is.

For the datasets above, detection results are shown in Table 1. It is shown that FB has a higher accuracy and lower MSE than corresponding traditional algorithms.

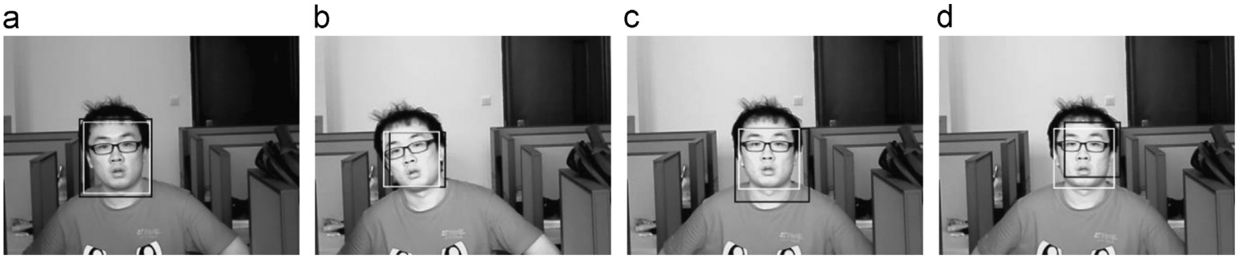

Fig. 13. Face detection results of image sequence using traditional Adaboost algorithm and FB method.

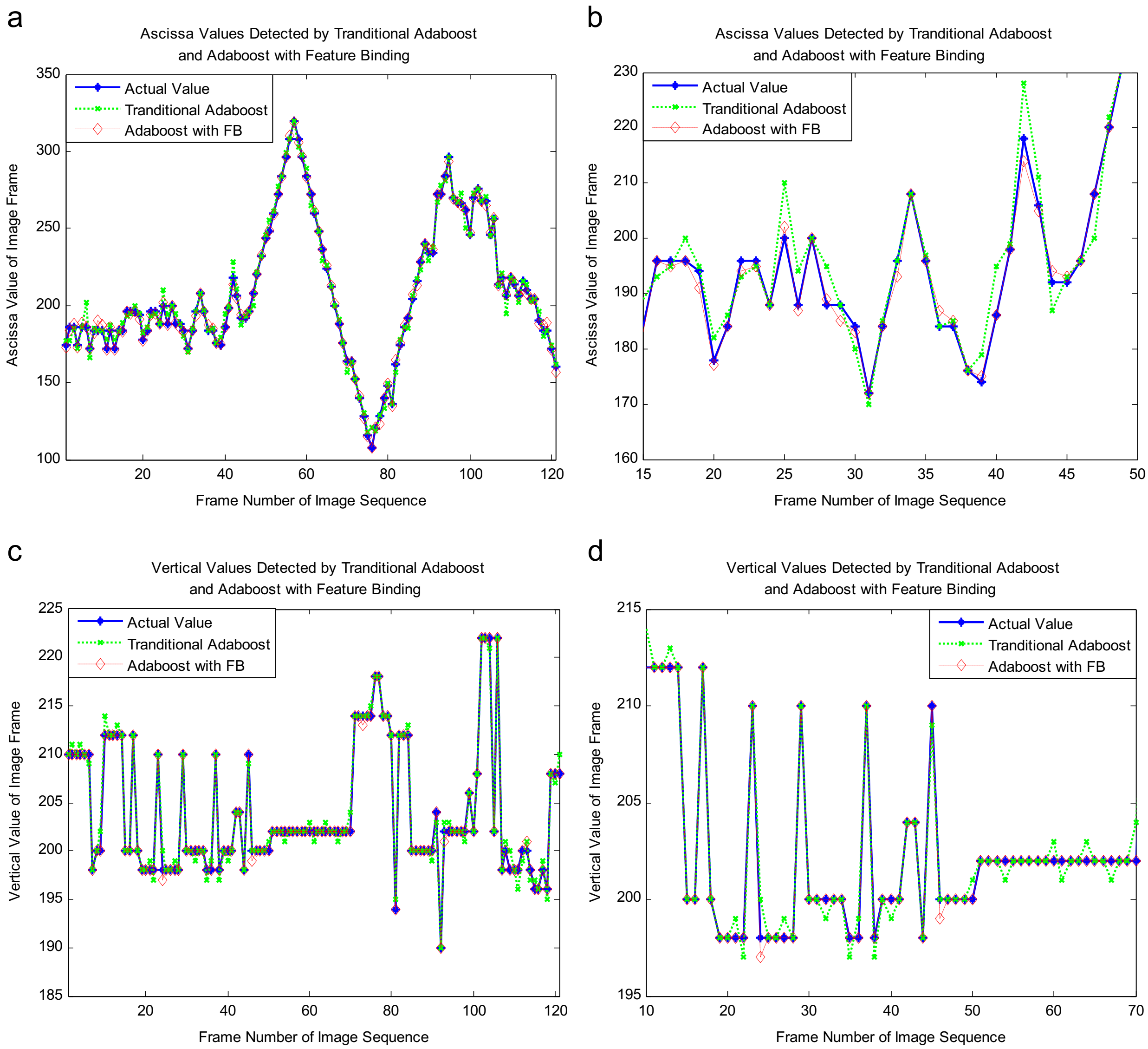

Fig. 14. Face location comparison of traditional Adaboost algorithm, FB method and the actual position.

Fig. 17 is Receiver Operating Characteristic (ROC) curves of detection methods. It is shown that FB method has larger Area Under Curve (AUC) which means better detection performance.

5. Conclusion

In this paper, we propose the concept of Feature Binding in pattern recognition and use it to detect and

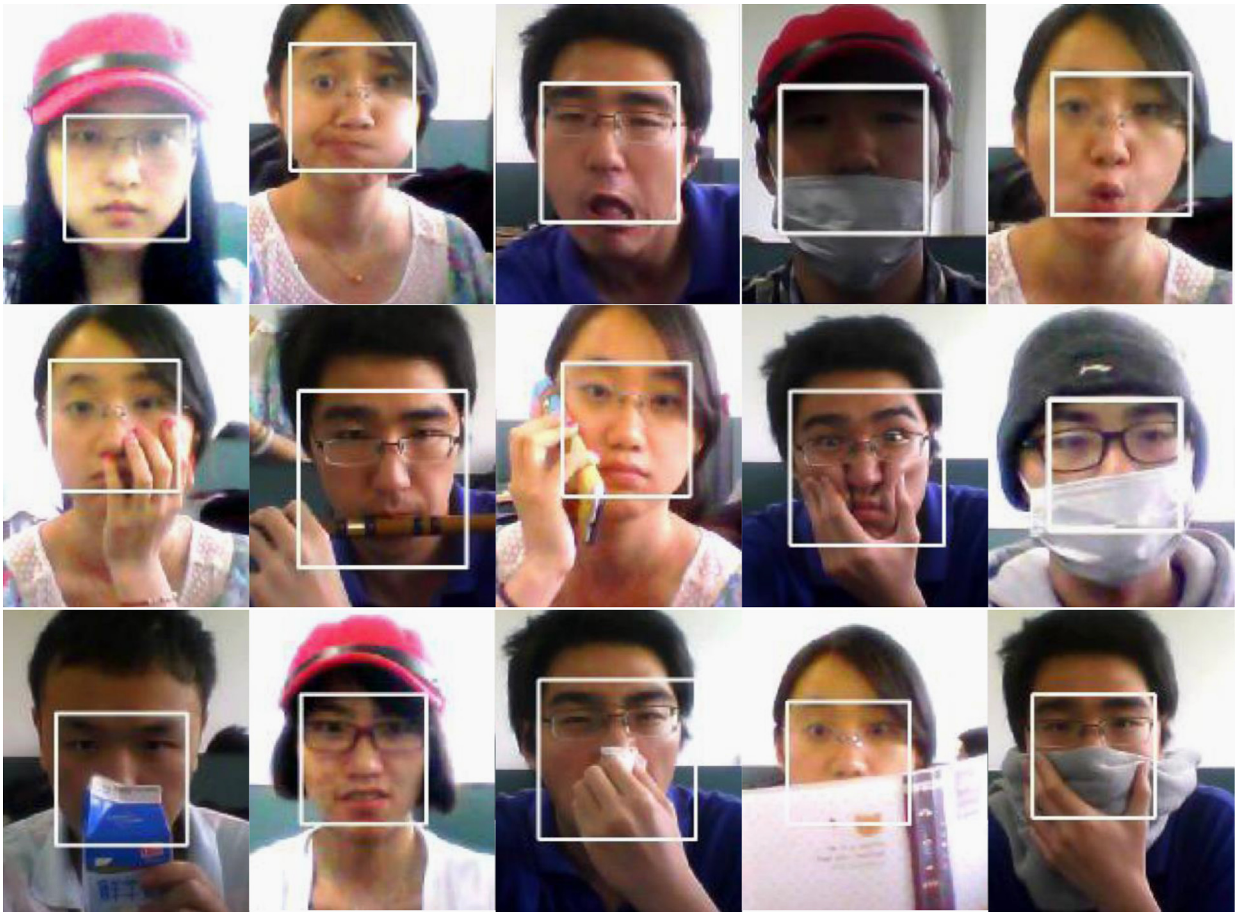

Fig. 15. Face detection results with FB method.

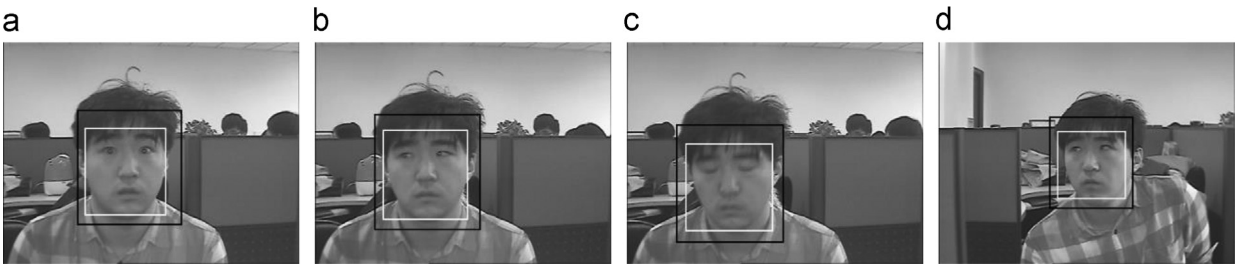

Fig. 16. Face detection results of image sequence using traditional ASM algorithm and FB method.

Table 1
Face detection results.

	Accuracy	MSE
Traditional Adaboost	95.92%	5.1733
Adaboost with FB	98.12%	3.3563
Traditional ASM	97.97%	4.7579
ASM with FB	98.95%	3.1345

locate human face. The basic idea of FB is dividing all features into several groups according to certain rules and binding each group as a feature set. We divide and bind features into left eye set, right eye set, nose set and mouth set on the basis of their distribution on the face. Each feature set is detected separately and has independent confidence coefficient. A whole face is identified by the sum of confidence coefficients of each feature set. In addition, FB has no specificity so that can be used in a variety of traditional algorithms. In experiments, it is used for Haar feature and ASM feature which are based on the statistical theory and template matching. Detection results have improved accuracy obviously. Moreover, FB is also effective for the obscured face.

In order to improve our method, there are still several goals we should reach in our future work. Firstly, our method should be robust when faces rotate substantially. Secondly, our method should detect faces correctly when

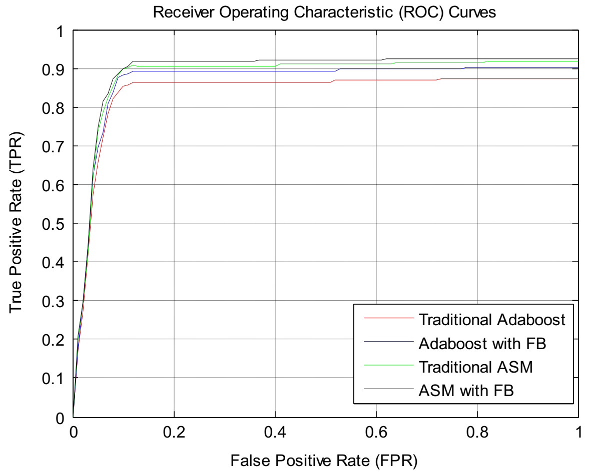

Fig. 17. Receiver Operating Characteristic (ROC) curves.

there are more than one user in the image. For different eyes, FB should distinguish them from different users. Thirdly, for each feature set detection module, it can be accelerated using cascades structure in theory. However, it needs to be proved in experiments quantitatively.

Acknowledgments

This research is sponsored by National Science and Technology Major Project (AHJ2011Z001). The authors would like to acknowledge the contributions of Nanjing University stereo imaging technology (SIT) laboratory 3D imaging team.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.image.2015.06.010.

References

[1] Yalan Xue, Yuanqing Wang, Multi-User autostereoscopic 2D/3D switchable flat-panel display, J. Disp. Technol. 10 (9) (2014) 737–745.
[2] Guangzheng Yang, Thomas S. Huang, Human face detection in complex background pattern recognition, Pattern Recognit. 27 (1) (1994) 53–63.
[3] Lu Chunyu, Changshui Zhang, Fang Wen, Yan Pingfan, Regional feature based fast human face detection, J. Tsinghua Univ. (Sci. Technol.) 39 (1) (1999) 101–105.
[4] Zhang Pinzheng, Zhao Hongyu, Liang Xiaoyun, Shu Huazhong, Xu Qinzhen, Face detection and validation method in complex background, J. Data Acquis. Process. 19 (1) (2004) 10–15.
[5] Huang Chun-Hsian, A reconfigurable point target detection system based on morphological clutter elimination, J. Syst. Archit. 60 (8) (2014) 644–654.
[6] Lorenzo Livi, Antonello Rizzi, Alireza Sadeghian, Optimized dissimilarity space embedding for labeled graphs, Inf. Sci. 266 (2014) 47–64.
[7] Poonam Sharma, K.V. Arya, R.N Yadav, Efficient face recognition using wavelet-based generalized neural network, Signal Process. 93 (6) (2013) 1557–1565.
[8] Jun Miao, Baocai Yin, Kongqiao Wang, Lansun Shen, Xuecun Chen, A hierarchical multiscale and multiangle system for human face detection in a complex vackground using gravity-center template pattern recognition, Pattern Recognit. 32 (7) (1999) 1237–1248.
[9] Chao Yan, Yuangqing Wang, Zhaoyang Zhang, Robust real-time multi-user pupil detection and tracking under various illumination and large-scale head motion, Comput. Vis. Image Underst. 115 (1) (2011) 1223–1238.
[10] T F Cootes, C J Taylor, D H Cooper, Active shape model-their training and application, Comput. Vis. Image Underst. 61 (1) (1995) 38–59.
[11] Xiujie Zhao Xianmei Wang, Liying Jia, Ping Yang, Pupil localization for multi-view eyeballs by ASM and eye gray distribution, Procedia Eng. 15 (2011) 2993–2998.
[12] T.F. Cootes, G.J Edwards, and C.J. Taylor, Active appearance models, in: Proceedings of the 5th European Conference on Computer Vision, vol. 2, 1998, pp. 484–498.
[13] T F Cootes, G V Wheeler, K N Walker, C J Taylor, View-based active appearance models, Image Vis. Comput. 20 (2002) 657–664.
[14] T F Cootes, C J Twining, C J Taylor, Diffeomorphic Statistical Shape Models, Br. Mach. Vis. Conf. 1 (2004) 447–456.
[15] Xiaogang Wang, Xiaoou Tang, Random sampling for subspace face recognition, Int. J. Comput. Vis. 70 (1) (2006) 91–104.
[16] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade, Neural network-based face detection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996, pp. 203–208.
[17] Cha Zhang, and Zhengyou Zhang, Improving multiview face detection with multi-task deep convolutional neural networks, in: Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), 2014, pp. 1036–1041.
[18] N.H. Dardas, Georganas, D Nicolas, Real-time hand gesture detection and recognition using bag-of-features and Support Vector Machine techniques, IEEE Trans. Instrum. Meas. 60 (11) (2011) 3592–3607.
[19] P., Kumawat, A., Khatri, and B, Nagaria, Comparative analysis of offline handwriting recognition using Invariant moments with HMM and combined SVM–HMM classifier, in: Proceedings of the 2013 International Conference on Communication Systems and Network Technologies (CSNT), 2013, pp. 140–143.
[20] Zhang Zhiwei, Yi Dong, Lei Zhen, Stan Z. Li, Regularized transfer boosting for face detection across spectrum, Signal Process. Lett. IEEE 19 (3) (2012) 1070–9908.
[21] Anne Treisman, Feature binding, attention and object perception, Philos. Trans. R. Soc. B: Biol. Sci. 353 (1998) 1295–1306.
[22] Di Lollo Vincent, The feature-binding problem is an ill-posed problem, Trends Cogn. Sci. 16 (6) (2012) 317–321.
[23] Hommel Bernhard, Event files: feature binding in and across perception and action, Trends Cogn. Sci. 8 (11) (2004) 494–500.
[24] Hongwei Zhao, Xiao Chen, Liu Pingping, Geng Qingtian, Adapative segmentation for visual salient object, Opt. Precis. Eng. 21 (2) (2013) 531–538.
[25] Ma Runing, Tu Xiaopo, Ding Jundi, Yang Jingyu, To evaluate salience map towards popping out visual objects, Acta Autom. Sin. 38 (5) (2012) 870–876.
[26] M Sezgin, B Sankur, Survey over image thresholding techniques and quantitative performance evaluation, J. Electron. Imaging 13 (1) (2004) 146–168.
[27] Alessandro Sperduti, Antonina Starita, Speed up learning and network optimization with extended back propagation, Neural Netw. 6 (3) (1993) 365–383.
[28] Sudheer Ch, Nitin Anand, B.K Panigrahi, Shashi Mathur, Streamflow forecasting by SVM with quantum behaved particle swarm optimization, Neurocomputing 101 (2013) 18–23.
[29] P. Viola, and M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2001, pp. 511–518.
[30] Zhou Jin, Yuan-qing Wang, Ke-feng Fan, Face detection method for real-time anti-interference, Comput. Eng. Des. 34 (4) (2013) 1399–1403.
[31] L. Fei-Fei, R. Fergus and P. Perona, Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories, in: Proceedings of the IEEE, CVPR Workshop on Generative-Model Based Vision, 2004.
[32] Vidit Jain and Erik Learned-Miller, FDDB: a benchmark for face detection in unconstrained settings. Technical Report UM-CS-2010-009, Deptartment of Computer Science, University of Massachusetts, Amherst, 2010.