Efficient Face Candidates Selector for Face Detection

Jianxin Wu [a] Zhi-Hua Zhou [a,1]

[a] *National Laboratory for Novel Software Technology*
Nanjing University, Nanjing 210093, P.R.China
wujx2001@yahoo.com zhouzh@nju.edu.cn

Abstract

In this paper an efficient face candidates selector is proposed for face detection tasks in still gray level images. The proposed method acts as a selective attentional mechanism. Eye-analogue segments at a given scale are discovered by finding regions which are roughly as large as real eyes and are *darker* than their neighborhoods. Then a pair of eye-analogue segments are hypothesized to be eyes in a face and combined into a face candidate if their placement is consistent with the anthropological characteristic of human eyes. The proposed method is robust in that it can deal with illumination changes and moderate rotations. A subset of the FERET data set and the BioID face database are used to evaluate the proposed method. The proposed face candidates selector is successful in 98.75% and 98.6% cases respectively.

Key words: Face candidates selector, Face detection, Focus of attention, Eye-analogue segment.

[1] Corresponding author. Tel.: +86-25-3593163; fax:+86-25-3300710

List of Figures

List of Tables

1 Introduction

Automatic face recognition has attracted significant attention in image analysis and understanding, computer vision and pattern recognition for decades. Although it is very easy for human beings to track, detect, recognize, and identify faces in complex background, building an automatic face recognition system remains very difficult because the face images available can vary considerably in terms of facial expression, age, image quality and photometry, pose, occlusion and disguise [1].

In general, face recognition algorithms can be divided into two categories [2] [2]. The first category of the algorithms is based on geometric, feature-base matching which uses the overall geometrical configuration of the facial features, e.g. eyes and eyebrows, nose and mouth, as the basis for discrimination. It is clear that in order for those algorithms to work, face region and salient facial features should be detected before any other processing can take place. The second category of the recognition algorithms is based on template matching, in which recognition is based on the image itself. In those algorithms, faces should be correctly aligned before recognition, which is usually performed based on the detection of eyes. So, proper face and eye detection are vital to face recognition tasks.

Many face and eye detection methods have been developed in the literature. Most of the existing face detection algorithms can be put into a two-stage framework. In the first stage, regions that may contain a face are marked, i.e. this stage focuses attention to face candidates. In the second stage, the possible regions, or, face candidates, are sent to a "*face verifier*", which will

[2] Many new methods have been proposed in the literature since the publication of [2], however, most of them still can be categorized according to the taxonomy described in [2].

decide whether the candidates are real faces. Different methods put emphasis on different stages. An extreme is that if the face verifier is powerful enough to discriminate between face patterns and non-face patterns in nearly all cases, the candidate selection stage may be omitted. In this case the algorithm can move through the image from left to right, and from top to bottom, i.e. treat each sub-region in the image as a face candidate. On the contrary, if in the first stage most non-face regions are eliminated and all face regions are selected, the face verifier might be dramatically simplified or even omitted.

This paper proposes a new face candidates selector to efficiently select possible face regions from still gray level images for face detection purposes. Since color images can be turned into gray level ones easily, the proposed method can be applied to color images too. However, color images contain more useful information than gray level ones and can be processed more efficiently [3]. In the proposed method, eye-analogue segments are discovered by finding regions which are roughly as large as eyes and are *darker* than their neighborhoods. Then a pair of eye-analogue segments are treated as left and right eyes in a face pattern if their placement is consistent with the anthropological characteristic of human eyes. The image patch containing these segments is selected as a face candidate. Face candidates are searched in successively scaled images to find faces in different sizes. Since the selector only uses two robust cues in discovering eye-analogue segments, i.e. the lower intensity value and size of eyes, it is expected to be robust to illumination changes and moderate rotation in depth and rotation on-the-plane. Experiments on two face image data sets show that the proposed method can attain very good results even when it is accompanied with a weak face verifier, i.e., a simple variation of the template matching method.

The rest of this paper is organized as follows. A brief review on existing face detection methods is presented in Section 2. The proposed method is given in Section 3. Experimental results are reported in Section 4 and conclusions are

drawn in Section 5.

2 Brief review on face detection methods

This section briefly reviews the literature on face detection in still gray level images [3].

Some methods try to model distribution of the high dimensional face space. Turk and Pentland [6] applied principal component analysis to face detection where a set of eigenfaces were generated by performing principal component analysis on some training images. The reconstruction error of the principal component representation, referred to as *distance-from-face-space*, was computed for all locations in the image. The global minimum of this distance was marked as a face region. Moghaddam and Pentland [7] extended this idea by using density estimation in the eigenspace. They estimated the complete probability distribution of a face's appearance using an eigenvector decomposition of the image space. The desired target density was decomposed into two components: a density in the principal subspace and its orthogonal complement which was discarded in Turk and Pentland's method. Then a maximum likelihood detection scheme was applied in which the image location with highest likelihood was chosen as the face region. In those methods, the *distance-from-face-space* was computed for every image location, so those methods do not select the focus of attention.

The idea of estimating the distribution using a few linear subspaces is very useful. Yang, Ahuja and Kriegman [8] presented two methods of using mixture of linear subspaces for face detection in gray level images. The first method used mixture of factor analysis to model the distribution and the parameters

[3] For a more detailed review on the subject of face detection, please refer to [4]. For a review on face processing, please refer to [1] and [5].

of the mixture model were estimated by an EM algorithm [9]. Factor analysis (FA) is a method for modelling the covariance structure of high dimensional data using only a few latent variables. Mixture of factor analysis overcomes a serious shortcoming of PCA, i.e. PCA does not define a proper density model. The second method used Fisher's Linear Discriminant (FLD), where face and non-face images were divided into several sub-classes using Kohonen's Self Organizing Map, and then the within- and between-class scatter matrices were computed and the optimal projection based on FLD was also computed. In each sub-class, the density was modelled as a Gaussian. Maximum likelihood decision rule was used to determine whether a face was present. Both methods in [8] used brute force searching strategy. So they also lack mechanism for focus of attention. Sung and Poggio [10] developed an example-based learning approach for locating vertical frontal views of human faces in complex scenes. They built a distribution-based face model which was trained using 4150 face patterns and 6189 face like non-face patterns. These patterns were grouped into six face and six non-face clusters using the *elliptical k-means* clustering algorithm. Each cluster was represented by a prototype pattern and a covariance matrix. Two distance metrics, a Euclidean distance and a Mahalanobis distance, were defined for each cluster. Thus a pattern yielded 24 values. A multi-layer perceptron (MLP) was trained to determine whether a test pattern was actually a face based on these distance metrics. This method detects faces by exhaustively scanning an image. Note that in this method, a bootstrap strategy was used to find representatives of non-face patterns. This technique is followed in many other works.

Neural networks are often used in face detection. Rowley, Baluja and Kanade [11] proposed a face detection method based on neural networks. They trained neural networks that could discriminate between face and non-face patterns on a large set of face and non-face image patterns. Each neural network returned a score ranging from –1 to 1. This score being close to 1 indicated a

face pattern while –1 for a non-face pattern. A pyramid of images was built to detect faces in different scales. To improve performances of the system, they trained several neural networks and used various arbitration schemes such as logic operations and voting to combine the detection results of those networks. Such an approach was extended by Rowley, Baluja and Kanade [12] to detect rotated faces. In the extended version, the image patch was sent to a router network at first, which determined the rotation angle of the image patch. Then the image patch was rotated back to the upright direction and sent to the detector network which was the same as that in [11]. Feraud [13] presented a detection method based on a generative neural network model constrained by non-face patterns. A four-layer neural network was used to perform non-linear dimensionality reduction. Each non-face pattern was constrained to be reconstructed as the mean of the n nearest neighbors of face patterns. Multiple networks were also used to improve performances. Feraud tried ensemble, conditional mixture and conditional ensemble of CGMs. The conditional mixture and conditional ensemble acted as gate networks evaluating the probability that whether the input pattern was an upright face, a rotated face, or non-face. SVM (Support Vector Machines), a new way to train polynomial, neural networks, or radial basic function classifiers, has also been applied to face detection recently [14]. The method of creating representative sets of non-face patterns which was proposed in [10] was used in [11–14]. In these neural network based methods, emphasis were put on the face verifier part.

Template matching methods were used as early attempts to solve the face detection problem [2,15,16]. A template (a standard face pattern) was defined at first. Then different parts of an input image was compared with those in the template, usually correlation values were computed. Decision was based on those correlation values. Templates may be defined as the image itself, or edges, contours, invariants, Gabor wavelets, and in many other forms. How-

ever, simple template matching method is not robust since it can not deal with faces in different pose, scale, expression, and illumination condition. In [2] multi-scale templates were used to detect eyes in different scales. Yuille, Hallinan and Cohen [17] have used deformable templates [18] to extract facial features. They designed parametric models for eyes and mouth templates using circle and parabolic curves. The best fit of these models was found by minimizing an energy function of parameters of these models, such as the center and radius of the circle and coefficients of the parabola.

Burl, Weber and Perona [19] modelled the human face class as a set of characteristic parts arranged in a variable spatial configuration and applied this model to face detection in cluttered scenes. In their system, the face detector was composed of two terms: the first term measured how well the hypothesized parts (local facial features) in the input pattern matched the model parts, while the second term measured how well the hypothesized spatial arrangement of these parts matched the ideal model arrangement. To achieve translation, rotation, and scale invariance, a joint probability density [20] over several parts which is TRS-invariant was used. There was a trade-off between having the parts match well and having the shape match well. No exact solution was presented in [19], instead, a heuristic approach was suggested. In this approach, candidate parts were grouped into hypotheses as in the shape-only model, but the parts' match scores were retained and combined with the shape likelihood.

In the aforementioned face detection methods, efforts are concentrated on finding powerful face verifiers. However, there are also some methods which try to build efficient focusing mechanisms. Using an attention mechanism allows for intelligent control to deal with the allocation of computational resources in terms of *where*, *what*, and *how* to sense and process. In face detection applications with such mechanism, since only those regions of interest go through the complex and computationally expensive recognition stages, the system runs

faster, and perhaps, more accurately [21].

Yang and Huang [22] proposed a hierarchy knowledge-based method to detect faces. The system consisted of three levels. The highest two levels were based on mosaic images at different resolutions. A mosaic image was constructed by decreasing the resolution of the original picture. In the highest level, rules such as "The center part of the face (the quartet [4]) has four cells with a basically uniform gray level" which describe some common constraints on human face patterns were used to select face candidates. For level 2 a set of face detection rules were established on the basis of the octet [5]. Another set of more detailed rules were used to further screen out some non-face patterns. In level 3 local histogram equalization, multi-binarization, and bi-directional edge tracking were used. Then the features of the edges of eyes and mouth were extracted. If the extracted characteristics fitted well with the characteristics of eyes and mouth, the system would recognize it as a face pattern. This system's detection rate is not very high but some of its ideas such as mosaic images, coarse to fine searching and focus of attention mechanism are useful in designing successful face detection systems.

Han et al. [23] used morphological operations to locate eye-analogue pixels in the original image. Since eyes and eyebrows are the most salient and stable features of human face, a labelling process was used to generate the eye-analogue segments that guided the search for potential face regions. Amit, Geman and Jedynak [24] based their focusing mechanism on spatial arrangements of edge fragments. Following examples of faces, they selected particular arrangements that were more common in faces than in general background. Takács and Wechsler [25] proposed a focusing mechanism motivated by the *early vision*

[4] A quartet is defined as a local mosaic image, which has 4×4 cells and occupies the main part of a face [22].

[5] An octet is defined as a local mosaic image, which has 8×8 cells and occupies the main part of a face [22].

system. Huang, Liu, and Wechsler [26] used genetic algorithms to evolve some finite state automata that could navigate and discover the most likely eye locations, then they selected optimal features using genetic algorithms and built decision trees to classify whether the most salient locations identified earlier were eyes.

3 The proposed method

3.1 Overview of the proposed face candidates selector

Eyes are the most important facial features in face detection and recognition systems [23,26,27]. The proposed face candidates selector locates possible face patterns by combining two eye-analogue segments. At first all eye-analogue segments are found by finding regions that are roughly the same size of an real eye and are darker than their neighborhoods. If locations of a pair of these segments conform to the geometrical relationship of a human face's two eyes, the face candidates selector concludes that a face pattern *may* exist. The image patch that contains these two eye-analogue segments is rotated to the upright direction, i.e. the line connecting the two eye-analogue segments is parallel to the horizontal axis. Then it is re-scaled to a fixed size, and becomes a face candidate. The block diagram of the face candidates selector is shown in Figure 1.

3.2 Finding eye-analogue segments

Eyes are darker than other part of the face [28,29], so the eye-analogue segments are found by searching small image patches in the input image that are roughly as large as an eye and are darker than their neighborhoods. The proposed block diagram for finding eye-analogue segments is shown in Figure

2.

Let $P(x, y)$ be an intensity image of size $N_1 \times N_2$, where $x \in [1, N_1]$, $y \in [1, N_2]$, $P(x, y) \in [0, 1]$, in which x is the row index and y is the column index. Let $\text{avg}(P, x, y, h, w)$ be the average intensity of the image patch whose upper-left corner is (x, y) and whose size is $h \times w$, i.e.

$$\text{avg}(P, x, y, h, w) = \frac{\sum_{i=x}^{x+h-1} \sum_{j=y}^{y+w-1} P(i, j)}{h \times w} \tag{1}$$

Each pixel's intensity in the input image is compared to the average intensity of its eight neighborhood image patches (see Figure 3). Heuristically if a pixel is darker than most of its neighbors, it is marked as an eye-analogue pixel. Precisely, if eyes of size $h_e \times w_e$ is to be found, pixel (x, y) is preliminarily marked as an eye-analogue pixel if and only if six [6] or more of the following constraints are satisfied:

$$P(x, y) < 0.9 * \text{avg}(P, x - \lfloor h_e/2 \rfloor, y - \lfloor w_e/2 \rfloor, \lfloor h_e/2 \rfloor, \lfloor w_e/2 \rfloor) \tag{2}$$
$$P(x, y) < 0.9 * \text{avg}(P, x - \lfloor h_e/2 \rfloor, y, \lfloor h_e/2 \rfloor, 1) \tag{3}$$
$$P(x, y) < 0.9 * \text{avg}(P, x - \lfloor h_e/2 \rfloor, y + 1, \lfloor h_e/2 \rfloor, \lfloor w_e/2 \rfloor) \tag{4}$$
$$P(x, y) < 0.9 * \text{avg}(P, x, y - \lfloor w_e/2 \rfloor, 1, \lfloor w_e/2 \rfloor) \tag{5}$$
$$P(x, y) < 0.9 * \text{avg}(P, x, y + 1, 1, \lfloor w_e/2 \rfloor) \tag{6}$$
$$P(x, y) < 0.9 * \text{avg}(P, x + 1, y - \lfloor w_e/2 \rfloor, \lfloor h_e/2 \rfloor, \lfloor w_e/2 \rfloor) \tag{7}$$
$$P(x, y) < 0.9 * \text{avg}(P, x + 1, y, \lfloor h_e/2 \rfloor, 1) \tag{8}$$
$$P(x, y) < 0.9 * \text{avg}(P, x + 1, y + 1, \lfloor h_e/2 \rfloor, \lfloor w_e/2 \rfloor) \tag{9}$$

in which $\lfloor\ \rfloor$ is the greatest integer function (floor function).

We use P' to represent the image formed by eye-analogue pixels of P. $P'(x, y)$ is 1 if $P(x, y)$ is an eye-analogue pixel, and 0 if otherwise.

Now we use an example to illustrate the following face candidates selection

[6] Due to the possible different illumination directions, we should not require an eye-analogue pixel to be darker than its all eight neighborhoods. Six is a reasonable tradeoff.

process. Figure 4(a) is the original input image [7], eye-analogue pixels are marked in Figure 4(b) as white pixels. If a group of eye-analogue pixels forms an image patch that is large enough, e.g. its bounding rectangle's length is larger than $2h_e$ or the width is larger than $2w_e$, it is not likely to be an eye, so such connected pixels are removed. The eye-analogue pixels after this operation is shown in Figure 4(c).

Next, sparse pixels are removed. Nearly all pixels between the upper and lower eyelids in a face image are eye-analogue pixels (refer to Figure 4(a)). If in an eye-analogue pixel's neighborhood region there are too few eye-analogue pixels, it must not be a pixel in the eye region and thus should be removed. Precisely, an eye-analogue pixel $P'(x, y)$ is removed, i.e. not marked as an eye-analogue pixel, if

$$\text{avg}(P', x - \lfloor h_e/2 \rfloor, y - \lfloor w_e/2 \rfloor, h_e, w_e) < 0.2 \tag{10}$$

After these pixels have been removed, the eye-analogue pixels are shown in Figure 4(d). In equation (10), the threshold value, i.e. 0.2, is very small because this operation only intends to eliminate pixels in extremely sparse regions.

In the eye region, the upper and lower eyelids, pupil and sunken eye sockets are relatively darker than other parts of the eye. So there may be a few pixels near these parts that have not been marked as eye-analogue pixels after the aforementioned operations. So a pixel is marked as eye-analogue pixel if there is enough eye-analogue pixels in its neighborhood. Precisely, a pixel $P(x, y)$ is marked as eye-analog-pixel if

$$\text{avg}(P', x - \lfloor h_e/4 \rfloor, y - \lfloor w_e/4 \rfloor, \lfloor h_e/2 \rfloor, \lfloor w_e/2 \rfloor) > 0.35 \tag{11}$$

The result after this operation is shown in Figure 4(e). As in equation (10),

[7] This image comes from the BioID Face Database, http://www.bioid.com/research /index.htm [30].

in equation (11) the parameter 0.35 is not strict either.

The last step is to remove those image patches formed by connected eye-analogue pixels that are not likely to be eyes: blocks that are too large, too small, or blocks whose orientation are near vertical. If an image patch's bounding rectangle is of size $h \times w$, it will be removed if any of the followings equations is satisfied:

$$h/h_e < 0.5 \quad \text{or} \quad h/h_e > 2.0 \tag{12}$$
$$w/w_e < 0.5 \quad \text{or} \quad w/w_e > 2.0 \tag{13}$$
$$h/w > 0.8 \tag{14}$$

Equation (12) and (13) exclude extremely small or extremely large blocks. Note that the parameters 0.5 and 2.0 are quite relaxed because we do not want to lose any face pattern, and false candidates can be eliminated by the followed face verifier. The average value for h/w of human eyes are roughly 0.5. Equation (14) excludes blocks of large h/w values. Then every remaining image patch formed by connected eye-analogue pixels is an eye-analogue segment. The final result is shown in Figure 4(f). The eye-analogue segments are imposed on the original input image to make observation easier.

It is obvious that the operator $\text{avg}(P, x, y, h, w)$ is used intensively in the process illustrated in Figure 4. But since it is very easy to prove the following equations

$$\text{avg}(P, x, y, h, w) = \frac{\sum_{i=x}^{x+h-1} (\sum_{j=y}^{y+w-1} P(i, j)/w)}{h} \tag{15}$$
$$\text{avg}(P, x, y+1, 1, w) = \text{avg}(P, x, y, 1, w) + \frac{P(x, y+w) - P(x, y)}{w} \tag{16}$$
$$\text{avg}(P, x+1, y, h, 1) = \text{avg}(P, x, y, h, 1) + \frac{P(x+h, y) - P(x, y)}{h} \tag{17}$$

The summation over a rectangular area can be factored into a row summation followed by a column summation. Thus, given the input image P of size $N_1 \times N_2$, no matter how large the summation range h and w are, $\text{avg}(P, x, y, h, w)$

can be calculated for each pixel position (x, y) in $O(N_1 N_2)$ time.

3.3 Determining Face Candidates

In the proposed method, some in depth rotation of the face depth or rotation on-the-plane of the image are permitted as long as both eyes of the face are visible.

The block diagram of determining face candidates from eye-analogue segments is shown in Figure 5. An image patch in the input image is marked as a possible face if two eye-analogue segments's relative distance is like that of two eyes in a face. Let (x_i, y_i) and (x_j, y_j) be centroids of two eye-analogue segments. A face candidate may exist if all of the following constraints are satisfied:

$$d_{ij} < 2.5 w_e \tag{18}$$
$$d_{ij} > 1.5 w_e \tag{19}$$
$$|x_i - x_j| < h_e \tag{20}$$

in which d_{ij} is the distance between the two centroids,

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{21}$$

Parameters in equation (18) and (19) are chosen according to the fact that averagely $d_{ij} = 2w_e$ [31].

The rotation angle of this face candidate is determined by the line connecting the two centroids (see Figure 6):

$$\theta = \arctan \frac{x_i - x_j}{y_i - y_j} \tag{22}$$

Once the rotation angle is determined, the input image is rotated back to the upright direction. The sub-image containing these two eye-analogue segments is extracted from the original image and re-scaled to a normalized size. A

normalized image pattern is of size 30×30, and centroids of the two eye-analogue segments take the positions $(6, 8)$ and $(6, 22)$ respectively. Then a face mask is applied to remove some pixels near-boundary pixels in the normalized image patch. The mask is shown in Figure 7.

The masked image patch is histogram equalized to accommodate for illumination changes in the input image. Then the best fit linear plane is subtracted.

In order to detect faces in different scales, the input image is repeatedly scaled by a factor of 1.2 for 7 times. In each scale, all face candidates are marked and verified by the face verifier.

4 Experimental results

4.1 Data sets

The proposed method is tested on two face data sets. One is a subset of the FERET [32] data set and the other is the BioID face database [30].

We randomly selected 400 frontal view (**fa** or **fb**) face images from the FERET face data set. These images form the first data set. Images in this data set are eight-bit gray level images, 256 pixels in width and 384 pixels in height. The interocular distance [8] ranges from 40 to 60 pixels. The images primarily contain an individual's head, neck and shoulder [32]. There are nearly no complex background in these images.

The BioID face database is also a head-and-shoulder image face database. However, it stresses "real world" conditions. The BioID face database features a large variety of illumination and face size. Background of images in the face

[8] Interocular distance is distance between the eyes.

database is very complex. The images were recorded during several sessions at different places. The database consists of 1521 frontal view gray level images of 23 different test persons with a resolution of 384x286 pixel [33].

The BioID face database is believed to be more difficult than some commonly used head-and-shoulder face database without complex background, e.g. the extended M2VTS database (XM2VTS) [34]. In [30], when the same detection method and evaluation criteria were applied to both XM2VTS and BioID face databases, the successful detection rates are 98.4% and 91.8% respectively.

4.2 The face verifier

In order to get the successful face detection rate, the proposed method is followed by a very simple face verifier, which is a variation of the template matching method used in [2]. Eight face templates are defined (shown in Figure 8). For every face candidate, eight correlation values with the face templates are calculated. The largest correlation value is this face candidate's matching score and the face candidate with the largest matching score is selected as a face. The correlation value between a face candidate I_T and one of the template T is computed as [2]:

$$C_N(I_T) = \frac{< I_T T > - < I_T >< T >}{\sigma(I_T)\ \sigma(T)} \tag{23}$$

in which $<>$ is the average operator, $I_T T$ represents the pixel-by-pixel product and σ is the standard deviation over the area being matched.

The eight face templates are created using the soft clustering capacity of the *EM algorithm for mixture of probabilistic principal component analyzers* (mixture of PPCA) proposed by Tipping and Bishop [35]. Using this method, a mixture model consisting of eight probabilistic principal component analyzers are trained on a training face data set containing 200 face images. Then

the centers of each component become face templates. The mixture of PPCA clustering would result in more localized clusters, but with final reconstruction error inferior to that of a clustering model based explicitly on a reconstruction criterion. This property is preferred in our face detection task since localized clustering is more important than smaller reconstruction error. For details of EM algorithm for the PPCA mixture model, please refer to [35].

We only use the centers of the probabilistic principal component analyzers and discard all the second-order information. This choice is based on two considerations: firstly, the emphasis of our work is the face candidates selector and the face verifier is only used to test the successful detection rate, so a simple face verifier is enough to serve our aim; secondly, using the full mixture of PPCA model may be computationally intensive and thus time consuming. However, if a stronger face verifier is used, we may expect higher detection rates.

4.3 Evaluation criterion of the results

We used the criterion of [30,33] in our experiments. The criterion is a relative error measure based on the distances between the expected and the estimated eye positions. Let C_l and C_r be the manually extracted left and right eye positions of a face image, $\tilde{C}_l$ and $\tilde{C}_r$ be the estimated positions by the face detection method, d_l be the Euclidean distance between C_l and $\tilde{C}_l$, d_r be the Euclidean distance between C_r and $\tilde{C}_r$, and d_{lr} be the Euclidean distance between C_l and C_r. Then the *relative error* of this detection is defined as:

$$err = \frac{\max(d_l, d_r)}{d_{lr}} \tag{24}$$

If $err < 0.25$, the detection is considered to be correct. Notice that $err = 0.25$ means the bigger one of d_l and d_r roughly equals half an eye width.

Similarly, if one of the face candidates selected from an input image has a relative error less than 0.25, i.e. the face candidate is a face, the face candidates selector is considered to be successful in this input image.

4.4 Results and discussions

Two series of experiments are performed. The first series of experiments evaluate the ability of the face candidates selector, and the second series of experiments are designed to test the face detection rate of the proposed face candidates selector followed by the simple face verifier. The criteria for successful candidates selection and face detection are defined in section 4.3. Results on the two face data sets are listed in Table 1.

The face candidates selector is successful in most cases. It can select possible face patterns in different pose, scale, expression, and illumination conditions. Only two cues are used for finding eye-analogue segments, i.e. the eyes are darker than their neighborhood areas and the eye-analogue segments should be roughly as large as the eyes in a particular scale. These two cues are very robust. This robustness may lead to the robustness of the face candidates selector. Figure 9 shows some cases in which the face candidates selector fails. In Figure 9(a), one eye is near the image border so that it is not marked as eye-analogue segment; in Figure 9(b), glisten of the glasses misleads the face candidates selector; in Figure 9(c), the whole face in the input image is too dark to discriminate from the background; and in Figure 9(d), rotation angle of the face pattern is so large that it can not pass equation (20).

Some face detection results are shown in Figure 10 and Figure 11. Figure 10(a) to Figure 10(d) are examples of successful detections. An eye is depicted by a small white circle. Figure 10(a) to Figure 10(d) vary greatly in background, scale, expression and illumination. Figure 11(a) to Figure 11(d) are typical

examples of erroneous detections. Many false detections are around the correct face region, but the relative error *err* is slightly larger than 0.25 (Figure 11(a)). The two corner points of a mouth and the pair of eyebrows are sometimes mistaken for eyes (Figure 11(b) and Figure 11(c)). Figure 11(d) is a special case because among the 84 erroneous detections in the BioID face database, 54 images are of the same identity as Figure 11(d). It is possible that the simple face verifier does not fit well with face images of this identity.

In section 4.3, the threshold value for successful detection is set to 0.25. Distribution function of the relative error against successful detection rate is shown in Figure 12, (a) for the subset of FERET data set and (b) for the BioID face database. In Figure 12, successful detection rate are tracked when the threshold value on *relative error* varies from 0 to 0.4.

4.5 Comparing with other methods

Many face detection methods used the FERET face data set to evaluate their performances. The related data are listed in Table 2.

The FERET face data set is relatively easy since it contains nearly no background. All of the methods listed in Table 2 obtained satisfactory detection rates. The four methods' detection rates were computed using 400, 2000, 426, and 2340 images in the FERET face data set respectively.

The BioID was created later than FERET and less results are available on this face database. On this database, the proposed method is compared with the method proposed by Jesorsky et al. [30]. The proposed method successfully detected 94.5% faces (1437 out of 1521) while method in [30] is successful in 91.8% cases.

5 Conclusion

In this paper an efficient face candidates selector for detecting faces in still gray level images is proposed. Only two cues are used for finding eye-analogue segments, i.e. the eyes are darker than their neighborhood areas and the eye-analogue segments should be roughly as large as the eyes in a particular scale. Eye-analogue segments are then combined to yield face candidates. Since the two cues are robust, the proposed method can deal with illumination changes and moderate rotations. In a subset of the FERET face data set containing 400 face images and the BioID face database containing 1521 face images, the face candidates selector is successful in 98.75% and 98.6% cases respectively.

Experimental results show that the proposed face candidates selector may fail in the following conditions:

- One of the eyes of a face is too close to border of the input image,
- Glisten of glasses or occlusion of eyes,
- The image is too dark to discriminate between eye and other part of a face, and
- The face's rotation angle is too large.

Further improvements can be made on the face candidates selector, however, in order to get higher detection rates or detect faces in images that contain multiple faces, a powerful face verifier is required.

Acknowledgements

The comments and suggestions from the anonymous reviewers greatly improved this paper. The National Natural Science Foundation of China, the Natural Science Foundation of Jiangsu Province, China, and the 10th Five

Years Plan of Jiangsu Province, China, supported this research.

References

[1] R. Chellappa, C.L. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, Proceedings of the IEEE, 83(5), 1995: 705-741.

[2] R. Brunelli, T. Poggio, Face Recognition: Features versus Templates, IEEE Transcations on Pattern Analysis and Machine Intelligence, 15(10), 1993: 1042-1052.

[3] R. Kjeldsen, J. Kender, Finding skin in color images, In Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996: 312-317.

[4] M.-H Yang, D. Kriegman, N, Ahuja, Detecting Faces in Images: A survey, to appear in IEEE Transcations on Pattern Analysis and Machine Intelligence, 2001.

[5] R. Chellappa, W. Zhao, Face Recognition: A Literature Survey, to be submitted to ACM Journal of Computing Surveys, 2000.

[6] M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neurosicence, 3(1), 1991: 71-86.

[7] B. Moghaddam, A. Pentland, Probabilistic Visual Learning for Object Representation, IEEE Transcations on Pattern Analysis and Machine Intelligence, 19(7), 1997: 696-710.

[8] M.-H Yang, N. Ahuja, D. Kriegman, Mixtures of linear subspaces for face detection, In Proceedings of the fourht International Conference on Automatic Face and Gesture Recognition, pp 70-76, 2000.

[9] Z. Ghahramani, G.E. Hinton, The EM algorithm for mixture of factor analyzers, Technical Report CRG-TR-96-1, Department of Computer Science, University of Toronto, 1996.

[10] K.K. Sung, T. Poggio, Example-based learning for View-based Human Face Detection, Technical Report AIM-1521, MIT AI Lab, 1994.

[11] H. Rowley, S. Baluja, T. Kanade, Neural Network-based face detection, IEEE Transcations on Pattern Analysis and Machine Intelligence, 20(1), 1998: 23-38.

[12] H. Rowley, S. Baluja, T. Kanade, Rotation Invariant Neural Network-based Face Detection, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 38-44, 1998.

[13] R. Feraud, PCA, Neural Network and Estimation for Face Detection, In H. Wechsler, P.J. Phillips, V. Bruce et al eds, Face Recognition: From Theory to Applications, NATO ASI Series F, volumn 163, Springer-Verlag, 1998, pp. 424-432.

[14] E. Osuna, R. Freund, F. Girosi, Training Support Vector Machines: An Application to Face Detection, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1997, pp. 130-136.

[15] I. Craw, H. Ellis, J. Lishman, Automatic extraction of face features, Pattern Recognition Letters, 5, 1987: 183-187.

[16] I. Craw, D. Tock, A. Bennett, Finding face features, In Proceedings of teh second European Conference on Computer Vision, 1992: 92-96.

[17] A.L. Yuille, P.W. Hallinan, D.S. Cohen, Feature extraction from faces using deformable templates, International Journal of Computer Vision, 8(2), 1992: 99-111.

[18] A.K. Jain, Y. Zhong, M.-P. Dubuisson-Jolly, Deformable template models: A review, Signal Processing, 71(2), 1998: 109-129.

[19] M.C. Burl, M. Weber, P. Perona, A Probabilistic Approach to Object Recognition Using Local Photometry and Global Geometry, In Proceedings of the European Conference on Computer Vision, 1998, pp. 628-641.

[20] M.C. Burl, P. Perona, Recognition of planar object classes, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, California, Jun. 1996, pp. 223-230.

[21] B. Takaćs, H. Wechsler, A Saccadic Vision SYstem for Landmark Detection and Face Recognition, In H. Wechsler, P.J. Phillips, V. Bruce et al eds, Face Recognition: From Theory to Applications, NATO ASI Series F, volumn 163, Springer-Verlag, 1998, pp. 627-636.

[22] G. Yang, T.S. Huang, Human face detection in a complex background, Pattern Recognition, 27(1), 1994: 53-63.

[23] C.-C Han, H.-Y. M. Liao, K.-C. Yu, L.-H. Chen, Fast face detection via morphology-based pre-processing, In Proceedings of the Ninth International Conference on Image Analysis and Processing, 1998, pp. 469-476.

[24] Y. Amit, D. Geman, B. Jedynak, Efficient Focusing and Face Detection, In H. Wechsler, P.J. Phillips, V. Bruce et al eds, Face Recognition: From Theory to Applications, NATO ASI Series F, volumn 163, Springer-Verlag, 1998, pp. 157-173.

[25] B. Takaćs, H. Wechsler, Face Locaion Using a Dynamic Model of Retinal Feature Extraction, In Proceedings of the International Workshop on Automatic Face and Gesture Recogntion, Zurich, Switzerland, June 1995, pp. 243-247.

[26] J. Huang, H. Wechsler, Visual routines for eye location using learning and evolution, IEEE Transctions on Evolutionary Computation, 4(1), 2000: 73-82.

[27] K.M. Lam, Y.L. Li, An Efficient Approach for Facial Feature Detection, In Proceedings of the IEEE International Conference on Signal Processing, Beijing, China, 1998, pp. 1100-1103.

[28] K. Sobottka, I. Pitas, Face localization and facial feature extraction based on shape and color information, In Proceedings of the IEEE International Conference on Image Processing, Lausanne, Switzerland, vol. III, pp. 483-486, 1996.

[29] G.C. Feng, P.C. Yuen, Multi cues eye detection on gray intensity images, Pattern Recognition, 34(5), 2001: 1033-1046.

[30] O. Jesorsky, K. Kirchberg, R. Frischholz, Robust Face Detection Using the

Hausdorff Distance, In J. Bigun and F. Smeraldi, editors, Audio and Video based Person Authentication - AVBPA 2001, Springer, 2001, pp. 90-95.

[31] A.M. Alattar, S.A. Rajala, Facial features localization in front view head and shoulders images, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 6, 1999, pp. 3557-3560.

[32] P.J. Phillips, H. Wechsler, J. Huang, P.J. Rauss, The FERET database and evaluation procedure for face-recognition algorithms, Image and Vision Computing, 16(5), 1998: 295-306.

[33] http://www.bioid.com/research/index.htm, The BioID Face Database, BioID-Technology Research, June 2001.

[34] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In Proceedings of the Second International Conference on Audio and Video-based Biometric Person Authentication, March 1999, pp. 72-77.

[35] M.E. Tipping, C.M. Bishop, Mixtures of Probabilistic Principal Component Analysers, Neural Computation, 1999 11(2): 443-482.

[36] J. Huang, S. Gutta, H. Wechsler, Detection of Human Faces Using Decision Trees. In Proceedings of the Second International Conference on Automatic Face and Hand Gesture Recognition, 1996, pp. 248-252.

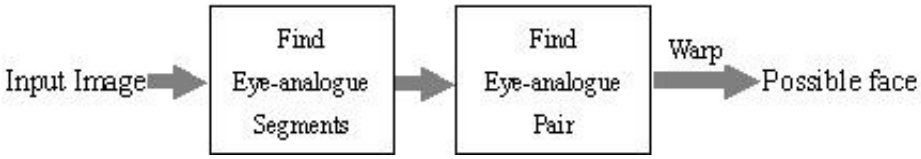


Fig. 1. Block diagram of the face candidates selector

Fig. 2. Block diagram for finding eye-analogue segments

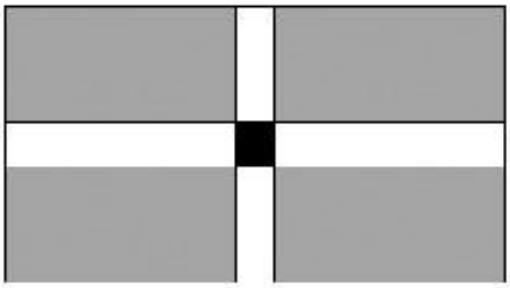
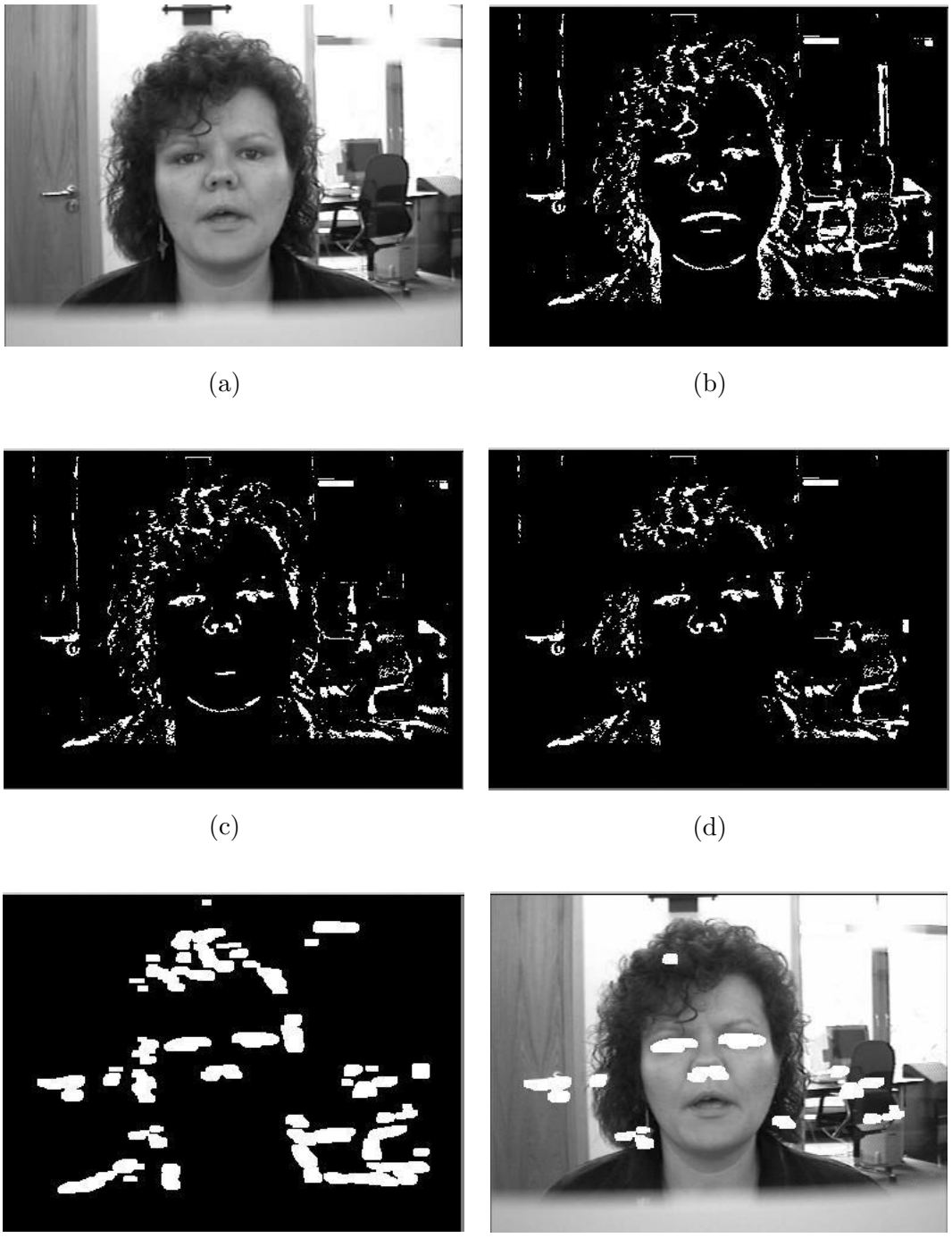


Fig. 3. An pixel (the black box) and its eight neighborhood image patches.



(a) (b)

(c) (d)

(e) (f)

Fig. 4. An example of the process of finding eye-analogue segments.

Fig. 5. Block diagram of determining face candidates from eye-analogue segments.

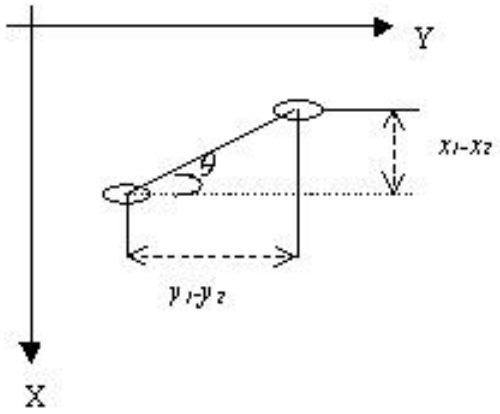


Fig. 6. The rotation angle of a face candidate.

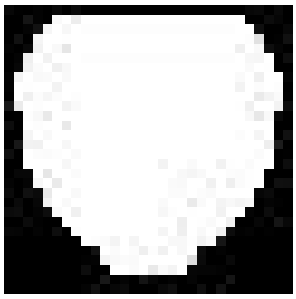


Fig. 7. The face mask of size 30×30.

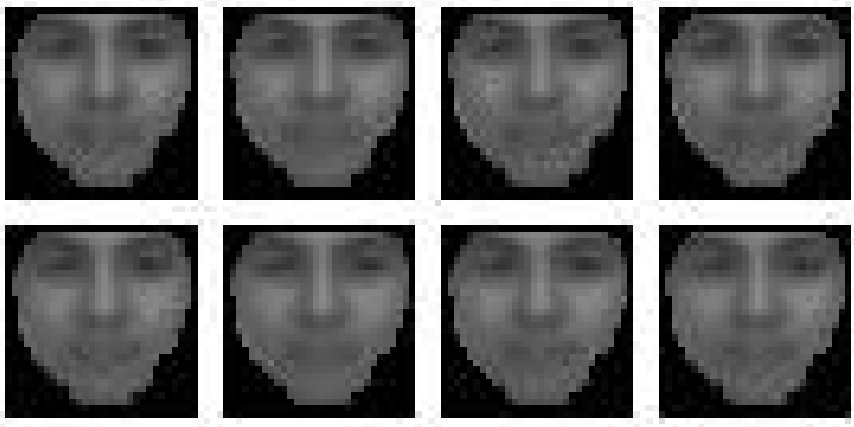
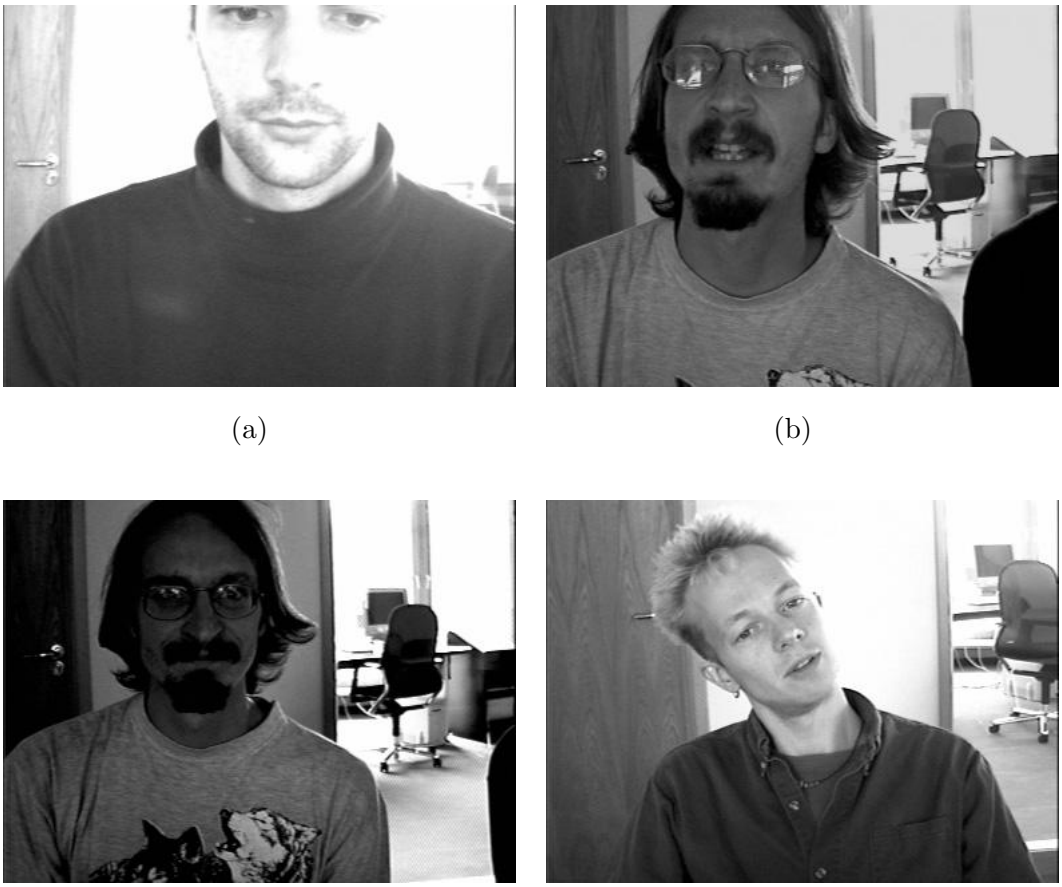


Fig. 8. The eight face templates.



(a) (b)

(c) (d)

Fig. 9. Some examples in which the face candidates fails.

(a) (b)

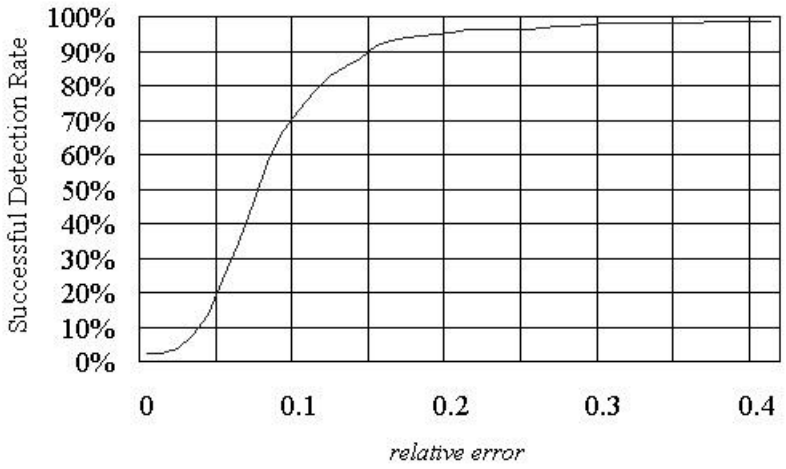
(c) (d)

Fig. 10. Some successful detection results.

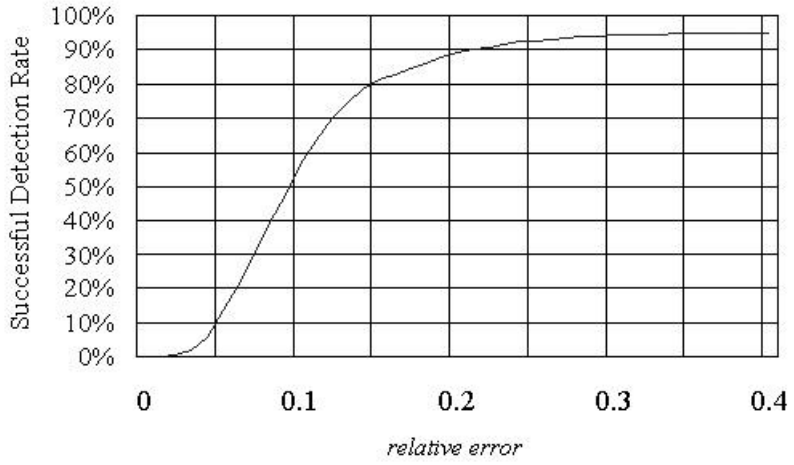
(a) (b)

(c) (d)

Fig. 11. Some erroneous detection results.



(a) FERET face data set



(b) BioID face database

Fig. 12. Distribution function of the relative error against successful detection rates.

Table 1

Successful detection rates

Test face data set	Face candidates selector rate	Face detection rate
subset of FERET data set	98.75%(395/400)	97.25%(389/400)
The BioID face database	98.6%(1499/1521)	94.5%(1437/1521)

Table 2

Successful detection rates on the FERET face data set

Detection method	Face detection rate
Proposed method	97.25%
Moghaddam[7]	97.0%
Takács[25]	95.3%
Huang[36]	96.0%

About the author

Jianxin Wu received his B.S degree in Computer Science from Nanjing University, China, in 1999. He is currently a graduate student in the Department of Computer Science & Technology at Nanjing University. His research interests include machine learning, pattern recognition, and face recognition.

Zhi-Hua Zhou received the BSc, MSc and PhD degrees from Nanjing University, China, in 1996, 1998 and 2000 respectively, all with the highest honor. He joined the Computer Science & Technology Department of Nanjing University as a lecturer in 2001, and was promoted to associate professor in 2002. He has won the Mircrosoft Fellowship Award in 1999. His current interests are in machine learning, neural computing, evolutionary computing, pattern recognition, and data mining. He has chaired the organizing committee of the 7th Chinese Workshop on Machine Learning, and acted as program committee members for several international conferences. He is on the editorial board of the journal Artificial Intelligence in Medicine, is an executive committee member of Chinese Association of Artificial Intelligence (CAAI) and CAAI Machine Learning Council, and is a member of IEEE and IEEE Computer Society.