

Fast Face Detection via Morphology-Based Pre-processing¹

Chin-Chuan Han[†], Hong-Yuan Mark Liao[†]Gwo-Jong Yu[‡], and

Liang-Hua Chen[§]

[†]Institute of Information Science,

Academia Sinica, Nankang, Taipei, Taiwan

[‡]Institute of Computer Science and Information Engineering,

National Central University, Chung-Li, Taiwan

[§]Department of Computer Science and Information Engineering,

Fu Jen University, Taiwan

Abstract. An efficient face detection algorithm which can detect multiple faces in cluttered environment is proposed. First of all, morphological operations and labeling process were performed to obtain the eye-analogue segments. Based on some matching rules and the geometrical relationship on a face, eye-analogue segments were grouped into pairs and used to locate potential face regions. Finally, the potential face regions were verified via a trained neural network and the true faces were determined by optimizing a distance function. Since the morphology-based eye-analogue segmentation process can efficiently locate the potential eye-analogue regions, the subsequent processing only has to deal with 5-10% area of the original image. Experiments demonstrate that an approximately 94% success rate is reached and the relative false detection rate is very low.

Key Words: Face Detection, Morphological Opening/Closing Operation

1 Introduction

Human face detection and recognition have long being a difficult research topic. In the last two decades, researchers have devoted much effort to these two problems and have received some satisfactory results. Part of these previous efforts were focused on face recognition[1–3]. However, an accurate and efficient method for human face detection is still deficient. Govindaraju *et al.*[4] presented a system which could locate human faces in photographs of newspapers, but the approximate size and expected number of faces must be known in advance. Sirohey[5] utilized an elliptical structure to segment the human heads from cluttered images. Yang and Huang[6] utilized a three-level hierarchical knowledge-based method to locate the human faces in complex backgrounds. Sung and Poggio[7] applied two distance metrics to measure the distance between the input image and the cluster center. A feature vector consisting of 12 values was inputted into a multi-layer perceptron network for the verification task. Moghaddam and Pentland[8] designed a distance function called *distance from feature space*(dffs)

¹ This work is supported by National Science Council of Taiwan under grant no. NSC86-2213-E-001-023.

as a metric to evaluate the difference between the input face image and the reconstructed face image for the face detection.

Face detection can also be achieved by detecting geometrical relationships among facial components such as nose, eyes, and mouth. Juell and Marsh[9] proposed a hierarchical neural network to detect the multiple human faces in the scene. Leung *et al.*[10] coupled a set of local feature detectors via a statistical model to find the facial components for face locating. Their approach was invariant with respect to translation, rotation, and scale. Besides, they could also handle partial occlusion of faces.

Most of the above mentioned systems limit themselves to deal with human faces in frontal view. Namely, the orientation problem which is a potential trouble in this type of problem was not seriously considered. Besides, Rowley[11] adopted a small window(20×20) to slide over all portions of an image at various scales. This brute force search is no doubt a time-consuming procedure. Jeng *et al.*[12] proposed a geometrical face model to solve the above mentioned problem. The average execution time for locating a face using a Sparc-20 machine was less than 5 seconds. However, the major drawbacks of their approach are: (1) the size of a face must be larger than 80×80 and, (2) the accuracy of using a geometrical face model may be significantly influenced due to the change of lighting conditions.

In this paper, an efficient face detection system is proposed. The proposed system consists of three main steps including *eye-analogue segmentation*, *generation of potential face regions*, and *face verification* modules. The extraction of eye-analogue region using morphological operations is described in Section 2. In Section 3, eye-pairs and normalized images are generated based on some geometric rules. A neural network-based verifier and a cost function are presented to verify the face region in Section 4. Some experimental results to show the validity of our detection system are demonstrated in Section 5. Finally, concluding remarks are given in Section 6.

2 Eye-Analogue Segmentation

As mentioned in [12,13], eyebrows, eyes, nostrils and mouth always look darker than the rest part of a face. Among these facial features, eyebrows, nostrils, and mouth frequently vary much according to the lighting, the occludence, and the facial expressions. In comparison with the above mentioned unstable facial features, the eyes can be considered a salient and relatively stable feature on a face. Therefore, the eye-analogue pixels in a cluttered image are segmented first by the morphological operations. Then, the small segments of eye-analogue regions are grouped together by conditional morphological dilation operations and labeled by a traditional labeling process.

In 1993, Chow and Li[13] employed a morphological *opening residue* operation to extract the intensity valleys as potential eye-analogue pixels via a 5×5 circle structuring element. In our approach, we apply the morphological closing (\bullet) and clipped different (\sqcup) operations to find the candidate pixels of eye-analogue. Let X be the original image and $X_{1/2}$ be the half image with scale 0.5. A horizontal

structuring element S_h with size 1×7 and a vertical structuring element S_v with size 7×1 are, respectively, operated on X and $X_{1/2}$. The set of operations which are able to identify the eye-analogue pixels are expressed as follows:

$$\begin{aligned} E_1 &= T_1 (X \cup (X \bullet S_h)), & E_2 &= T_2 (X_{1/2} \cup (X_{1/2} \bullet S_h)), \\ E_3 &= T_3 (X \cup (X \bullet S_v)), & E_4 &= T_4 (X_{1/2} \cup (X_{1/2} \bullet S_v)), \\ E &= E_1 \vee E_2^2 \vee E_3 \vee E_4^2, \end{aligned} \quad (1)$$

where the superscript, 2, of E_2 and E_4 are used to enlarge them into twice of the original size. T_1, T_2, T_3 , and T_4 are four threshold functions whose values are the average values of the images E_1, E_2, E_3 and E_4 , respectively.

In practice, when the scene is complex, some pixels of this kind may not be the eye-analogue pixels. Therefore, we apply the conditional morphological dilation operation at this stage to remove these unwanted pixels in the background. Then, a conventional labeling process is performed to locate the eye-analogue segments. The eye-analogue detection algorithm is described in detail as follows.

Eye-analogue Detection Algorithm:

Step 1 : Perform labeling process on image E and compute a set of geometrical data from each segment including the lengths of the major and minor axes, the orientation, the center point, and the minimal bounding rectangle.

Step 2 : If the length of the major axis of segment i is larger than $0.6N$ (where N is the smallest width of a face region), terminate the conditional dilation operation for segment i and eliminate segment i from image E . Otherwise, go to next step.

Step 3 : Perform conditional dilation operation on every segment i by the structuring element $SE = \{1_{(x,y)} | (x,y) \in Z\}$, where

1. if the orientation of segment i is located at $[-\frac{\pi}{8}$ to $\frac{\pi}{8}]$, set $SE = \{1_{(-1,0)}, 1_{(0,0)}, 1_{(1,0)}\}$,
2. if the orientation of segment i is located at $(-\frac{\pi}{8}$ to $-\frac{3\pi}{8}]$, choose the element $\{1_{(1,-1)}, 1_{(0,0)}, 1_{(-1,1)}, 1_{(-1,0)}, 1_{(1,0)}\}$ as the structuring element SE ,
3. if the orientation of segment i is located at $(\frac{\pi}{8}$ to $\frac{3\pi}{8}]$, the structuring element SE is defined as $\{1_{(-1,-1)}, 1_{(0,0)}, 1_{(1,1)}, 1_{(-1,0)}, 1_{(1,0)}\}$,
4. if the orientation of segment i is located at $(-\frac{3\pi}{8}$ to $-\frac{4\pi}{8}]$ or $(\frac{3\pi}{8}$ to $\frac{4\pi}{8}]$, set $SE = \{1_{(0,-1)}, 1_{(0,0)}, 1_{(0,1)}\}$.

Step 4 : Repeat steps 1 to 3 for $N/5$ times.

An example demonstrating the eye-analogue segmentation process is shown in Fig. 1. The main advantage of the two above mentioned processes is to combine those non-eye analogue segments whose sizes are small. These processes will reduce the number of potential eye-analogue segments.

3 Generation of Potential Face Regions

After performing the segmentation process, each eye-analogue segment can be considered a candidate of one of the eyes. In this section, we propose four matching rules to guide the merging of potential eye-analogue segments into pairs. In

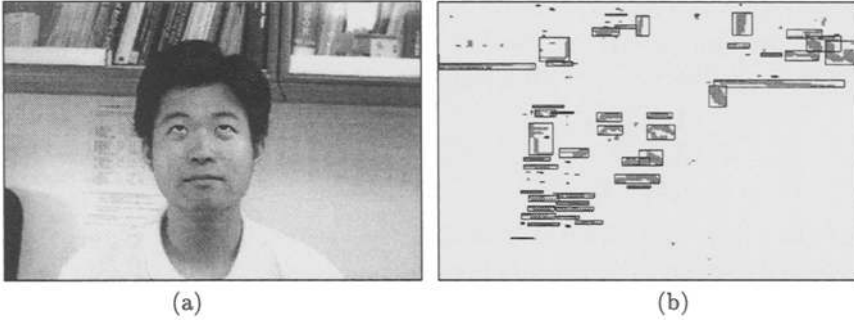


Fig. 1. The eye-analogue segmentation process. (a) The original image, (b) the eye-analogue segments after the closing, clipped different, conditional dilation, and labeling operations.

the image E , there are at most M^2 potential combinations of face regions generated via the M eye-analogue segments. In order to reduce the execution time, the geometrical constraint between a real eye pair is introduced to screen some impossible pairs as follows:

- (a) *The length ratio between the major and the minor axes of segment i must be smaller than 10.*
- (b) *The distance between the center points of two eye-analogue segments must be larger than $0.6N$. Here, the value 0.6 is a rough estimate of the ratio between the distance between the two pupils of eyes and the width of a face, and N is the shortest width of a face.*
- (c) *Each eye must be located by extending a small range from the other eye as shown in Fig. 2(a).*
- (d) *The area of segment i should be larger than 10 pixels.*

Once the potential eye pairs are determined, their corresponding face regions can be easily extracted and each of these faces is normalized into the 20×20 standard size. As shown in Fig. 2(b), (x_i, y_i) and (x_j, y_j) are two center points of segment i and j , respectively. (x_1, y_1) , (x_2, y_2) , (x_3, y_3) , and (x_4, y_4) are four corner points of a normalized face region. Let $x_i + x_j = A$, $x_i - x_j = B$, and $y_i - y_j = C$. The the coordinates of the four corner points can be calculated as follows.

$$\begin{aligned}
 x_1 &= \frac{1}{2}A - \frac{c_2}{c_1}B + \frac{c_3}{c_1}C, & y_1 &= \frac{1}{2}A - \frac{c_2}{c_1}C - \frac{c_3}{c_1}B, \\
 x_2 &= \frac{1}{2}A + \frac{c_2}{c_1}B + \frac{c_3}{c_1}C, & y_2 &= \frac{1}{2}A + \frac{c_2}{c_1}C - \frac{c_3}{c_1}B, \\
 x_3 &= \frac{1}{2}A - \frac{c_2}{c_1}B + \frac{c_4}{c_1}C, & y_3 &= \frac{1}{2}A - \frac{c_2}{c_1}C - \frac{c_4}{c_1}B, \\
 x_4 &= \frac{1}{2}A + \frac{c_2}{c_1}B + \frac{c_4}{c_1}C, & y_4 &= \frac{1}{2}A + \frac{c_2}{c_1}C - \frac{c_4}{c_1}B,
 \end{aligned} \tag{2}$$

where values c_1 , c_2 , c_3 , and c_4 are, respectively, 12.5, 10, 4, and 16 in a normalized 20×20 face image (see Fig. 2(b)) due to the training samples used in the experiments. The potential eye pairs which satisfy the matching rules are linked via solid line segments as shown in Fig. 2(c). According to Eq. (2), the potential face regions which cover the potential eye pairs in Fig. 2(c) are extracted and normalized to 20×20 standard size as shown in Fig. 2(d).

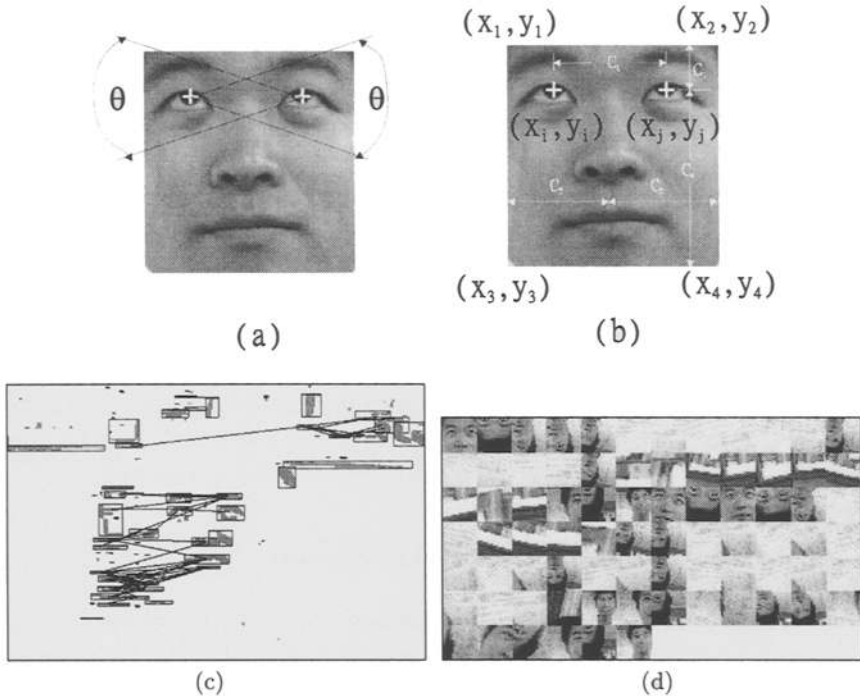


Fig. 2. The extraction of face region. (a) The matching rule (c), (b) the geometrical relationship of face region, (c) the potential eye pairs of Fig. 1(c), (d) the potential face regions.

4 Face Verification

In the previous section, all potential faces are normalized to a fixed size and rotated into frontal position. In this section, we propose a coarse-to-fine face verification process to locate the real positions of faces in an image. In coarse verification, a trained backpropagation neural network[11] is applied to decide whether a potential region contains a face. In the fine verification process, we apply a cost function to perform final selection. In what follows, the detailed procedure, from coarse to fine, will be described.

Before the neural network is applied to execute coarse verification, a pre-processing step consisting of masking, illumination gradient correction, and histogram equalization has to be performed first. A set of 11200 face images generated from 700 face samples are collected as the positive samples by randomly, slightly rotating (up to 10°), scaling (90% to 110%), translating (up to half a pixel), and mirroring. In addition, 5000 non-face images are collected from 20 scenery images as negative samples. The trained backpropagation net is then applied to filter out those non-face regions. If the neural network outputs a positive response, the potential region may contain a face. Otherwise, this region will be eliminated from the potential candidate face list. This coarse verifica-

tion process using the trained neural networks filters out significant number of non-face regions.

In the fine verification process, an evaluation function ϵ is designed to eliminate the above mentioned overlapping detections as well as those previously retained non-face regions. The evaluation function ϵ is defined as follows:

$$\epsilon^2 = \sum_{i=M+1}^N y_i^2 = \|x - \bar{x}\|^2 - \sum_{i=1}^M y_i^2, \quad (3)$$

where N is the size of an image x which is to be checked, M is the number of principal components used to reconstruct the original images. Since the value ϵ denotes the difference between the input image and the reconstructed image via PCA, one has to choose the region with the positive response and the local smallest value as a face region. Once a face region is confirmed, the last step is to eliminate those regions that overlap with the chosen face region.

5 Experimental Results and Discussion

A set of experimental results are demonstrated to show the effectiveness and efficiency of the proposed system. In the test images, all human faces were oriented into various directions and positioned arbitrarily in cluttered background. In this research, the minimum size of a face which could be detected was 50×50 .

Figs. 3(a)-(f) show 10 test images which have correct detection results. The bounding rectangle that bounds a face region is used to justify whether the detection is correct or not. One thing worth noticing is that the test image shown in Fig. 3(f) contained two faces with a nearly 180 degree difference, it is obvious that our system worked perfectly in dealing with this kind of problem. For an overall evaluation, 122 faces were detected successfully out of the total of 130 faces. Therefore, the success rate for detecting face was roughly about 94%. On the other hand, the proposed system also detected totally 25 fake faces from the cluttered backgrounds of the test images. Fig. 4 shows some unsuccessful detections including fake faces(the left bounding rectangle of Fig. 4(a)) and missed faces(the right-hand side face of Fig. 4(b)).

As to the execution time problem, the time required to locate the precise locations of the faces in the test image set is dependent upon the size and complexity of images. For example, a 512×340 image shown in Fig. 3(e) needed less than 18 seconds to locate the correct face position using a Sparc 20 workstation. As to the case shown in Fig. 3(f), the execution time under the same environment was about 23 seconds.

6 Concluding Remarks

In this paper, we have proposed an efficient face detection algorithm to find multiple faces in cluttered images. The proposed face detection technique may

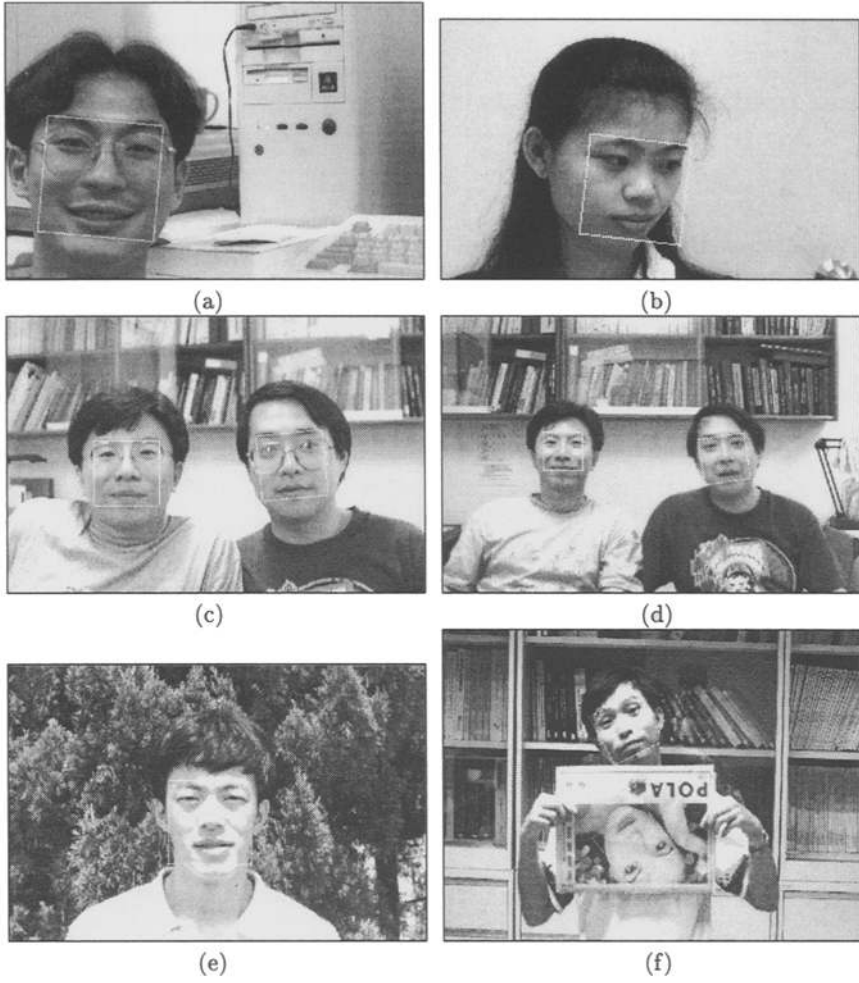


Fig. 3. Testing examples.



Fig. 4. The mis-detected examples.

locate multiple faces oriented in any directions. Besides, the morphology-based eye-analogue segmentation process is able to reduce the background part of a cluttered image up to 95%. This process significantly speeds up the subsequent face detection procedure because only 5-10% regions of the original image are left for further processing. Therefore, the execution time is much less than most of the existing systems.

References

1. S. Lawrence, C. Lee Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach", *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 98-113, January 1997.
2. R. Brunelli and T. Poggio, "Face recognition: Features versus templates", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, 1993.
3. M. Turk and A. Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.
4. V. Govindaraju, S. N. Srihari, and D. B. Sher, "A computational model for face location", in *Proc. Computer Vision and Pattern Recognition*, 1990, pp. 718-721.
5. S. A. Sirohey, "Human face segmentation and identification", Master's thesis, University of Maryland, 1993.
6. G. Yang and T. S. Huang, "Human face detection in a complex background", *Pattern Recognition*, vol. 27, no. 1, pp. 53-63, 1994.
7. K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection", in *Proc. Image Understanding Workshop*, Monterey, Calif., Nov. 1994, pp. 843-850.
8. B. Moghaddam and A. Pentland, "A subspace method for maximum likelihood target detection", in *Proc. IEEE International Conference on Image Processing*, Washington DC, Oct. 1995.
9. P. Juell and R. Marsh, "A hierarchical neural network for human face detection", *Pattern Recognition*, vol. 29, no. 5, pp. 781-787, 1996.
10. T. K. Leung, M. C. Burl, and P. Perona, "Finding faces in clustered scenes using random labeled graph matching", in *Proc. Computer Vision and Pattern Recognition*, Cambridge, Mass., Jun. 1995, pp. 637-644.
11. H. A. Rowley, S. Baluja, and T. Kanade, "Human face detection in visual scenes", Tech. Rep., Carnegie Mellon University, 1995.
12. S. H. Jeng, H. Y. Mark Liao, Y. T. Liu, and M. Y. Chern, "An efficient approach for facial feature detection using geometrical face model", in *Proc. 13th International Conference on Pattern Recognition*, Vienna, Austria, Aug. 1996, pp. 426-430.
13. G. Chow and X. Li, "Towards a system for automatic facial feature detection", *Pattern Recognition*, vol. 26, no. 12, pp. 1739-1755, 1993.