A Robust Two Stage Approach for Eye Detection

Jing-Wein Wang and Chin-Chun Kuo

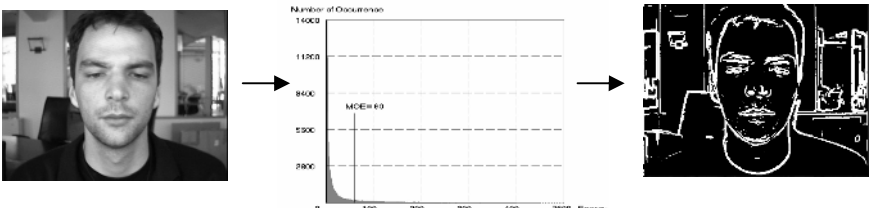
Institute of Photonics and Communications,
National Kaohsiung University of Applied Sciences,
415 Chien-Kung Road, Kaohsiung, Taiwan 807, R.O.C.
Tel. +886-7-3814526 Ext. 3350
Fax. +886-7-3832771
jwwang@cc.kuas.edu.tw

Abstract. This paper adopts face localization to eye extraction strategy for eye detection in complex scenes. First, an energy analysis is applied to enhance face localization performance by removing most noise-like regions rapidly. According to anthropometry, the face-of-interest (FOI) region is located using signatures derived from the proposed head contour detection (HCD) approach that searches the best combinations of facial sides and head contours. Second, via the de-edging preprocessing for facial sides, a wavelet subband inter-orientation projection method is devised to generate and select eye-like candidates. By utilizing the geometric discrimination information among the facial components, such as the eyes, nose, and mouth, the proposed eye verification rules verify the eye pair selected from the candidates. The experimental results demonstrate the significance performance improvement using the proposed method over others on three head-and-shoulder databases.

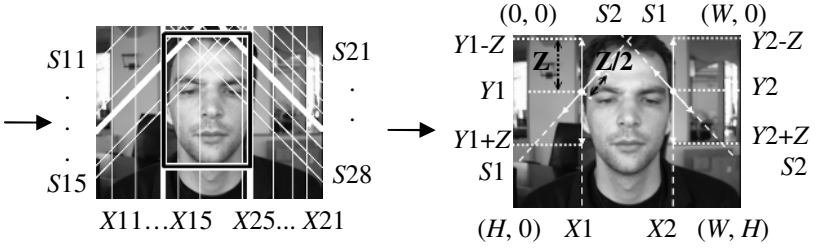
1 Introduction

Biometric technology such as eye detection in an image is a challenging problem because it involves locating eye with no prior knowledge about image content [1]. In this work, we propose the use of complementary techniques which are based on head contour geometry characterization and wavelet subband inter-orientation projection. The technique aims at providing an efficient system to operate with complex backgrounds and must tolerate illumination changes, scale variations, and small head rotations, say 30°. The presented eye detection framework composed of the face localization and eye extraction stages as shown in Fig. 1. This paper is organized as follows. The next section is dedicated to HCD approach being as a justified way of locating FOI region. Using FOI size estimate, we define a wavelet subband inter-orientation projection method for generating and selecting the eye-like candidates in Section 3. Eye extraction is achieved by examining the correlation between eyes and detecting geometrical relationships among the facial components such as the eyes, nose, and mouth. Finally, Section 4 gives the experimental results on three head-and-shoulder image databases and Section 5 concludes the paper.

F. Roli and S. Vitulano (Eds.): ICIAP 2005, LNCS 3617, pp. 431 – 437, 2005.

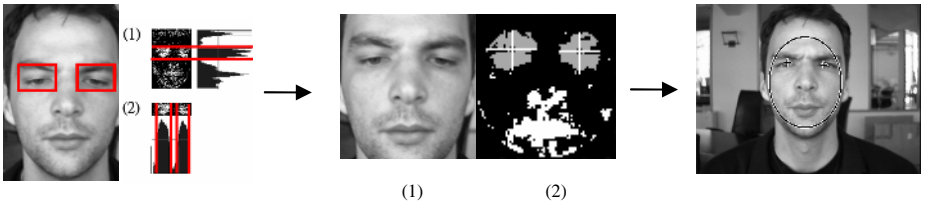


(a) Input image, (b) Histogram and MOE value of the energy image, (c) Result of energy thresholding,



(d) Head contour detection, (e) Edge confirmation criterion,

(f) Located FOI region, (g) One-level wavelet decomposition, (h) LH subband binary thresholding along with HL subband de-edging for facial sides,



(i) Candidate eye-pair generation and selection by using wavelet subband inter-orientation projection, (j) Both correlation coefficient and facial geometry relationship used for eye verification, (k) Output image.

Fig. 1. Block diagram of eye detection steps

2 Face Localization

Prior to face object extraction, a 3 × 3 smoothing filter is used to move the center from pixel to pixel in an image to guarantee the desired local edge enhancement. This

continues until all pixel locations have been covered and a new energy image is to be created for storing the response of the linear mask. Fig. 1(a) shows the example image of size 384 × 286 pixels with 256 gray levels, and Fig. 1(b) shows its energy histogram. To extract the interested object from the background, the threshold is set as mean of energy (MOE) initially and gradually changed until the observed pixel-of-thresholding (point) density of the whole image is around 10% ~ 20% which is considerably good representation for face to be detected as shown in Fig. 1(c) denoted as $B(x, y)$. Human head contour, which contains relatively concentrated information in face image, is more reliable than the feature of eyes or other face organs especially detecting unclear images or small face images. Locating eye with feature from head contour has the advantage of scale invariance and simplification. The algorithm consists of the following details as shown in Fig. 1(d). Below, we show the details of the proposed HCD algorithm, how effectively locates the head-face profile. Beginning, if

$$Nr(B(x,y)) = \begin{cases} 1, & \text{if } B(x,y) = 255 \\ 0, & \text{if } B(x,y) = 0 \end{cases}, \tag{1}$$

and $0 \le k \le H$, for each left diagonal projection, D_{lk} is computed by

$$D_{lk} = \sum_{(x,y)=(0,k)}^{(k,0)} Nr(B(x,y)), \tag{2}$$

and each left vertical projection, V_{lk} is computed by

$$V_{lk} = \sum_{(x,y)=(0,0)}^{(k,H)} Nr(B(x,y)). \tag{3}$$

Next, for each right diagonal projection, D_{rk} is computed by

$$D_{rk} = \sum_{(x,y)=(W,k)}^{(W-k,0)} Nr(B(x,y)), \tag{4}$$

and each right vertical projection, V_{rk} is computed by

$$V_{rk} = \sum_{(x,y)=(W,0)}^{(W-k,H)} Nr(B(x,y)). \tag{5}$$

According to the point projection waveforms from equations (2)-(5), the diagonal head boundary candidates and horizontal facial side candidates are determined from the peak responses with wide enough spreading which is defined as projection relief slope greater than 1.0, respectively. A pseudo line drawn from the valley to the peak of response measures the slope. Afterwards, the lines marked by white solid lines denote all candidate locations of head-face as presented in Fig. 1(d), i.e.

$$S1 \in \{S11, S12, S13, S14, S15\}, \tag{6}$$

$$S2 \in \{S21, S22, S23, ..., S28\}, \tag{7}$$

$$X1 \in \{X11, X12, X13, X14, X15\}, \tag{8}$$

$$X2 \in \{X21, X22, X23, X24, X25\}. \tag{9}$$

For each pair of head candidates $S1\times$ and $S2\times$, the algorithm gives the cost for each pair of facial side candidates $X1\times$ and $X2\times$, where $\times$ stands for a digit 1, 2, 3,... Next, the proposed edge confirmation criterion measures the fit of line candidates to the input image. We cite one illustrated example for localization as displayed in lines $S1$, $S2$, $X1$, and $X2$ of Fig. 1(e). The cost of the possible head-face boundary, lines $S1$ and $X1$, on the left side of the image is given by

$$\#LV1 = \sum_{i=Y1-Z}^{Y1-1} Nr(B(x,i)), \tag{10}$$

$$\#LV2 = \sum_{i=Y1}^{Y1+Z-1} Nr(B(x,i)), \tag{11}$$

$$\#LD1 = \sum_{i=0}^{\frac{Z}{2}-1} Nr(B(X1-i,Y1+i)), \tag{12}$$

$$\#LD2 = \sum_{i=0}^{\frac{Z}{2}-1} Nr(B(X1+i,Y1-i)), \tag{13}$$

$$\#LE = \#LV2 + \#LD2, \tag{14}$$

$$Nr(B(x,y)) = \begin{cases} 1 & if\ B(x,y) = 255 \\ 0 & if\ B(x,y) = 0 \end{cases}, \tag{15}$$

where $\#LV1$ and $\#LV2$ are number of points on the upper and lower line segments of $X1$, $\#LD1$ and $\#LD2$ are number of points on the lower and upper line segments of $S1$, and $\#LE$ is the sum of the line segments $\#LV2$ and $\#LD2$. Similarly, the cost of the possible head-face boundary, lines $S2$ and $X2$, on the right side of the image are given by $\#RV1$, $\#RV2$, $\#RD1$, $\#RD2$, and $\#RE$. While the cost of each boundary candidate measures the boundary possibility of each candidate, the filtering condition evaluated by using the equation (16) screens the candidate FOI regions and the decision condition made on both sides is based on the largest numbers #*LE* and #*RE* to complete detecting which the face locates, respectively.

$$\begin{aligned} &(\#LV2 > \#LV1)\ \&\&\ (\#LD2 > \#LD1)\ \&\& \\ &(\#RV2 > \#RV1)\ \&\&\ (\#RD2 > \#RD1)\ \&\& \\ &(\#LV2 > Z/4)\ \&\&\ (\#LD2 > Z/8)\ \&\& \\ &(\#RV2 > Z/4)\ \&\&\ (\#RD2 > Z/8), \end{aligned} \tag{16}$$

where Z denotes the region of check points for computation and we set Z to a value $H/4$. The result of Fig. 1(f) shows that a successful face localization is obtained based on the aspect ratio of the face shape, which has been set to be 6:4 in this algorithm. Once the HCD approach fails to detection, in what follows, the face localization will renew and search one of nine subimages with size two-thirds of the input image, which were given by overlap subdivision. The localization will not stop until an eye pair in the input image is found or the scans are over all the subimages.

3 Eye Extraction

The algorithm for the extraction and verification of the eye pair is described as follows:

Step 1: Perform discrete wavelet transforms (DWT) using D4 scalar wavelet for the FOI subimage as shown in Fig. 1(f) and take both the LH detail subband D_{LH} and HL detail subbad D_{HL} (Fig. 1(g)), since the separable sampling in DWT provides divisions of spectrum with sensitivity to horizontal eye edges and vertical facial sides, respectively.

Step 2: Remove the associated facial edge on the LH subband to avoid a false alarm in eyes projection, provided there is a significant peak response on either side of the HL subband.

Step 3: Binarize the FOI subimage obtained from the previous step to delete noise-like coefficients by adaptively selecting a reasonable threshold using the wavelet histogram of this region (Fig. 1(h)). Let B_{LH} represent the thresholded output.

Step 4: Project *y*-axis integrally and pick at most the first three peak responses as bases. The point projections of $B_{LH}(i, j)$ along its rows is given by

$$P_h(i) = \sum_{j=1}^{3n/4} B_{LH}(i, j) \;\; 1 \le i \le m \,. \tag{16}$$

The *y*-axis projection area is within the confines of three fourths $R^{m \times n}$ to avoid the possible false alarm caused by the mouth region. Then we can use *j* that has a maximum value of $\{P_h(j)\}_{\max}$ from $P_h(j)$ to restrict the *x*-axis projection region. We note that *j* in $\{P_h(j)\}_{\max}$ slightly varies according to the appearance within the candidate as shown in (1) of Fig. 1(i). Therefore, as shown in (2) of Fig. 1(i), we can decide on the basis of the coordinate $\{P_h(j)\}_{\max}$ to constraint the *x*-axis projection around the area delimited by the projective upper and lower valleys, $P_h(j)\}_{u\_\min}$ and $\{P_h(j)\}_{l\_\min}$, respectively. So the point projections of $B_{LH}(i, j)$ along its column is given by

$$P_v(i) = \sum_{i=1}^{m} B_{LH}(i, j) \;\; \{P_h(j)\}_{u\_\min} \le j \le \{P_h(j)\}_{l\_\min} \tag{17}$$

The above-presented inter-orientation projection does capture characteristics of the vertical profile and the horizontal symmetric distribution of human eye. Symmetrical projection relieves will encourage eye candidates to be identified, while asymmetrical projection relieves will favor others. According to anthropometry, we could circumscribe a circle area on the eye-mouth triangle by cropping the face of Fig. 1(c) for further screening the primary selection. For successful extraction of eyes, the eye-mouth circumscribed circle area (Fig. 1(j)) must pass over all examinations for verification. Based on facial component geometry eye-pair verification rules including eyes matching by correlation and mouth checking, which serves to tell us how like to each other they are, are adopted to verify the extracted eyes. The details are described in Step 5 below.

Step 5: The requirements in the eye verification procedure consist of the following:

1). The points of eye region are more than the ones of cheek region ((1) of Fig. 1(j)).
2). The points of both nose and mouth regions are more than the ones of two cheeks, or the points of mouth region are more than the ones of nose region ((1) of Fig. 1(j)).
3). The correlation coefficient is calculated to match the eye pair ((2) of Fig. 1(j)). Before obtaining the matching score between eyes, it is necessary to obtain normalizing for both size and rotation, which involves spatial scaling and finding of the eye pair centroids. Search the centroid in the eye region, we approach region segmentation by finding meaningful boundary based on point aggregation procedure. Choosing the center pixel of the eye region is a natural starting point and grouping points to form the region of interest with paying attention to 4-connectivity would yield a segmentation result, when no more pixels for inclusion in the region. The region segmentation result in general does not only give the required eye region, but also eyebrow or glasses, since there are still some thresholded wavelet coefficients located within the growing path between eye socket and eyebrow. In other words, besides the true eye landmarks, eyebrow and glasses, if any, will be included as well due to projection response spreading. After growing, the region centroid is relocated. By overlapping two centroids and with the help of two region borders we simply translate and rotate two regions so that they align themselves. The eye pattern decision is to perform matching via the correlation. The correlation factor was empirically determined as 0.5. The higher correlation value indicates that the two regions shapes are the best match. In determining the potential eye candidate using correlation matching, we aim to minimize the false acceptance rate with zero false rejection one. After the eyes are detected as described, the actual nose and the mouth locations, and the face size are also known.

Step 6: Refine the size and position of the FOI bounding box as an ellipse with the "+" symbol representing hitting the bull's-eye that is the most eyeball location. In order to make a compensation for region growing in the previous step, the "+" coordinates were modified by shifting one-third diameter of the grown region down in the vertical direction from the detected centroid coordinates. If the algorithm is failure to extract eyes, on the other hand, then it just outputs the earlier result of face localization (Fig. 1(k)).

4 Experimental Results

In order to evaluate the performance of the proposed system, we report on three different head-and-shoulder image sets, which are BioID [3], JAFFE [4], and Caltech [5]. The successful eyes detection was defined as excluding the rest face regions from the eye sockets. To make the results compared with the related works on the BioID and JAFFE test sets, we adopt the relatively error measure introduced by Jesorsky et al. [6]. The detection rate is hereby defined as the ratio of the number of correct detections to that of all images in the gallery. In Table 1, three test sets have been evaluated the performance of our method. From the experimental results on BioID, our system outperforms the results achieved by [6], [7], and [8]. On the JAFFE test set, our method compares favorably with the result reported in [7]. On the other hand, more false positives (erroneous) and false negatives (dismissal) than the other two sets are found on the Caltech test set, which has a little lower correct rate than the other two sets. For an overall evaluation, the experimental results show that the proposed eye detector can handle frontal view, facial variations (e.g., eye close or mouth open), pose, scale, and

view-independent rotation present in the database quite well. The correct rates for detecting eyes are all over 90%. However, the proposed system reports unsuccessful detection including false and missed eyes. We observe that our detector failed mainly for faces of too dark. If the facial components are not found because of lighting effect, which complicate the face localization task considerably, it is hardly to further extract eyes even though the face can be located. In terms of speed, the execution time of the presented detector is directly related to size and complexity of the images. For example, our system is operating at an average processing time 2.0 sec per BioID image (384 × 286 pixels) on a 1.4 GHz Pentium PC.

Table 1. Eye detection rates on three datasets using the HCD algorithm

Dataset (Sample number)	False positive	False negative	Correct
BioID (1521)	0.79% (12)	0.98% (15)	98.23% (1494)
JAFFE (213)	0% (0)	0.94% (2)	99.06% (211)
Caltech (450)	1.33% (6)	8.67% (39)	90.00% (405)

5 Conclusions

Experiments have shown high detection rates with a low number of false alarms on three datasets. To conclude, the proposed system demonstrates its robustness and high accuracy to that of the related works.

Acknowledgements

The financial support provided by the NSC 94-2622-E-151-014-CC3 is gratefully acknowledged.

References

[1] Jain, A. K., Ross, A., and Prabhakar, S.: An introduction to biometric recognition. IEEE Trans. Circuits and Systems for Video Technology, 14 (2004) 4-20.
[2] Mallat, S. (ed.): A Wavelet Tour of Signal Processing. Academic Press (1999).
[3] http://www.humanscan.de/
[4] http://www.mis.atr.co.jp/~mlyons/jaffe.html
[5] http://www.vision.caltech.edu/html-files/archive.html
[6] Kirchberg, K. J., Jesorsky, O., and Frischholz, R. W.: Genetic model optimization for Hausdorff distance-based face localization. Proc. ECCV2002, Denmark, LNCS-2359 (2002) 103-111.
[7] Zhou Z.-H. and Geng, X.: Projection functions for eye detection. Pattern Recognition, 37 (2004) 1049-1056.
[8] Wu, J. and Zhou, Z.-H.: Efficient face candidates selector for face detection. Pattern Recognition, 36 (2003) 1175-1186.