Face Detection and Precise Eyes Location

Weimin Huang and Robert Mariani

RWCP* Multi-Modal Functions KRDL † Laboratory,
21 Heng Mui Keng Terrace, Singapore 119613
Email: {wmhuang,rmariani}@krdl.org.sg

Abstract

This paper presents a robust and precise scheme for face detection and precise facial feature location. Multiscale filters are used to obtain the pre-attentive features of objects, based on which different models are investigated to locate the face and facial features such as eyes, nose and mouth. The structural model is used to characterize the geometric pattern of facial components. The texture and feature models are used to verify the face candidates detected before. Since the eyeballs are the only features that are salient and have strong invariant property, the distance between them will be used to normalize faces for recognition. Motivated from this, with the face detected and the structural information extracted, a precise eyes location algorithm is applied using contour and region information. It detects, with a subpixellic precision, the center and the radius of the eyeballs of a person's eyes. The detected result can be used as an accurate normalization of images, which reduces greatly the number of possible scales used during the face recognition process.

1 Introduction

On-line face detection in a scene is the first step in Automatic human face recognition. It is still a problem considering the variation of illumination, skin tone, face scale and orientation and the complex background of the image. And obviously, the face pattern detection and normalization play critical roles since the errors of recognition are caused partially by the errors of the face detection and components detection.

Existing face detection methods include template-based[1], neural network-based[2, 3], model-based[4], color-based[5] and motion-based approaches[6].

In this paper we present the model-based approach to obtain the face location and facial components positions. Compared with the template-based methods, model-based approach is faster and more flexible.

Once we located the face and fixed the facial components area, further studies on the precise components detection can be carried out. The detected positions will be used for normalization or recognition directly. A salient and often used feature for normalization is the distance between eyes. Many ways including template matching[1], and feature searching[7, 8] are proposed up to now for eye detection. However, in order to improve the performance of recognition, the robust and accurate eyes detection should be a must.

With the precise detection of the eyeballs, we are able to reduce greatly the number of the scales under consideration during the recognition, and therefore to improve the performance and the speed of the recognition process. The algorithm detects the center and radius of the eyeballs at subpixellic accuracy.

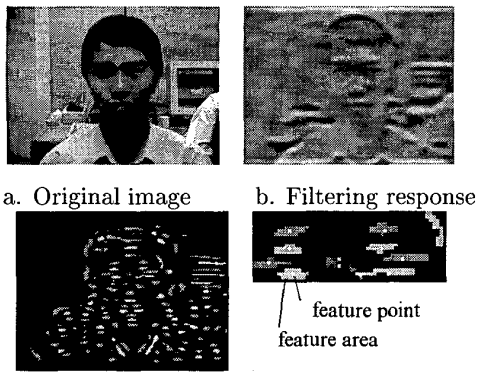
2 Pre-attentive feature detection

In low resolution, the eye or eyebrow in the face image usually are dark bars which can be easily detected by the elongated second derivative Gaussian filter. In fact, the nose and mouth are also dark bars when face is in low resolution. So the response of the filtering is a peak or valley in the center of such a feature.

Given a filter, only the features in the same scale and same orientation could be detected. However, we found that even with a fixed scale filter the pre-attentive features in certain range can also be detected correctly. As an example, given the 2nd derivative Gaussian filter that is 13 taps in y-direction (25 taps in x-direction), a face's components can be detected as pre-attentive features when the distance between two eyes of the face is from 18 to 37 pixels in experiments.

*Real World Computing Partnership
†Kent Ridge Digital Labs

0-7695-0750-6/00 $10.00 © 2000 IEEE



a. Original image b. Filtering response

c. Pre-attentive points d. Zoomed part

Figure 1. Pre-attentive feature detection

Of cause in order to detect features in different scales, multiscale filters in multiple orientations should be applied to the image. The face candidates screening is done from the smallest scale. If a face is detected, other scale space will not be searched. It can cover the scale range from 18 pixels to 74 (37×2) pixels and tilt range from -30 to +30 degree when detected in two discrete scales and three discrete orientations, which are enough for our system.

3 Facial image analysis

With the feature candidates obtained above, three models are investigated to search face pattern and facial components. The structure model is used to group feature points into face. It provides information on whether the area is face-like in structure. The texture model is used for similarity measurement with gray or color information for whether the pattern is face-like in texture. The feature model is used for feature measurement for whether a component is a facial feature.

3.1 Structure Model

The structure model groups the feature focusing points into face candidates using the geometric relationships defined in [4]. In this paper, a simplified version of face structure model is presented as follows.

For eyes area, two possible sub-structures are shown in Fig.2.a, two pairs of eye-eyebrow are shown, and Fig.2.b, one pair of eyes are shown. For mouth area, with the information of eye-pair, there is at least one local minimum feature point in the corresponding area. Considering the different expressions and other noise, there may be two or more local minima in the area. Shown in Fig.3, the mouth/nose will be one of the sub-structures.

All the relationships among the substructures should satisfy certain geometric conditions. In system, the sub-structures of eye-pair are detected first. With a sub-structure of eye-pair, the corresponding sub-structure of mouth-nose are searched. The *real world face* candidate is composed of the two sub-structures. Then the affine transformation are applied to the *face structure model* to fit the real world face structure.

a. Eye-eyebrows detected separately (left)
b. Eye-eyebrows detected as one maximum (right)

Figure 2. Eye area sub-structure, in the figure, gray area is the corresponding mouth area

a. Mouth area, nose-mouth structure (left)
b. More maxima detected (right)

Figure 3. Mouth area sub-structure, in the figure, gray area is the corresponding eyes area

3.2 Texture Model

The texture model measures the gray or color similarities of a candidate with face model, including the variation between the facial regions (eyes, nose etc.), the symmetry of the face and the color/gray texture similarity between two regions of the face. Moreover, in the model, the symmetry of face in the filtered domain is enhanced because the related brightness is kept almost the same.

To measure the texture similarity, we use the two cheek areas, which are defined as the areas below eyes and at the side of nose. One example is shown in Fig.4.

The texture measurement is on two features. One is the gray level variance in the area:

$$V_l = \sqrt{\sum_{i,j}(I(i,j) - M)^2},$$

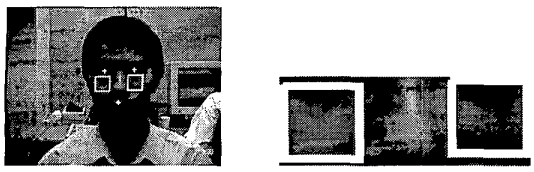


Figure 4. Cheek areas defined by the eye positions

where $I(i,j)$ is the intensity value at pixel (i,j) and M is the mean of the gray values in the area. The other kind of feature is gradient variance in the area.

$$V_G[n] = \sqrt{\sum_{i,j} G_I[n](i,j)^2},$$

where $G_I[n] = I * G[n]$, $G[n], n = 0,1,2,3$ are the 2D Sobel operators.

Let V_l^{left} be the V_l in the left cheek, V_l^{right} be the V_l in the right cheek, $V_G^{left}[n]$ be the $V_G[n]$ in left cheek and $V_G^{right}[n]$ be the $V_G[n]$ in the right cheek. The texture symmetry of the two cheek areas is then characterized by

$$R_l = \frac{|V_l^{left} - V_l^{right}|}{V_l^{left} + V_l^{right}},$$

$$R_G[n] = \frac{|V_G^{left}[n] - V_G^{right}[n]|}{V_G^{left}[n] + V_G^{right}[n]}, n = 0,1,2,3.$$

The ratio values are ideally 0. However in order to reduce the influence of spectacles and the variation of illumination and skin tone, we set

$$R_l < 0.30$$
$$R_G[n] < 0.33,\ n = 0,1,2,3$$

as the face texture model. The texture measurement itself can also be a feature of face. We believe that V_ls and V_Gs in all of the subregions that are below the eye region have the characteristics that can distinguish face pattern and many other non-face patterns.

3.3 Feature Model

The feature model compares the feature area to specific facial feature. Here we use the eigen-eyes method [9] combined with image feature analysis for eyes detection. Since the scale information has been obtained for each face candidate, via the structure model, the eigen-eye method can be applied here efficiently.

The normalized horizontal and vertical projections of the image of eye areas are the first kind of feature. The correlation of the projection with the template that is trained by samples is taken as a parameter of the similarity measurement.

Another important feature is the direction of the detected preattentive feature. Because the two eyes are consistent with each other in the direction, the directions detected along the eyes should be almost the same. The feature is used in combination with structure model (cf. section 3.1). More details of using image features to eye detection are presented in section 4.

3.4 Elimination of Conflicting Candidates

Let f_1 and f_2 the two conflicting face candidates, and $T^1 = (R_l^1, \{R_G^1[n], n = 0,1,2,3\})$ and $T^2 = (R_l^2, \{R_G^2[n], n = 0,1,2,3\})$ are respective texture measurement vectors (cf. section 3.2). Because the elements of the vector T^1 (resp. T^2), indicate the flatness of the cheek of f_1 (resp. f_2), we use the difference between the two vectors, denoted by D_{flat}, to indicate the similarity of flatness:

$$D_{flat} = \frac{1}{N}\sum_i (T_i^1 - T_i^2), N = 5.$$

The similarity in feature model is characterized by the eigen-eyes similarity. Let D_{DFFS}^1 and D_{DFFS}^2 the measurement of eigen-feature [9] similarity in the feature model for the eye pair 1 in face f_1 and the eye pair 2 in face f_2. We define

$$D_{eigen} = D_{DFFS}^1 - D_{DFFS}^2.$$

Using D_{flat} and D_{eigen}, we define the distance measure as

$$D(f_1, f_2) = \omega D_{flat} + (1-\omega) D_{eigen},$$

where ω=0.25. Notice that D_{flat} and D_{eigen} are already normalized, $0 \leq D_{eigen}, D_{flat} \leq 1$. The decision is

$$Detected\ face = \begin{cases} f_1, & if\ D(f_1,f_2) > \alpha \\ f_2, & if\ D(f_1,f_2) < -\alpha \end{cases}.$$

Empirically, the value of α is a very small positive number.

4 Precise Eye Location

Here, we propose a method to detect the center and the radius of the eyeballs with a subpixellic precision, considering the face recognition is done on face images that are normalized with the eyes position decided previously by the face model. We start from the initial eye

position (x, y), and we look for the homogeneous circular regions as eyeballs centered at (x^*, y^*) and having a radius r^* in the subimage centered in (x, y) containing the eye. This research is realized in the zoomed image, and by the reverse coordinates transformation, we obtain the real coordinates with a precision of half a pixel.

We combine two complementary measures based on the edge information, namely the hough transform for the circle and the contour-correlation of circle model with the image, and we eliminate the invalid hypothesis using a measure of homogeneity associated to the image defined by the current disc. Finally, we use a robust method to select the best circle among all the possible circles.

4.1 Preprocessing

This preprocessing consists in three ordered steps. First, we construct a zoomed image of the eye, centered in (x, y), and we normalize it, in order to cancel the linear changes of contrast and brightness [10]. Then, we extract the edges from this normalized image, and finally, we improve the quality of the eyeball region, by reducing the quantity of light which is reflected within it.

The zoom factor applied to the face is determined, using the previous position of the two eyeballs, says (x_1, y_1) and (x_2, y_2), such that the final distance separating two eyes in the zoomed image is equal to 100 pixels. For the face recognition, we work with normalized images, such that the distance between two eyes is equal to 50 pixels. In these images, we noticed, before the proposed precise eyes detection, a position error of ± 3 pixels of the eyeball centers, crucial points for a good geometrical normalization and thus for a good recognition. Therefore, the precision obtained in the zoomed image is of half-pixel, in respect to the normalization for the recognition.

Figure 5. The Canny-deriche edge detection

4.2 Hypothesis Evaluation

A hypothesis $H(x, y, r)$ is a possible eyeball centered at (x, y) with radius r. To build the hypothesis set, we proceed in four steps: 1) using *a priori* knowledge, we select the set of likely hypothesis; 2) we compute the hough transform; 3) we compute the reciprocal operation using a contour correlation technique; 4) we keep only the homogeneous regions.

Figure 6. Cancelling the light reflected in the eye

original image (left); filtered image (middle); smoothed image (right)

Using an automatic thresholding method, we get the eyeball region by segmentation of the dark pixels, which is defined as the pixels having gray level less than a threshold τ, so that all the dark pixels occupy a small part, say 15%, of the eye area.

Finally, we obtain the set $H = \bigcup_{r=6}^{10} H(r)$ of the likely hypothesis, using the dilation (neighborhood of 2 pixels),

$$H(r) = \{(x, y, r) : \exists (i, j) : I(x + i, y + j) \leq \tau\}$$

where $-2 \leq i, j \leq 2$.

Figure 7. Automatic Thresholding Results

4.2.1 Circular Detection

Both the Hough transform and the contour correlation are used for finding discs on the contour image. The parameters are the disc center (x, y) and the radius r. Then, the hypothesis (x, y, r) is kept for subsequential analysis if

$$H = \{(x, y, r) \in H : acc(x, y, r) > M\}$$

where $acc(x, y, r)$ is the accumulator in Hough transform and $M = 3$.

The correlation of a hypothesis (x, y, r) with the digital circle $C(x, y, r)$ is computed as

$$cor(x, y, r) = \frac{1}{N} \sum_{(a,b) \in C(x,y,r)} \alpha(a, b)$$

where N is the number of pixels in the circle model,

$$\alpha(a, b) = \begin{cases} 0 & \text{if } \beta(a, b) <= K \\ 1 & \text{if } \beta(a, b) > K \end{cases}$$

$\beta(a,b)$ is the number of points having a nonzero edge magnitude in a 3x3 neighborhood of (a,b), which allows more robustness, against the variability of the shape of the observed circle, and against the possible breaks of the contour,

$$\beta(a,b) = \sum_{i=-1}^{1} \sum_{j=-1}^{1} contour(a+i, b+j).$$

Here $K = 3$.

The hypothesis are rejected by contour correlation, when the obtained score is too weak,

$$H = \{(x,y,r) \in H : cor(x,y,r) > th_{cor}\}.$$

with the threshold th_{cor} fixed to 0.6, which means that we accept an hypothesis only if at least 60% of the circle is present in the edge image.

4.2.2 Region Homogeneity

A hypothesis is an eyeball if 1) it is a homogeneous (monochromatic) region, 2) it is darker than the white region surrounding it. In order to compensate the presence of light within the eyeball, we use the grey level image, filtered in the preprocessing step.

A normalized standard deviation is used here as the homogeneity measure,

$$hom(x,y,r) = \frac{1}{255}\sigma\left(\frac{r_{min}}{r}\right)^2,$$

where σ is the standard deviation of the gray level in the region delimited by the digital disc $D(x,y,r)$ and r_{min} is the minimum radius allowed (here 6 pixels). The smaller the value of $hom(x,y,r)$, the better the homogeneity. With the intensity mean μ to measure the darkness of the region, we have

$$H = \{(x,y,r) \in H : \mu < \tau \wedge \mathrm{hom}(x,y,r) < th_\sigma\},$$

where τ has been fixed in the pre-selection step (4.2), and th_σ is defined for the disc of radius r_{min}

$$th_\sigma = \frac{5}{255} = 0.02.$$

It means that we tolerate a standard deviation of 5 grey levels around the mean value within 100 pixels.

4.3 Optimal Decision of the Location

We obtained three values for each valid hypothesis (x,y,r): 1) $acc(x,y,r)$; 2) $cor(x,y,r)$; 3) $hom(x,y,r)$. In order to get the best hypothesis of eyeball (x^*,y^*,r^*), we first select the best center (x^*,y^*) which minimizes a cost function, and secondly, we research the best radius.

4.3.1 Selecting the Best Center

The set V_r of the likely disc hypothesis is defined as the set of disc for which reasonable values of the contour correlation and the homogeneity values have been obtained

$$\begin{aligned} V_r &= \{(x,y,r) \in H : \\ &\quad cor(x,y,r) > \alpha \wedge hom(x,y,r) < \beta\}. \end{aligned}$$

for $\alpha = 0.6$ and $\beta = 0.015$. The potential centers set V is then given by

$$V = \{(x,y) : \exists (x,y,r) \in V_r\}$$

To select the best center in V, we define a cost function which combines the scores of Hough transform and contour correlation,

$$N(x,y) = \gamma \widehat{H}^2(x,y) + (1-\gamma)\widehat{C}^2(x,y)$$

where

$$\widehat{H}(x,y) = 1 - \frac{H(x,y)}{H_{max}}\ ,\ \widehat{C}(x,y) = 1 - \frac{C(x,y)}{C_{max}}$$

with

$$\begin{array}{ll} H(x,y) = \int_r acc(x,y,r) & H_{max} = \max_{(x,y)} H(x,y) \\ C(x,y) = \int_r cor(x,y,r) & C_{max} = \max_{(x,y)} C(x,y) \end{array}$$

Here $\gamma = 0.3$ weights the relative importance.

The values of N are normalized, and the optimal center position is then provided by the minimum value of $N(x,y)$, that is

$$(x^*,y^*) = \arg\min_{(x,y)} N(x,y).$$

4.3.2 Selecting the Best Circle

Let V_r^*, the set of the likely hypothesis centered in the neighborhood of (x^*,y^*), with $\epsilon_x = \epsilon_y = 1$,

$$V_r^* = \{(x,y,r) : |x - x^*| \leq \epsilon_x \wedge |y - y^*| \leq \epsilon_y\}.$$

The optimal hypothesis (x_c,y_c,r_c) is searched in V_r^*. For two hypothesis $h_1 = (x_1,y_1,r_1)$, $h_2 = (x_2,y_2,r_2)$, we define the order relation $h_1 >^* h_2$, if

$$cor(h_1) > cor(h_2) + \epsilon_{cor}$$

or

$$|cor(h_1) - cor(h_2)| < \epsilon_{cor} \wedge hom(h_1) < hom(h_2)$$

where $\epsilon_{cor} = 0.001$. We then obtain the optimal point $h_c = (x_c,y_c,r_c) \in V_r^*$ such that

$$h_c >^* h, \forall h \in (V_r^* - \{h_c\})$$

and the hypothesis is ultimately accepted if

$$\pi(h_c) > \pi,\ with\ \pi(h_c) = cor(h_c) * (1 - hom(h_c)).$$

For the prefixed thresholds, namely α for the correlation and β for the homogeneity, the rejection threshold π has been fixed such that $\pi > \alpha(1-\beta)$. With $\alpha = 0.6$ and $\beta = 0.02$ (cf. 4.2.1,4.2.2), we can have $\pi = 0.6$.

5 Results and Discussion

We have proposed a scheme for face location and accurate eyes detection, which is based on multiple evidences, including facial components structure, texture similarity, component feature measurement and contour matching.

Applying the multiscale and multi-orientation filters bank to images, the proposed method can detect faces in the size from 18 pixel to 74 pixel and tilted from -30 degree to +30 degree.

The original image size is 384 × 288. For one image, it takes about 5.0 seconds to capture the eyes on Sun Ultra1 workstation, considering to search all the scale and orientation space. It can be faster if the color information is used to segment the image first. Upon the first face captured, the face tracking can be implemented for the on-line detection and human-machine interface. Total 852 images are captured, in which some persons provided several times for capturing their faces with different conditions, such as that of the illumination, background, with or without glasses, a little expressions etc.

Figure 8. Faces detected with eyes marked

There are 84 errors in face detection which are caused by tilt face (tilt angle larger than 30 degree), too much rotation in depth (so not front view), illumination unbalance including the bright reflection on the glasses, and model unfitting, the feature model in the current system can not fit too *dark* skin.

Based on the face detected, we also proposed a precise eye detection (half-pixel) combining contour and region information extracted from a zoomed image. The accuracy of the detection is defined by the difference between the one detected automatically and the one fixed manually. Here we have tested it against two databases. One is built from video images that contains more than 200 images. The other is a photo image database with more than 5000 frontal view faces.

Noticeable improvements have been realized, especially in the rotation and scale normalizations. In video image database, we got 100% eyes located accurately provided the face detected correctly before. On the photo database, only 2% failure is reported for either no eyes detected or wrong eyes position obtained, which are mainly caused by the poor quality of the images.

Another use of this algorithm is to evaluate, before the recognition, if a detected face is suitable for a good normalization or not, and therefore, for a successful recognition or not. To a specific hardware configuration, if the detected eyeballs are judged too small, too close or too far from each other, we expect that the normalization will be very unstable. Here the probability of confusion with another face is reduced, of cause with the increase of the rejection rate.

References

[1] S. Gutta et. al. Face recognition using ensembles of networks. In *Proc. of Int. Conf. on Pattern Recognition.* Vienna, Austria, Aug. 1996.

[2] H. A. Rowley, S. Baluja and T. Kanade. Human face detection in visual scenes. Technical Report CMU-CS-95-158, Dept of Computer Science, Carnegie Mellon University, July, 1995.

[3] S.-H. Lin, S.-Y. Kung and L.-J. Lin. Face recognition/detection by probabilistic decision-based neural network. *IEEE Trans. on Neural Networks*, 8(1):114–132, 1997.

[4] K. C. Yow and R. Cipolla. Feature-based human face detection. Technical Report CUED/F-INFENG/TR 249, Dept of Engineering, University of Cambridge, Cambridge, England, Aug, 1996.

[5] Q. Chen, H. Wu and M. Yachida. Face detection by fuzzy pattern matching. In *Proc. of 5th Int conf on Computer Vision*, pages 591–596. MIT, Boston, 1995.

[6] J. L. Crowley and J. Coutaz. Vision for man machine interaction. In *Proc. of Working Conference on Engineering for Human-Computer Interaction*, pages 187–217. (Grand Targhee Resort), 1995.

[7] J. Bala et. al. Visual routine for eye detection using hybrid genetic architectures. In *Proc. of Int. Conf. on Pattern Recognition.* Vienna, Austria, Aug. 1996.

[8] A. Yuille, D. Cohen and P. Hallinan. Feature extraction from faces using deformable templates. In *IEEE Computer Soc. Conf. on Computer Vision and Patt. Recog.*, pages 104–109, 1989.

[9] A. Pentland, B. Moghaddam and T. Starner. View-based and modular eigenspaces for face recognition. Technical Report No. 245, Perceptual Computing Section, Media Laboratory, MIT, 1994.

[10] T. H. Reiss. Recognizing planar objects using invariant image features. In *Lecture Notes in Computer Sciences, Springer Verlag, Vol. 676*, pages 14–15, 1993.