Eyes Location by Neural Network-Based Face Segmentation

Xiao-yi Feng

College of Electronics and Information, Northwestern Polytechnic University, Xi'an, China

Summary
This paper proposed a neural network-based face segmentation method for eyes location. After face is detected, it is segmented automatically by a competitive Hopfield neural network (CHNN) and the facial feature candidates such as eyes, eyebrows and mouth are obtained, and then eyes are located by facial features evaluation and validation, which is based on face's geometrical structures. To enhance the difference between facial features and other face area, the color face image is converted into a color-ratio grey image before face segmentation. Experimental results show that the proposed system performs well.
Key words:
Competitive Hopfield neural network, Eyes location, Face segmentation, color-ratio image

Introduction

Robust non-intrusive eyes location plays an important role in vision based man-machine interaction including automotive applications, such as driver inspection, face recognition and facial expression recognition, etc. In the past years, many works were addressed on this area. The most known methods include color-based approaches, neural network approaches, genetic algorithm approaches, and principle component analysis approaches [1-6].

Most eyes location methods contain three steps: face detection, facial features candidate detection and features validation, of which, facial feature candidates' detection is the most important one.

In this paper, a novel method for eyes location is proposed, which is based on a neural network based face segmentation and which is mainly for improving the performance of facial features candidate detection,. First, an illumination invariant skin model suggested in [7] is applied to extract face area effectively; Then, face image (color image) is converted into a color-ratio image (grey image); to enhance the difference between facial features and other face area; Then, both upper part and lower part of the color-ratio image are segmented separately, using a Hopfield neural network. As a result, we get the collection of facial features such as eyes, mouth, eyebrows candidates. These facial candidates are verified by face's geometric structure during eyes validation procedure.

The rest of the paper is organized as follows. Face detection method and face transformation is presented in section 2. Face segmentation using CHNN is introduced in section 3. In section 4, Facial feature evaluation and eyes detection is presented. Finally in section 5 we conclude the paper.

2. Face Detection and Face Image Transformation

2.1 Skin Locus Based Face Detection

In our work, face-like regions in an input image are detected using the skin detection method proposed by Martinkauppi et al. [7], who had found that the Normalized Color Coordinates (NCC) combined with skin locus most appropriate for skin detection under varying illumination. To detect face-like area, the image presented in RGB color space is converted to the NCC space *r*, *g* and *b*. Since $r+b+g = 1$, only two chromaticity *r* and *b* are used for detection. If *r* and *b* of a pixel fall into the area of the skin locus, the pixel belongs to skin. Skin detection result is shown in Fig.1 (b). This result is enhanced by morphological erode and dilate. Considering real applications, the largest skin component is regarded as the most likely face area (see the part inside the green box in Fig.1 (b)).

(a) An input image (b) Face area

Fig. 1 Face detection

2.2 Face Image Transformation

Based on the knowledge that facial features are darker than their surroundings, morphological valley detectors are usually used for features detection. While under bad lighting conditions, facial features are not quite different from other part of face area, and so this features detection methods will be failed. Fig.2 describes this case. Fig.2 (a) is the gray face image, which was obtained by converting a color image into the gray one directly and cropping it to erase non-face area. It is obviously that its facial features are not too different from other face area in gray level. Fig.2 (b) shows its valley detection result. It can be seen that the facial features are not well detected. Fig.2 (c) is its eyes location result using geometric structure-based feature validation. In this case, eyes are not located correctly.

(a) Grey image (b) Valley detection (c) Eyes location

Fig.2. Valley-based eyes detection

To make facial feature candidates can be detected more precisely, it is obviously that they should be quite different from other face area. Here we propose a new method for this purpose: The color face region is converted to a gray level image named color-ratio image as follows, base on the observation that eyes, eyebrows contain less red elements and lip contains more than the skin part.

$$f(x,y) = \min(255, {b \times 255}/{r}) \qquad (1)$$

Here $f(x,y)$ is the gray level of a pixel in position (x,y) in the color-ratio image, and r and b are chromaticity in the NCC space. The color-ratio image corresponding to the image in Fig. 1(a) is shown in Fig. 3(a). Compare images in Fig. 2(a) and Fig. 3(a), we can see that features are more different from other face area after the proposed face image transformation.

3. Face Segmentation

It can be seen from the color-ratio image that eyes and eyebrows are with higher gray levels than their surroundings and lips are with lower gray levels. This knowledge is used for face segmentation.

To make our method insensitive to noise, both pixels' gray levels and their local mean gray values are taken into account during face segmentation, so a two-dimensional gray vector is composed of them in our work. Due to its fast convergence, the competitive Hopfield neural network is used for clustering.

Let L denote the total gray level and N denote the number of pixels in the image. The gray level of a pixel (suppose as i) and its mean gray value (suppose as j) forms a group of (i, j). Let f_{ij} and p_{ij} are the number and the associated probability density of group (i, j).

$$\begin{aligned} & p_{ij} = f_{ij} / N , \\ & \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p_{ij} = 1 , \; p_{ij} \geq 0 , \qquad (2) \\ & i, j = 0,1,2,3 \cdots L-1 \end{aligned}$$

Distance between group (x_1, y_1) and (x_2, y_2) is defined as

$$d_{\mathrm{x_1y_1x_2y_2}} = \frac{1}{2} \times \left[(x_1 - x_2)^2 + (y_1 - y_2)^2 \right] \qquad (3)$$

Suppose the two-dimensional vectors are classified into two kinds so CHNN consists of $L \times L \times 2$ neurons. Let V_{xyc} denote a neuron's state at (x,y,c), where x and y are the first and second part of a two-dimensional vector and c is its classification. The following conditions will be met when the system is steady.

$$\begin{aligned} & V_{\mathrm{xyc}} \in \{0,1\}, \quad \sum_{\mathrm{c}=0}^{1} V_{\mathrm{xyc}} = 1 \\ & 0 < \sum_{x=0}^{\mathrm{L\text{-}1}}\sum_{\mathrm{y}=0}^{\mathrm{L\text{-}1}} V_{\mathrm{xyc}} < L^2 , \sum_{x=0}^{\mathrm{L\text{-}1}}\sum_{\mathrm{y}=0}^{\mathrm{L\text{-}1}}\sum_{c=0}^{1} V_{\mathrm{xyc}} = L^2 \end{aligned} \qquad (4)$$

The energy function is constructed as

$$\begin{aligned} E = & \frac{A}{2} \sum_{x_1=0}^{L-1}\sum_{y_1=0}^{L-1}\sum_{c-0}^{1}\sum_{x_2=0}^{L-1}\sum_{y_2=0}^{L-1} \frac{1}{\sum_{y=0}^{L-1}\sum_{x=0}^{L-1} p_{xy}V_{xyc}} t_{x_1y_1x_2y_2c} + \\ & + \frac{\mathrm{B}}{2} \sum_{x=0}^{L-1}\sum_{y=0}^{L-1}\sum_{c_1-0}^{1}\sum_{c_2=0}^{1} V_{xyc_1}V_{xyc_2} + \qquad (5) \\ & + \frac{\mathrm{C}}{2} [\sum_{c=0}^{1}\sum_{y=0}^{L-1}\sum_{x=0}^{L-1} V_{xyc} - L^2] \end{aligned}$$

Here $t_{x_1y_1x_2y_2c} = V_{x_1y_1c} d_{x_1y_1x_2y_2} p_{x_2y_2} V_{x_2y_2c}$.

Introduce the winner-takes-all(WTA) learning rule that is

$$V_{xyc} = \begin{cases} 1 & if \;\; Net_{xyc} = \max\limits_{0<j<2} \{Net_{xyc}\} \\ 0 & otherwise \end{cases} \quad (6)$$

Formula (4) is reduced to

$$E = \frac{1}{2} \sum_{x_1=0}^{L-1} \sum_{y_1=0}^{L-1} \sum_{c-0}^{1} \sum_{x_2=0}^{L-1} \sum_{y_2=0}^{L-1} \frac{1}{\sum\limits_{y=0}^{L-1} \sum\limits_{x=0}^{L-1} p_{xy} V_{xyc}} t_{x_1 y_1 x_2 y_2 c} \quad (7)$$

The input of neuron (x, y, c) is

$$Net_{xyc} = \frac{-1}{\sum\limits_{x_1=0}^{L-1} \sum\limits_{y_1=0}^{L-1} p_{x_1 y_1} V_{x_1 y_1 c}} \sum_{x_2=0}^{L-1} \sum_{y_2=0}^{L-1} tt_{xyx_2 y_2} \quad (8)$$

Here $tt_{xyx_2y_2} = d_{xyx_2y_2} \, p_{x_2y_2} V_{x_2y_2c}$.

After network's initialization, formula (6) and (8) are used to update every neuron's state and all the pixels are classified by the clustering result.

The upper part and the lower part of the color-ratio image are segmented respectively using the network and the results are shown in Fig.3 (b) and (c).

4. Facial feature evaluation

After the possible facial features are detected, a similar method as proposed in [8] is applied to evaluate feature constellations, using a geometrical face model including eyes, eyebrows, nostril and mouth.

We first select two facial features locating at the upper half of face area to form a possible eye pair and evaluate each possible eye pair as follows:

$$E_{eyepair} = 0.5\exp(-10(\frac{D_{eyes} - 0.4B_{width}}{D_{eyes}})^2) + \\ + 0.25\left|\theta_{eyeleft} + \theta_{eyeright} - 2 \times \theta\right| \quad (9)$$

Here B_{width} is the width of the face bounding box, D_{eyes} is distance between one possible eye pair and $D_{eyes} < 0.8B_{width}$. $\theta, \theta_{eyeleft}$ and $\theta_{eyeright}$ indicate directions of base line(The line passing through the centers of the eye candidates is called as base line) , left eye candidate and right eye candidate, respectively. The first item of this expression uses the distance between two eyes and the second item uses the direction of two eyes as eye pair constrains.

For each eye candidate pair, other facial features are evaluated as follows.

$$E_{feature} = \exp(-10(\frac{d_{feature} - D_{feature}}{D_{eyes}})^2) \quad (10)$$

Here $features = \{mouth, nostril, eyebrow\}$, $d_{feature}$ and $D_{feature}$ are real distance and reference distance from features to base line.

The total evaluation value is a weighted sum of values for each facial feature. The weights for eye pair, mouth, nostril pair, and eyebrow pair are 0.4, 0.3, 0.1 and 0.05, respectively. The constellation which has the largest evaluation value and which is bigger than a valve (for example, 0.4) is assumed to be real facial features (if the largest evaluation is less than the valve, it is regarded that there is no face in the image or no eyes in the image). Fig.2 (c) and Fig.3 (d) are detection results.

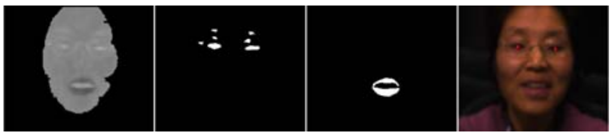

Fig.3. Neural-network based eyes detection. (a) Color-ratio image (b) Upper face segmentation (c) Lower face segmentation (d) Eyes detection

Compare Fig3. (b), Fig.3 (c) to Fig.2 (b), we can see that facial feature candidates can be more precisely detected in our method. As a result , eyes are located exactly in our method.

5 Conclusions

Effective facial feature candidates' detection is important for precise eyes location. This paper presents an automatic method for improving performance of facial features detection, using the combination of image transformation and neural network-based face segmentation. After face detection, image transformation is adopted to enhance the difference between facial features and other face area, then CHNN is used for detecting facial feature candidates. CHNN converges fast for its competitive learning mechanism, which assures its application in real time eyes location. Experimental results demonstrate the performance of the proposed method.

Acknowledgments

This work was supported by the natural science foundation of Shaanxi province and by the NWPU talent program.

References

[1] Wong, K.W., Lam, K.M., Siu W.C.: A Robust Scheme for Live Detection of Human Faces in Color Images. Signal Processing: Image Communication 18(2) (2003) 103-114
[2] Yu, L., Eizenman, M.: A New Methodology for Determining Point-of-Gaze in Head-Mounted Eye Tracking Systems. IEEE Trans. Biomedical Engineering 51(1) (2004) 1765-1773
[3] Paperno, E., Semyonov, D.: A New Method for Eye Location Tracking. IEEE Trans. Biomedical Engineering 50(1) (2003) 1174-1179
[4] Kapoor, A., Picard, R. W.: Real-Time, Fully Automatic Upper Facial Feature Tracking. In: Williams, A. D. (eds.) Fifth IEEE Int. Conf. Automatic Face and Gesture Recognition (2002) 10-15
[5] Huang, J., Wechsler, H.: Visual Routines for Eye Location Using Learning and Evolution. IEEE Trans. Evolutionary Computation 4(1) (2000) 73-82
[6] Yoo, D. H., Chung, M. J.: Non-Intrusive Eye Gaze Estimation without Knowledge of Eye Pose. In: Proc. Sixth IEEE Int. Conf. Automatic Face and Gesture Recognition (2004) 785-790
[7] Martinkauppi, B.: Face Color under Varying Illumination-Analysis and Applications. University of Oulu, Finland (2002)
[8] J.Hannuksela: Facial Feature based Head Tracking and Pose Estimation. University of Oulu, Finland (2003)

Xiaoyi Feng received the B.S. and M.S. degrees in Electrical Engineering from the Northwest University in 1991 and 1994, respectively. She received the doctor's degree from the Northwestern Polytechinic University in 2001. She now works in the Northwestern Polytechinic University, China.