Eye location under different eye poses, scales, and illuminations

Jinghe Yuan (袁景和)

Key Laboratory of Molecular Nanostructure and Nanotechnology, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, China

E-mail: jacobyuan@yahoo.com.cn

Received March 10, 2009

Robust non-intrusive eye location plays an important role in vision-based man-machine interaction. A modified Hausdorff distance based measure to localize the eyes is proposed, which could tolerate various changes in eye pose, shape, and scale. To eliminate the effects of the illumination variations, an 8-neighbour-based transformation of the gray images is proposed. The transformed image is less sensitive to illumination changes while preserves the appearance information of eyes. All the localized candidates of eyes are identified by back-propagation neural networks. Experiments demonstrate that the robust method for eye location is able to localize eyes with different eye sizes, shapes, and poses under different illuminations.

OCIS codes: 150.1135, 100.2000, 100.3008.

doi: 10.3788/COL20100801.0059.

Robust non-intrusive eye location plays an important role in vision-based man-machine interaction including automotive applications, such as driver inspection, face recognition, etc. In the past years, many works were addressed on this area. There are two major approaches for automatic eye detection. The first approach, the holistic one, conceptually relates to template matching, and attempts to locate the eye using global representations. Characteristic of this approach belongs to connectionist methods such as principal component analysis (PCA) using eigenrepresentations[1]. Although location by matching raw images has been successful under limited circumstances, it suffers from the usual shortcomings of straightforward correlation-based approaches, such as sensitivity to eye orientation, size, variable lighting conditions, noise, etc. The second approach for eye detection extracts and measures local facial features, while standard pattern recognition techniques are then employed for locating the eyes using these measurements. Yuille *et al.* described a complex but generic strategy[2]. The characteristic of this approach is the concept of deformable templates. Lam *et al.* extended Yuille's method to extract eye features by using corner locations inside the eye windows which are obtained by means of average anthropometric measures after the head boundary is located[3]. Deformable template is an interesting concept, but it is difficult in terms of learning and implementation to use them.

Hausdorff distance was originally defined as a dissimilarity measure on data sets. It later got wide acceptance in image comparison. Huttenlocher *et al.* proposed a partial Hausdorff distance (PHD) method for object detection and recognition[4]. This method gets distance measure between the most closely matching portions of the images being compared which in turn reduces the effect of occlusion in object matching. Guo *et al.* proposed spatially weighted Hausdorff distance (SWHD) as an improvement to conventional Hausdorff distance between edge images[5].

All the above-mentioned methods for feature recognition use edge images for finding Hausdorff distance or its variants. However, the appearance is more important than the edge maps. The intensity distribution of pixels captures this appearance information. However, direct comparison of gray images is unsuitable because the performance will be affected by illumination variations. To overcome this shortcoming, infrared imaging technique[6], method of single training image per person[7], and local binary patterns (LPBs)[8]were proposed.

In this letter, by using the average instead of the maximum in the directed Hausdorff distance, a modified Hausdorff distance (MHD) based measure[9] for comparing the appearance of eyes is proposed, which is able to tolerate changes in eye shape, pose, and size. To eliminate the effects of the illumination variations, we propose an 8-neighbour-based transformation of the gray images. The transformed eye image is less sensitive to illumination changes while preserves the appearance information of eyes. All the located eyes are identified by back-propagation neural network (BPNN) identifier.

Primarily, the Hausdorff distance is defined as a distance measure between two data sets. This distance gives a measure of dissimilarity between these two sets. The conventional Hausdorff distance between the two sets $A = \{a_1, a_2, \cdots, a_m\}$ and $B = \{b_1, b_2, \cdots, b_n\}$ is given by

$$H(A,B) = \max[h(A,B), h(B,A)], \tag{1}$$

where

$$h(A,B) = \max_{a\in A}\min_{b\in B}\|a-b\| \tag{2}$$

and

$$h(B,A) = \max_{b\in B}\min_{a\in A}\|b-a\| \tag{3}$$

are the directed Hausdorff distances from A to B and from B to A, respectively, and $\|\cdot\|$ is the norm of a vector.

$H(A,B)$ takes the maximum of the directed distances from A to B and from B to A. When the Hausdorff distance is measured between two images, the data sets are replaced with pixel value sets.

Compared with the template matching, the Hausdorff distance measure does not require the explicit pair of elements in the two data sets. However, when one applies the Hausdorff distance for shape comparison directly, the mismatch will be very large due to the object missing or outliers such as noise. To overcome this, by using the average instead of the maximum in the directed Hausdorff distance, we use the MHD[9] measure:

$$h_{\mathrm{MHD}}\left(A,B\right)=\frac{1}{N_A}\sum_{a\in A}\min_{b\in B}\|a-b\|, \tag{4}$$

$$h_{\mathrm{MHD}}\left(B,A\right)=\frac{1}{N_B}\sum_{b\in B}\min_{a\in A}\|b-a\|, \tag{5}$$

where N_A and N_B are the element numbers of sets A and B, respectively.

To compare two gray images with MHD directly, its performance will be affected by illumination variations. So most Hausdorff distance based methods for object comparison use edge image. However, the appearance of the object is more important than the edge maps. The intensity distribution of pixels captures this appearance information. To eliminate the effect of illumination variation, an 8-neighbour-based transformation of gray images is proposed before the Hausdorff distance measure.

Firstly, let us consider a pixel and its 8 neighbours which form a 3×3 window, for instance, as shown in Fig. 1(a). Then this window is transformed into another window according to the difference between each element value and this pixel value (213 in Fig. 1(a)). If this difference is larger than zero, set the corresponding element value of the second window to be 2; if the difference is equal to zero, set the element value to be 1; and if it is smaller than zero, set the element value to be 0, as shown in Fig. 1(b). Rearranging the 8 elements of the second window as a vector, a ternary number is formed, as shown in Fig. 1(c). Finally, according to

$$d=\sum_{0}^{7}3^i t_i, \tag{6}$$

where t_i is the ternary value of the ith pixel and d is the decimal number, this ternary number is transformed into decimal number as the pixel value of the transformed images (for instance, in Fig. 1, this decimal number is 153).

The advantage of this 8-neighbour transformation is

198	200	201
213	213	214
244	197	196

(a)

0	0	0
1	213	2
2	0	0

(b)

0	0	0	1	2	2	0	0

(c)

Fig. 1. Sketch of the 8-neighbour transformation of gray image.

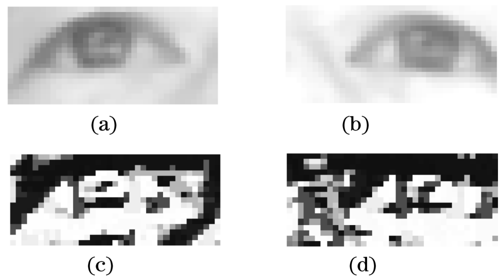


Fig. 2. 8-neighbour transformation results of eyes. (a),(b) Original eye images; (c) 8-neighbour transformation of (a); (d) 8-neighbour transformation of (b); (c) is used to calculate the MHD as a template eye in all the experiments of this letter.

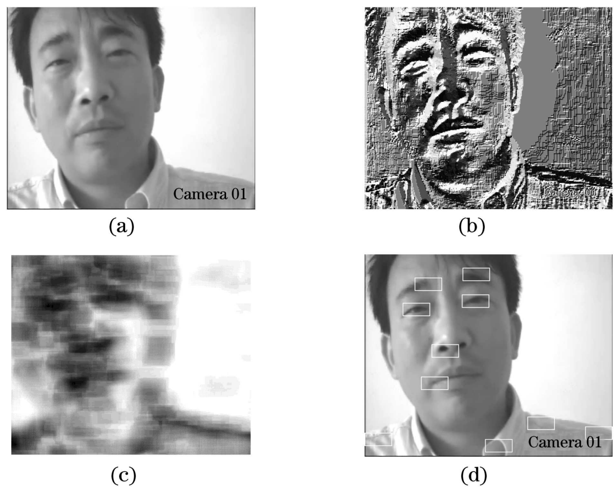


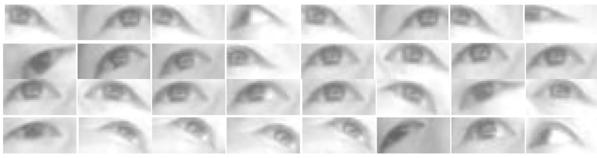
Fig. 3. MHD-based eye location process. (a) Original image; (b) transformed image of (a) with the 8-neighbour transformation; (c) MHD map between the transformed template eye and (b); (d) some selected candidates of eye (with the white rectangles).

that it eliminates the effects of the illumination variations, while remains the intensity distribution of the neighbours. Figure 2 illustrates the eye images before and after this transformation. We can see from Fig. 2 that the information of the intensity distribution is remained.

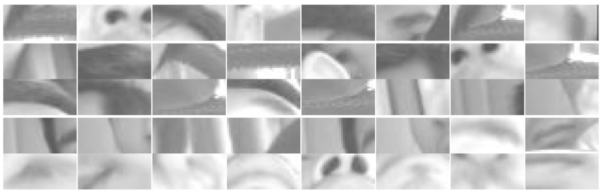
To localize an eye in a frame of image, we firstly transform this image and the template eye with the 8-neighbour transformation as mentioned above respectively, and then calculate the MHD between each patch of this transformed image and the transformed eye template image. It is not necessary that the size of each patch of this transformed image is the same as that of the transformed eye template image. Usually, several eye candidates with the smaller MHD are selected.

Figure 3 illustrates the locating process. It should be noticed that the template eye is not necessary the eye of the person in the image. As shown in Fig. 3(d), many candidates of eye have been selected with the above-mentioned MHD method. So this method is able to tolerate the slight shape, size changes, and the illumination variations.

To distinguish which one is the eye exactly, a common way is to use the geometric information of the eye. However, due to the scale and pose changes of the eye,



(a)



(b)

Fig. 4. (a) Positive and (b) negative training sets of eye image segmented with MHD method.

the geometrical method will be invalid. So we propose a BPNN classifier as eye identifier to automatically distinguish eyes from other candidates.

BPNN, as a type of multilayer perceptron[10] with error back propagation algorithm, has been applied in many pattern recognition tasks. In our application, the neural network includes an input layer, an output layer, and a hidden layer. Corresponding to the size of the candidate image ($39 \times 19 = 741$), the input layer consists of 741 input data elements and a bias element. The output layer includes one element, whose output represents the class that the input data belong to. One is for eye and minus one is for others. The activation function of this BPNN is a hyperbolic tangent function:

$$\varphi(x) = \tanh(x) = \frac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{\mathrm{e}^{x} + \mathrm{e}^{-x}}. \quad (7)$$

We collect the training images from five people with the same acquiring system, and then segment the eye candidate images with the MHD method mentioned above. Figures 4(a) and (b) display some images (with size $39 \times 19 = 741$) in the positive training set and the negative one, respectively. To make the convergence easy, before training the networks, we regularize the 8-bit data simply by

$$\hat{x} = \frac{2x}{255} - 1, \quad (8)$$

where x is the raw gray level and $\hat{x}$ is the regularized input value. Then the regularized matrix is rearranged as an input vector. Figure 5 shows the identified result of the frame in Fig. 3(d). Many false candidates have been removed effectively.

Usually, the identifying accuracy of BPNN is limited by the amount of training samples. For the misidentified sample, we should retrain the network repeatedly to obtain higher recognition accuracy.

To demonstrate the feasibility of our method, we used this algorithm on some videos. In the experiments, only one eye was identified, and the extension to two eyes identification was straightforward. Figure 6 shows the eye location results under different eye states and facial poses with our method. This eye location method tolerates these changes reliably. Figure 7 is for another

Fig. 5. Identified result of Fig. 3(d) with the BPNN identifier (with the white rectangles). Many false candidates are excluded.

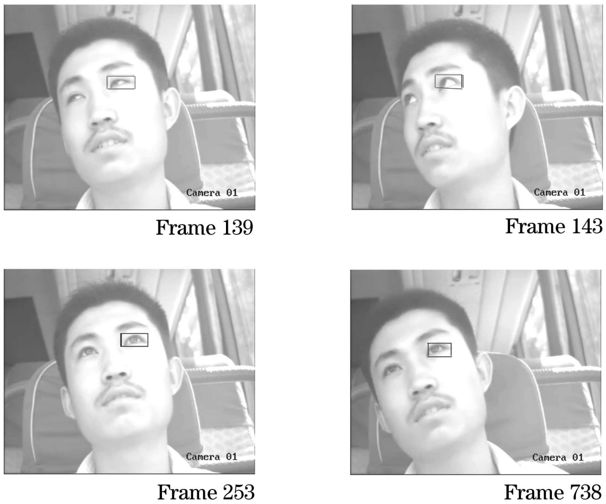


Fig. 6. Location results of 4 frames with different eye poses and head orientions (with the black rectangles).

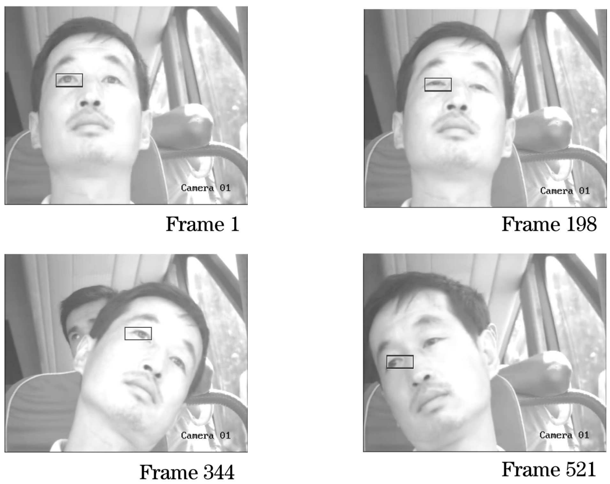


Fig. 7. Location results of 4 frames of another person with different eye poses and head orientions (with the black rectangles).

person who has different eye shapes, and the eye location method still works very well. Figure 8 illustrates the eye location results under different illuminations and different eye scales compared with Figs. 6 and 7. All experiments were operated with the same template eye

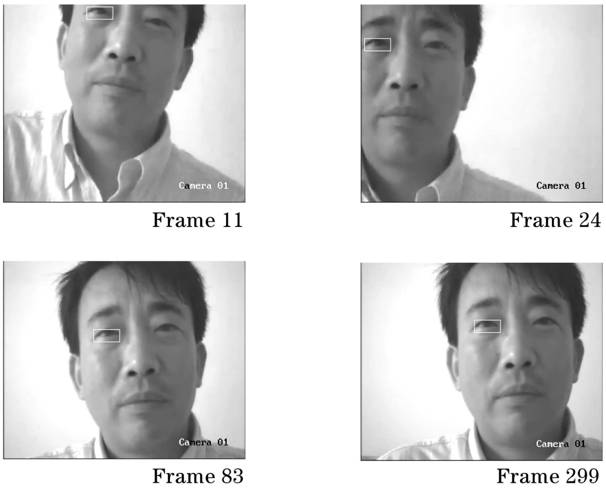


Fig. 8. Location results of 4 frames with different eye scales and illuminations in comparison with Figs. 6 and 7 (with the white rectangles).

shown in Fig. 2(b). From Figs. 6 – 8, we can see that our method is able to tolerate the changes in eye shape, scale, and pose very well. Meanwhile, it still works well under different illuminations.

In conclusion, the location method for eyes with MHD measure is able to tolerate the changes in eye shape, scale, and pose. With the 8-neighhour-transformation, this method is non-sensitive to the illumination variations. The combination of these two algorithms supplies an improved robustness to variable circumstance for eye location. The BPNN identifier distinguishes an eye from the candidates correctly. The location accuracy is enhanced considerably.

References

1. A. Pentland, B. Moghaddam, and T. Starner, in *Proceedings of IEEE Conference on Computer Vision & Pattern Recognition* 84 (1994).
2. A. L. Yuille, P. W. Hallinan, and D. S. Cohen, Int. J. Comput. Vis. **8,** 99 (1992).
3. K.-M. Lam and H. Yan, Pattern Recogn. **29,** 771 (1996).
4. D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, IEEE Trans. Pattern Anal. Mach. Intell. **15,** 850 (1993).
5. B. Guo, K.-M. Lam, K.-H. Lin, and W.-C. Siu, Pattern Recogn. Lett. **24,** 499 (2003).
6. X. Chen, Z. Jing, S. Sun, and G. Xiao, Chin. Opt. Lett. **2,** 694 (2004).
7. J. Li and J.-S. Pan, Chin. Opt. Lett. **6,** 255 (2008).
8. S. Marcel, Y. Rodriguez, and G. Heusch, International Journal of Image and Video Processing, Special Issue on Facial Image Processing 1 (2007).
9. K. J. Kirchberg, O. Jesorsky, and R. W. Frischholz, Lecture Notes in Computer Science **2359,** 103 (2002).
10. S. Haykin, *Neural Networks: A Comprehensive Foundation* (2nd edn.)(Prentice Hall, Englewood Cliffs, 1998).