

An Eye Location based Head Posture Recognition Method and Its Application in Mouse Operation

Zhe Chen¹, Bingbing Yang¹, Fuliang Yin¹

¹School of Information and Communication Engineering,
Dalian University of Technology, Dalian, 116024, China
[e-mail: {zhechen, flyin}@dlut.edu.cn]

*Corresponding author: Fuliang Yin

*Received August 27, 2014; revised November 17, 2014; revised December 30, 2014; accepted January 22, 2015;
published March 31, 2015*

Abstract

Abstract: An eye location based head posture recognition method is proposed in this paper. First, face is detected using skin color method, and eyebrow and eye areas are located based on gray gradient in face. Next, pupil circles are determined using edge detection circle method. Finally, head postures are recognized based on eye location information. The proposed method has high recognition precision and is robust for facial expressions and different head postures, and can be used in mouse operation. The experimental results reveal the validity of proposed method.

Keyword: head posture recognition, eye location, pupil correction, mouse operation, human computer interaction

1. Introduction

Human-Computer Interaction (HCI) has the intensive applications in some fields, such as computer, intelligent robot [1-3], mobile phone, consumer appliance [5-7], game [8-10], information retrieval [22-23], etc. Traditional HCI devices include mouse and keyboard. In recent few years, speech recognition and hand gestures have become gradually popular in some applications, however, in some occasions, speech recognition and hand gestures are limited. For example, speech recognition has poor performance in a car or aircraft due to background noise. It is difficult to apply hand gestures and speech recognition in the circumstance of high-altitude operations. In a surgery, doctor must use hands to take scalpel or other surgical instruments, and speech is not clear due to wear a mask, so the performances of speech recognition and hand gestures are poor in this circumstance. In addition, for language disorders or arm disabilities, speech recognition or hand gestures cannot be also used. Head and eyes are relatively stable organs, less affected by outside environment, such as background noise, face expression, and ambient environment, therefore, HCI techniques based on head posture and eye position have been developed in recent years.

Head pose estimation and recognition methods can be broadly divided into five categories: the template method, the appearance method, the classification methods, the regression methods, and the geometric methods. The template method [1-3] first creates a template for a head, then estimates and identifies the head pose; The appearance method [4, 5] usually assumes that there is a certain mapping between three- and two-dimensional face poses, and obtains the mapping relationship by training a large number of head or face images; The classification or regression methods [24-27] learn a mapping from the low dimensional manifold to the head angles, and then head postures are recognized using the mapping relation; The geometric methods [28-30] rely on the detection of facial features like corners of nose or mouth to estimate or recognize the head posture. However, these head posture recognition methods either are not robust to occlusion, scaling and other interferences, or require a heavy training burden. To cope with these problems, this paper tries to use eye location information to recognize head posture. Moreover, different from the depth information or multiple cameras-based head posture recognition method, the proposed method only uses one camera and 2D eye location information to recognize head posture.

Eyes are important feature organs in the head or face, and can play an important support role in head gesture recognition. Researches on eye location in head or face image have significant progress in recent years. In 2008, Bacivarov *et al* [12] used the appearance method to locate eye. The method extends the conventional statistical model of appearance to the component-based active appearance model (AAM) by combining a global model (the two eyes together) with sub-models (the two eyes separately), however, it needs a lot of training data, which results in a slow computational procedure. In 2010, Cristiani *et al* used inverse compositional algorithm (ICA) to build the best fitting active appearance model, then the AAM is used to detect and locate human eye [6]. In 2010, Kasinski *et al* used haar cascade classifiers augmented with some simple knowledge-based rules to locate human eye [9-11]. This method has low computational complexity, but it is easily interfered by eyebrows, face pose and noise. In 2012, Mehta *et al* used the eye shape and the template method to position eye [7, 8]. This method designs an eye generic template and then uses designed template to match eye in the eye search region, but it needs to design the different templates based on different illumination conditions and backgrounds, and the model could not adapt to different people. In 2013, Akinlar *et al* proposed an edge detection circle (ED-Circle) method [13] to

locate eyes, which has the low computational complexity and high location precision.

The existing head posture recognition methods are not robust to occlusion, scaling and other interferences, or require a heavy training burden. To remedy these problems, a head posture reorganization method based on eye location is proposed in this paper. Face is first detected by skin color model, and eyebrow and eye areas are located based on the characteristics of larger gray gradient in face image. Next, pupil circle is determined by edge detection circle (ED-Circle). Finally, head postures are recognized based on eye location.

The rest sections of this paper are organized as follows. Head posture recognition method based on eye location is proposed in Section II; Simulation results and analysis are given in Section III; Some conclusions are drawn in Section IV.

2. Head posture recognition method based on eye location

The existing head posture recognition methods are not robust to occlusion, scaling and other interferences, or require a heavy training burden. To resolve these problems, a head posture recognition method based on eye location is proposed. Block diagram of proposed head posture recognition scheme is shown in Fig. 1, which includes the five models: color correction, face detection, eyebrows and eye area location, pupil position, and head posture recognition, where color correction module corrects image color, eyebrow and eye area location module locates the eye region, pupil position module precisely locates eye, and the gesture recognition module identifies a head gesture based on eye location information.

Each module of the proposed method will be described in detail in following subsections.

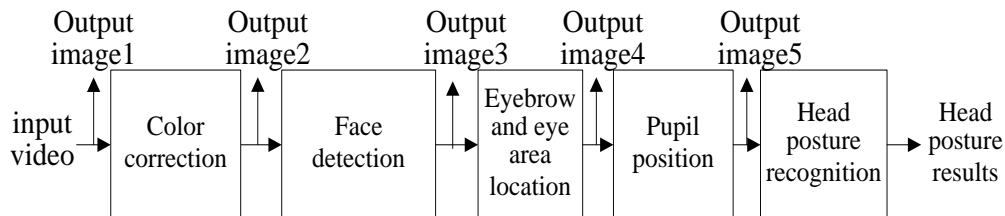


Fig. 1. Block diagram of head posture recognition system based on eye location

2.1 Color correction

Due to capture video images color cast, the performance of skin color-based face detection will degenerate. To remedy the problem, the raw image needs color correction. In proposed scheme, Gray World theory is used to correct image color.

Let $R(i,j)$, $G(i,j)$, and $B(i,j)$ denote the color values of Red, Green, and Blue of the pixel at (i, j) respectively; M and N are length and width of image respectively; R_a , G_a and B_a are the averages of R , G and B color components in image respectively. The conversion of RGB space of image is as follows

$$\begin{cases} R_a = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} R(i, j) \\ G_a = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} G(i, j) \\ B_a = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} B(i, j) \end{cases} \quad (1)$$

Let the gray world value be $V = R_a + G_a + B_a$. The values of $R(i, j)$, $G(i, j)$ and $B(i, j)$ are adjusted with the gray world value respectively

$$\begin{cases} R_1(i, j) = \frac{V}{R_a} R(i, j) \\ G_1(i, j) = \frac{V}{G_a} G(i, j) \\ B_1(i, j) = \frac{V}{B_a} B(i, j) \end{cases} \quad (2)$$

Let $f = \max(R_1(i, j), G_1(i, j), B_1(i, j))$. If $f > 255$, the $R_1(i, j)$, $G_1(i, j)$, and $B_1(i, j)$ should be scaled down

$$\begin{cases} R_2(i, j) = \frac{1}{f} R_1(i, j) \\ G_2(i, j) = \frac{1}{f} G_1(i, j) \\ B_2(i, j) = \frac{1}{f} B_1(i, j) \end{cases} \quad (3)$$

2.2 Face detection based on skin color model

In order to locate eye and estimate head posture, human face first needs to be detected and located. The appearance of skin color of human face is not easily affected by some factors such as facial expression, position and orientation, so a skin-color based face detection method is used.

To solve the color bias problem caused by colored light source, the values of R, G and B components of an input color image are adjusted by using the adjusted Y , C_b , C_g , and C_r components. The image from the RGB color space is converted into the $YC_bC_rC_g$ color space [14], i.e.

$$\begin{bmatrix} Y \\ C_r \\ C_b \\ C_g \end{bmatrix} = \begin{bmatrix} 0 \\ 128 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.289 & 0.578 & 0.114 \\ 0.500 & -0.4187 & -0.0813 \\ -0.1687 & 0.3313 & 0.500 \\ -0.316 & 0.500 & -0.84 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4)$$

To use ellipsoid skin model, the YCbCrCg color space is further transformed into the CbrCbgCgr color space [14], i.e.

$$\begin{cases} C_{br} = \frac{1}{2}(C_b + C_r) \\ C_{gr} = \frac{1}{2}(C_g + C_r) \\ C_{bg} = \frac{1}{2}(C_b + C_g) \end{cases} \quad (5)$$

Thus the ellipsoid skin model in space CbrCbgCgris the following [14]

$$l = \frac{(C_{br} - 135.5)^2}{12^2} + \frac{(C_{gr} - 132)^2}{6^2} + \frac{(C_{bg} - 120)^2}{10^2} \quad (6)$$

$$P(i, j) = \begin{cases} \frac{255}{2}l, & \text{if } l \leq 1 \\ 255, & \text{if } l > 1 \end{cases} \quad (7)$$

where l is the value of ellipsoid equation, and $P(i, j)$ is the pixel value. $P(i, j)$ is used to judge to the pixel whether is face pixel.

To trade off between the computational complexity and the detection accuracy, the fast face localization method based on skin color model [14] is adopted. The detailed information is described in the reference [14].

2.3 Position of eyebrows and eye area

In this paper, the gray-level gradient information is used to locate eyebrows and eye area from the face image. As eyes locate in the upper part of captured image, to reduce the computational load, the upper part of image only is traversed. Since eyebrow and eye areas have the feature of large gray gradient in face, this paper computes the gradient value of area B, and then moves area B by a certain step in area A to get area C, and computes the gradient value of area C. And then up to D. In Fig. 2, area B is total gray gradient value.

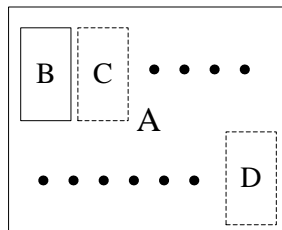


Fig. 2. Traversal diagram

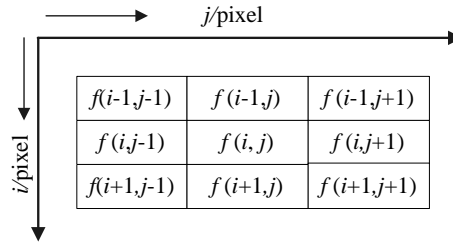


Fig. 3. Gray matrix of the image

The 3×3 gray-level matrix of an image at pixel (i, j) is shown in **Fig. 3**. The gradients at each direction at (i, j) are as follows

$$\begin{cases} G_{x1}(i, j) = f(i, j) - f(i-1, j) \\ G_{x2}(i, j) = f(i, j) - f(i+1, j) \\ G_{y1}(i, j) = f(i, j) - f(i, j-1) \\ G_{y2}(i, j) = f(i, j) - f(i, j+1) \\ G_{z1}(i, j) = f(i, j) - f(i-1, j-1) \\ G_{z2}(i, j) = f(i, j) - f(i-1, j+1) \\ G_{z3}(i, j) = f(i, j) - f(i+1, j-1) \\ G_{z4}(i, j) = f(i, j) - f(i+1, j+1) \end{cases} \quad (8)$$

where $f(i, j)$ is the gray value of image at point (i, j) , $G_{x1}(i, j)$, $G_{x2}(i, j)$, $G_{y1}(i, j)$, $G_{y2}(i, j)$, $G_{z1}(i, j)$, $G_{z2}(i, j)$, $G_{z3}(i, j)$ and $G_{z4}(i, j)$ are the gradient values in the x -direction, y direction, and 45° direction respectively. The average of absolute value of these eight gradients is computed as

$$avg(i, j) = \frac{1}{8} (|G_{x1}| + |G_{x2}| + |G_{y1}| + |G_{y2}| + |G_{z1}| + |G_{z2}| + |G_{z3}| + |G_{z4}|) \quad (9)$$

Selecting a sub-block k (such as B in **Fig. 2**) from the gradient matrix processed by Eq. (9), where size of sub-block is determined by image size, the total gray gradient value of sub-block k computed as

$$E = \sum_{j=1}^{M_1} \sum_{i=1}^{N_1} \log[1 + avg(i, j)] \quad (10)$$

where M_1 and N_1 are length and width of sub-block k , respectively.

The total gray gradient value of each sub-block is computed by the integral principle [16], and total gray gradient value of sub-block in the whole image is obtained by traversing image only one time, thus can reduce the computational load. Selecting some sub-blocks whose gray gradient values are larger than the others and merging these sub-blocks into a rectangle, a candidate of eyebrows and eye area can be obtained.

To further accelerate processing speed, the face area is downsampled [15], thus the positioning time of eyebrow and eye area can be reduced.

2.4 Pupil circle detection

2.4.1 Edge detection circle

To detect pupil circle, edge segments are first detected by edge drawing parameter free [17], and edge segments are converted into line segments [18]. The motivation of this step comes from the observation that any circular shape can be approximated by a consecutive set of lines, and these lines can easily be turned into circular arcs by a simple post-processing step. Then, the entire or part circles are detected by circular arc detection method. Finally, the candidate of circle are detected by arc join. The edge detection circle method is described in detail in reference [13].

2.4.2 Pupil circle discrimination

Usually, after processing of edge detection circle, there are many candidate circles, including pupil circles. It seems that pupil detection can be finished based on the fact that the total gray value of pupil circle is larger than others. In fact, the pupil circle can not be figured out only by this fact, because some circles in the eyebrow area also have large gray value. Thus in this paper, orientation of gradient is used to delete these eyebrow circles.

The pupil circle detection method is presented in the following:

After the eyebrow and eye area location, the output image 4 shown in Fig. 1 averages into two parts, $g(i_1, j_1)$ and $p(i_2, j_2)$; For the k -th candidate circle, the combination F_k of its gray value H_k and gray gradient value S_k are calculated as

$$F_k = (1 - w) \frac{H_k}{M_k} + w \frac{S_k}{N_k} \quad (11)$$

where w is the weight factor, $0 < w < 1$, N_k is the total number of pixels in the k -th circumference, S_k is the total gray gradient in the k -th circumference, M_k is the total number of pixels in the k -th inside circle, and H_k is the total gray value in the k -th inside circle.

Each candidate circle in the two parts $g(i_1, j_1)$ and $p(i_2, j_2)$ is substituted into Eq.(11). If the F_k value is the largest in all circles, the corresponding circle is identified as a pupil circle.

2.5 Head posture recognition method based on eye location

The current head posture recognition method mostly used the depth information or multiple cameras to recognize head posture, but these methods need expensive hardware devices or heavy training burden. In order to remedy these problems, the proposed method only uses one camera and 2D eye location information to recognize head posture.

Eyes are relatively stable organs in face, and can be used to recognize head posture. To estimate and recognize head posture, the geometry model of head posture recognition is designed as shown in Fig. 4. Seven head postures of of look at front, slant to left, slant to right, rise up, look down, gaze to right, and gaze to left are shown in Figs. 5(a)~(g).

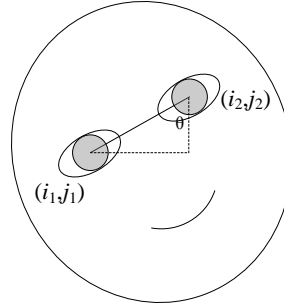


Fig. 4. Geometry structure of head model

If the judgment is not slant to left and right, the coordinates of left eye and right eye in the previous frame images is defined as (i_{pl}, j_{pl}) and (i_{pr}, j_{pr}) respectively, the coordinates of left eye and right eye in current frame images is defined as (i_{cl}, j_{cl}) and (i_{cr}, j_{cr}) respectively. Let size of face image be $P \times Q$ pixels.

(1) Recognitions of head slant to left and right

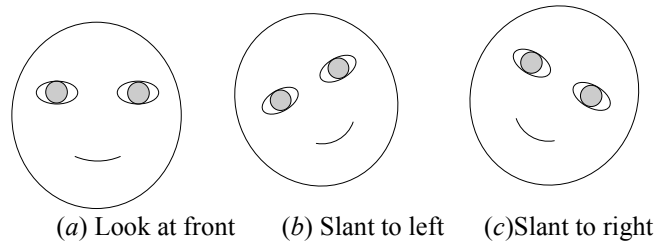
The geometry structures of head slant to the left or right are shown in **Figs. 5(b)** and **(c)**. After eyes are positioned, the Cartesian coordinates of left- and right eyes are available, including the left eye (i_{cl}, j_{cl}) and right eye (i_{cr}, j_{cr}) , ordering inclination $\theta = \arctan[(i_{cr} - i_{cl}) / (j_{cr} - j_{cl})]$. Assume that θ_1 , and θ_2 are slant thresholds which are used to judge head slants to the left or right, respectively. Thus if $\theta \in (\theta_1, \theta_2)$, head slants to the right; if $\theta \in (-\theta_2, -\theta_1)$, head slants to the left.

(2) Recognitions of head rise up and look down

The geometry structures of head rise up or look down are shown in **Figs. 5(d)** and **(e)**. If $(j_{cr} - j_{pr}) \geq T_1 P$ and $(j_{cl} - j_{pl}) \geq T_1 P$, the head posture is asserted to up; If $(j_{pr} - j_{cr}) \geq T_1 P$ and $(j_{pl} - j_{cl}) \geq T_1 P$, the head posture is asserted to down, where T_1 is the threshold.

(3) Recognitions of head gaze to left and right

The geometry structures of Gaze to right or left are shown in **Figs. 5(f)** and **(g)**. If $i_{cl} < i_{pl}$, $i_{cr} < i_{pr}$, $j_{cl} \approx j_{pl}$, $j_{cr} \approx j_{pr}$ and $(i_{cr} - i_{cl}) \leq T_2(i_{pr} - i_{pl})$, the orientation of eyes gaze is asserted to left; If $i_{cl} > i_{pl}$, $i_{cr} > i_{pr}$, $j_{cl} \approx j_{pl}$, $j_{cr} \approx j_{pr}$, and $(i_{cr} - i_{cl}) \leq T_2(i_{pr} - i_{pl})$, the orientation of eyes gaze is asserted to right, where T_2 is the threshold.



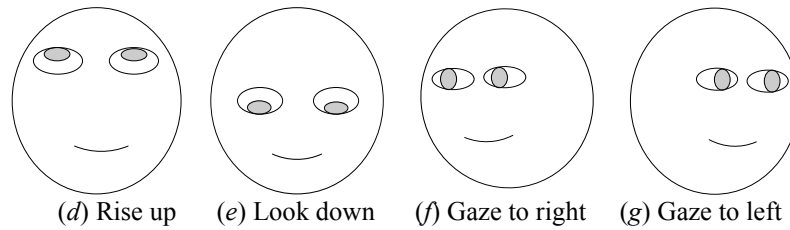


Fig. 5. Seven postures of head based on eye information

The proposed method can improve the head posture recognition precision, and is robust to occlusion, scaling and other interferences.

2.6 Seven postures of head and its application in mouse operation

Traditional mouse is an important HCI device, but as it needs arm to finish HCI operation, it has some limitations for user. Head postures can be used to simulate mouse operation. In this paper, the seven kinds head postures may define seven operations, i.e. UP, DOWN, LEFT, RIGHT, OK, CANCEL, and NO-OPERATION. The mapping relations between the head postures and the mouse operations are shown in **Table 1**. Even though the precision of the virtual mouse is not better than the traditional mouse, it is friendlier to users and easier to use than traditional mouse, and it is easy to realize interaction with computer when hands could not use in some reason, for example, the people whose arm is disable can operate simply computer by the virtual mouse.

Table 1. Mapping relations between head postures and mouse operations

Head posture	Mouse operation
Rise up	UP
Look down	DOWN
Gazes to left	LEFT
Gazes to right	RIGHT
Tilt to the left	OK
Tilt to right	CANCEL
The other gestures	NO-OPERATION

3. Simulations and Result Discussions

To verify the effectiveness of the proposed method, some experiments are carried out. In simulations, 120 images from IMM-FACE Database, issued by the Technical University of Denmark [19], are used. The size of each images is 640×480 pixels, and these images are front faces and have not occlusion in eyes and head. Self made with simple and complex background face databases are used. The size of images self made with simple background face database is 640×480 pixels, and size of images self made with complex background face database is 640×480 pixels and 760×460 pixels, respectively. The images from simple background face database are also front faces and have not occlusion in eyes and head. There are 600 face images in the complex background face database, including wearing glass, head covering and normal, and the number of each category is 200.

3.1 The experiments of pupil circle detection

Experiments of pupil circle detection are carried out under diffuse light, and simple and complex background conditions. The experiment of pupil circle detection with different weight w is shown in Fig. 6. It is observed from Fig. 6 that when w is approximately 0.5, the pupil detection performance is the best, so the value of w is set as 0.5 in following experiments. If the w value is not equal to 0.5, the pupil circle may be falsely located in eyebrow area.

Experiments of head posture recognition with different θ , T_1 , and T_2 are shown in Figs. 7(a)~(c). It is observed from Figs. 7(a)~(c) that when $\theta \in (50^\circ, 85^\circ)$, $T_1 \in (0.04, 0.1)$, and $T_2 \in (0.75, 0.95)$, the head posture recognition performance is the best. Considering that a person head has a certain extent tilt, the values of θ_1 , θ_2 , T_1 , and T_2 are set as 50° , 75° , 0.07, and 0.8 respectively.

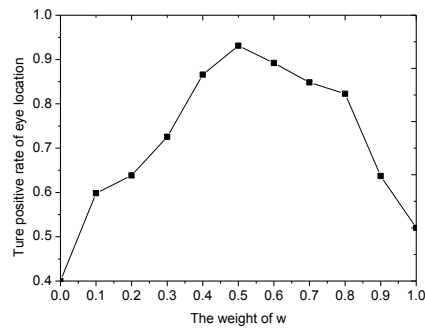
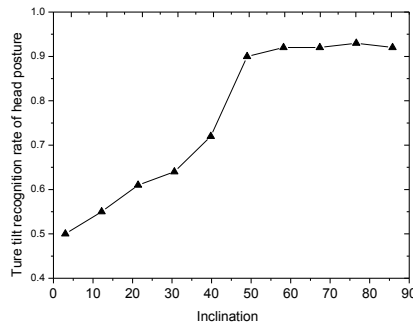
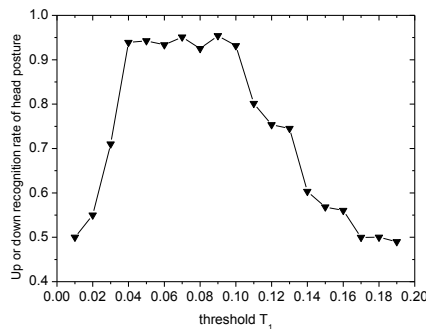


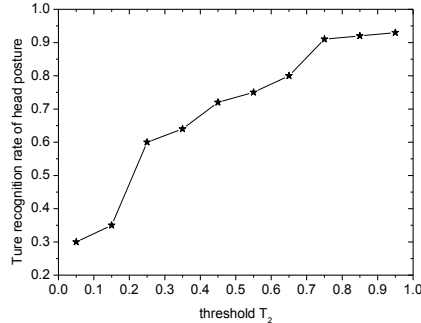
Fig. 6. Detect performance of eye location with different weight w



(a) Tilt recognition performance of head posture with different θ



(b) Recognition performance of head up or down postures with different threshold T_1

(c) Recognition performance of head Gaze posture with different threshold T_2 **Fig. 7.** Head posture recognition performance with different weight (w) and thresholds (θ , T_1, T_2)

3.2 The experiments of the human eye location

3.2.1 Experiments result of human face detection and location

In order to illustrate effectiveness of the proposed method, the results of face detection and location of Ref. [14] method are described in Table 2.

Table 2. Results of face detection and location of Ref. [14] method

Face database	Image numbers	Ref. [14] method	
		detection	location
IMM	120	99.2%	99.2%
Self-made wearing transparent glass in complex background	200	99.0%	98.0%
Self-made head covering in complex background	200	98.0%	97.5%
Self-made normal face in complex background	200	99.0%	98.3%
Self-made in simple background	330	99.4%	98.5%
Total	1050	99.1%	98.3%

Form the Table 2, the precision of face location of Ref.[14] method is high, and is enough for eye location and other operations, such as, face recognition, expression analysis, and so on. So the face location method is robust to different backgrounds, different genders, occlusion, etc.

3.2.2 Experiments of the human eye location

To illustrate the performance of the proposed method, the template method [7], the projection function method [9], and the appearance method [12] are also implemented, and are compared with proposed method, and Quantitative method is used to verify effectiveness of the proposed method,

For performance comparison, a quantitative method is defined as [20]

$$e = \frac{\max(d_l, d_r)}{d} \quad (12)$$

where d_l and d_r are the Euclidean distances between the found left and right eye centers. d is the Euclidean distance between the eyes in the ground truth, and $\max(\cdot)$ is the finding

maximum function.

Table 3. Comparison of normalized error scores of different method in different databases

Method	Accuracy of eye location		
	$e < 0.05$	$e < 0.10$	$e < 0.25$
Proposed method	90.3%	92.6%	98.3%
Ref.[7] method	85.2%	87.6%	97.2%
Ref.[9] method	83.4%	85.4%	96.0%
Ref.[12] method	80.7%	85.2%	94.7%

Table 4. Eye location of different methods in different face databases

Face database	Image numbers	Accuracy of eye location			
		Proposed method	Ref.[7] method	Ref.[9] method	Ref.[12] method
IMM	120	95%	90.0%	90.0%	88.3%
Self-made wearing transparent glass in complex background	200	91.0%	85.0%	79.0%	81.0%
Self-made head covering in complex background	200	92.0%	86.0%	84.0%	82.0%
Self-made normal face in complex background	200	94.0%	88.0%	88.0%	87.0%
Self-made in simple background	330	90.8%	89.2%	85.8%	87.5%
Total	1050	92.56%	87.64%	85.36%	85.16%

Eye location result in different face databases, when $e < 0.1$

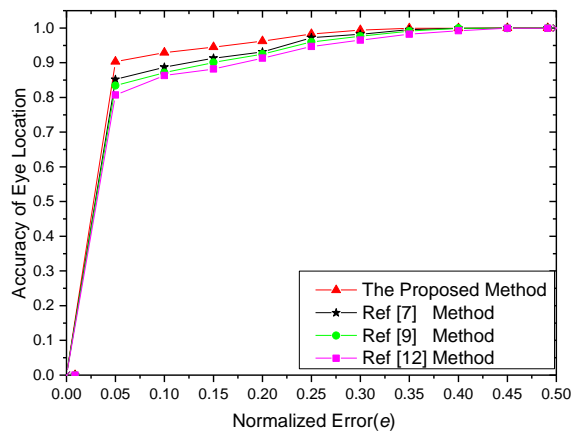


Fig. 8. Comparison results of normalized error of different methods

The experiments of human eye location are carried out under diffuse light, and simple and complex background conditions. The results of eye location of different methods in different databases are shown in Table 3 and Table 4 and Fig. 8. It is observed from Tab.3 that compared with Refs [7, 9, 12], the performance of the Ref. [7] method is obviously better than the Ref. [9] method and the Ref. [12] method, while the proposed method has more high position accuracy than the Ref. [7] method, that is, the performance of the proposed method is the best among these human eye location methods. Figure 8 also shows that the proposed method has the best location performance, no matter for high-precision localization ($e < 0.05$),

detection within the iris area ($e < 0.1$), and coarse detection of the eye area ($e < 0.25$).

When $e < 0.1$, the eye location results in different face databases are shown in Table 4. Simulation results of the proposed method for three face databases are shown in Figs. 9 and 10, where the green rectangle in Fig. 9 represents eyebrows and eye area; Fig. 9 shows the location results of eyebrow and eye region, and Fig. 10 shows the final eye location results. From Fig. 10, it is observed that the proposed method is robust to the head gesture (i.e. Figs. 10(c), 10(d), 10(f), 10(k), and 10(l)), facial expression (i.e. Fig. 10(b)), hair shelter (i.e. Fig. 10(g), 10(h), and 10(j)), mustache (i.e. Figs. 10(a), 10(c), and 10(e)). Fig. 10(g) and Fig. 10(h) show the eye location result of same person with different image size respectively. The simulation results in Fig. 10(g) and Fig. 10(h) show the proposed method is robust to image scaling; Figs. 10(i), 10(n), 10(o) show that the proposed method is robust to glasses. From Fig. 10(i), 10(n), the proposed method successfully locates eye, even though eye is partly occluded, so Figs. 10(i), and 10(n) shows that the proposed method is robust to occlusion. Fig. 10(o) shows that the proposed method is robust to wear hat. The Ref.[7] method has well location result with different temples, however, many temples must be designed based on different condition, i.e., it is not robust to different people or backgrounds. The Refs [9] method and the [12] method have bad eye location results in female images, especially if female has long hair in forehead.



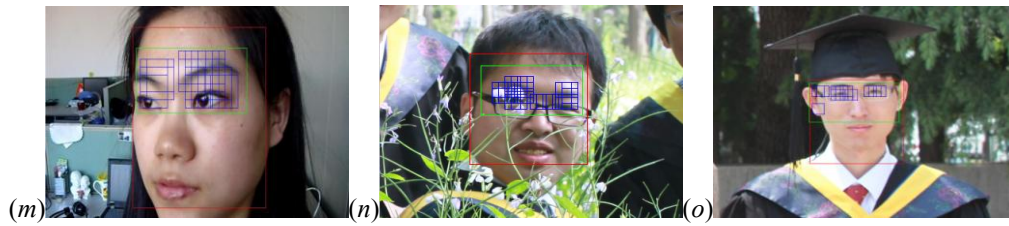


Fig. 9. Detect results of face, eyebrow and eye area: (a)~(f) simulation results of the IMM face database, (g)~(h) Self-made with simple background face database, and (i)~(o) self-made with complex background face database

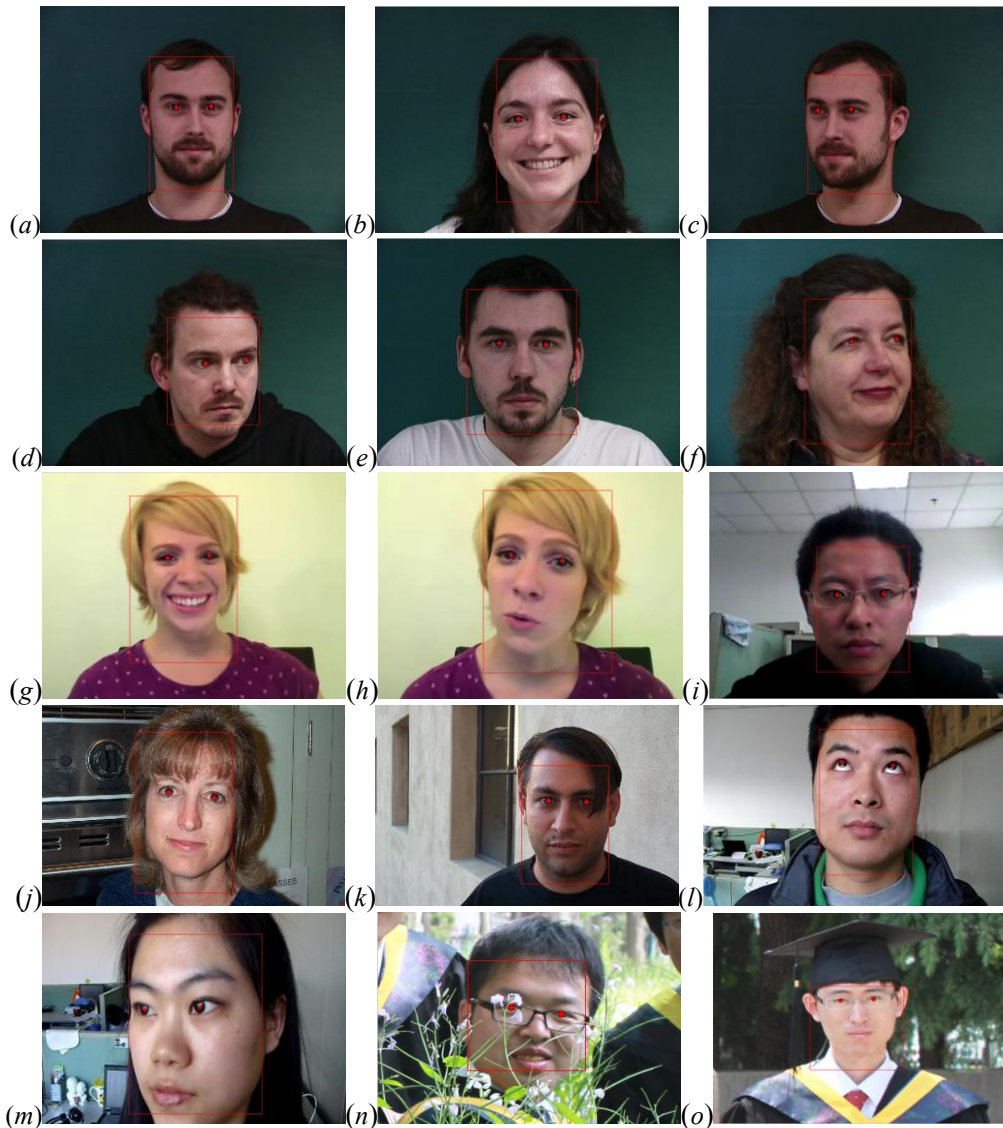


Fig. 10. Results of eye location: (a)~(f): IMM face database; (a) simulation result of a bearded man; (b) simulation result of a woman; (c) simulation result of a man whose eyes gaze to left; (d) simulation of a man whose eyes gaze to right; (e) simulation result of a man whose eyes gaze to right; (f) simulation result of a woman whose eyes gaze to right. (g)~(h): self-made with simple background face database; (g)~(h) simulation result of a woman; (i)~(o): self-made with complex background face database; (i) simulation result of man who wear glass; (j) simulation result of woman having long hair; (k) simulation

result of man whose hair occludes partly eye; (*l*) simulation result of a man whose head and eyes were up; (*m*) simulation result of a woman whose head or eyes gaze to right; (*n*) simulation result of a man whose eye is partly occlude; (*o*) simulation result of a man wearing glass and hat

3.3 Recognitions of head postures in mouse operation

Recognition experiments of seven head postures are carried out under the conditions with diffuse light, simple and complex backgrounds, and one people in video image. In order to verify the effectiveness of the proposed method, the three performance measures [21] are used.

The miss rate of head postures is defined as

$$M_r = \frac{m_r}{P_s + N_s} \quad (13)$$

The false positives rate of head postures is defined as

$$F_r = \frac{f_r}{N_s + P_s} \quad (14)$$

The true recognition rate of head postures is defined as

$$T_r = \frac{s}{P_s} \quad (15)$$

where m_r is the number of false negatives, f_r is the number of false positives, s is the number of true recognition head posture, P_s is the number of positive sample, and N_s is the number of negative sample; positive sample is these images including eyes and head, negative sample is these images including no eyes and head.

The recognition results of seven head postures are shown in Table 5. Each head posture in Table 5 includes 150 positive sample images and 100 negative sample images, and total test images include 1050 positive sample images and 700 negative sample images.

From Table 5, we can see that the Ref.[7] method has better recognition performance than the Ref.[9] method and the Ref.[12] method, while the recognition performance of the proposed method is better than that of the Ref.[7] method, i.e., the proposed method not only has good recognition precision, but also has low miss rate and false positive rate, this is, the performance of proposed method is stable. It is observed from Table 5 that when a head is down, the recognition precision for DOWN is a little low, because the pupil is easily obscured by eyelid, which leads to eye location fail and then results in a low head postures recognition precision

Table 5. Statistical recognition results of seven head postures for different methods

Head postures	Proposed method			Ref.[7] method			Ref.[9] method			Ref.[12] method		
	T_r	M_r	F_r	T_r	M_r	F_r	T_r	M_r	F_r	T_r	M_r	F_r
UP	95.3%	2.8%	1.2%	84.6%	9.2%	1.2%	90.0%	6.0%	1.2%	90.0%	6.0%	1.6%
DOWN	84.0%	9.6%	1.6%	84.6%	9.2%	3.2%	80.0%	12.0%	4.0%	85.3%	8.8%	3.2%
LEFT	97.3%	1.6%	0.8%	88.0%	7.2%	2.4%	85.3%	8.8%	1.6%	87.3%	7.6%	2.4%
RIGHT	98.0%	1.2%	1.6%	87.3%	7.6%	1.2%	82.0%	10.8%	1.6%	84.6%	9.2%	2.4%
OK	100 %	0.0%	0.8%	92.0%	4.8%	0.8%	80.0%	12.0%	0.8%	88.0%	7.2%	1.2%
CANCEL	96.7%	2.0%	0.8%	90.0%	6.0%	0.8%	84.6%	9.2%	1.2%	87.3%	7.6%	0.8%
NO OPERATION	96.0%	2.4%	1.2%	92.0%	4.8%	1.2%	90.0%	6.0%	1.2%	92.0%	4.8%	2.0%
Total	95.3%	2.8%	1.2%	88.4%	7.2%	1.5%	84.7%	9.2%	1.6%	88.0%	7.2%	2.0%

T_r : true recognition rate, M_r : miss rate, and F_r : false positive rate

4. Conclusions

In this paper, a head posture recognition method based on eye location is proposed. Eyes are first located using the larger gradient and gray values, and then head postures are recognized based on eye position information. Finally, the proposed method is used to emulate typical mouse operations. Simulation results show that the proposed method has high recognition precision, low miss rate and false positive rate. The proposed method can be used in human computer interaction applications

5. Acknowledgement

This work was supported by the National Natural Science Foundation of China (61172110, 61172107); Specialized Research Fund for the Doctoral Program of Higher Education of China (200801410015); Major projects in Liaoning province science and technology innovation(201302001); The Fundamental Research Funds for the Central Universities of China (DUT13LAB06).

References

- [1] Ma. B. P, Yang. X, Shan. S. G, "Head Pose Estimation Via Background Removal," in *Proc. of 5th Int. Workshop on Intelligent Science and Intelligent Data Engineering*, pp.140-147, Oct. 2013. [Article \(CrossRef Link\)](#).
- [2] Murphy C. E, Trivedi M M, "Head pose estimation and augmented reality tracking: an integrated system and evaluation for monitoring driver awareness," *IEEE Trans. on Intelligent Transportation Systems*, vol.11, no.2, pp. 300-311, Jun. 2010. [Article \(CrossRef Link\)](#).
- [3] Cai, Q, Sankaranarayanan. A, Zhang. Q, Zhang. Z, Liu. Z, "Real time head pose tracking from multiple cameras with a generic model," in *Proc. of 2010 IEEE Computer Society Conf. on Vision and Pattern Recognition Workshops*, pp. 25-32, Jun. 2010. [Article \(CrossRef Link\)](#).
- [4] Huang. D, Storer. M, DeLaTorre. F, Bischof. H, "Supervised local subspace learning for continuous head pose estimation," in *Proc. of 2011 IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2921-2928, Jun. 2011. [Article \(CrossRef Link\)](#).
- [5] Fu. Y, Huang. T. S, "Graph embedded analysis for head pose estimation." in *Proc. of 7th Int. Conf. on Automatic Face and Gesture Recognition*, pp. 6-8, Apr. 2006. [Article \(CrossRef Link\)](#).
- [6] Cristiani. A, Porta. M, Gandini. D, Bertolotti. G. M, Serbedzija. N, "Driver drowsiness identification by means of passive techniques for eye detection and tracking," in *Proc. of IEEE 4th Int. Conf. on Self-Adaptive and Self-Organizing Systems Workshop*, pp. 142-146, Sep. 2010. [Article \(CrossRef Link\)](#).
- [7] Mehta. R, Shrivastava. M, "Real time eye template detection and tracking," *Int. Journal of Advanced Computer Research*, vol.2, no.2, pp. 18-22, Jun. 2012. http://www.theaccents.org/ijacr/papers/current_jun_2012/3.pdf
- [8] Choi. J. K, Chung. S. T, "Reliable and fast eye detection," in *Proc. of 9th Int. Conf. on Signal Processing, Computational Geometry and Artificial Vision*, pp. 78-82, 2009. <http://dl.acm.org/citation.cfm?id=1627548>
- [9] Lu. L, Yang. Y, Wang. L, Tang. B, "Eye location based on gray projection," in *Proc. of IEEE 3th Int. Symposium on Intelligent Information Technology Application*, pp.58-60, Nov. 2009. [Article \(CrossRef Link\)](#).
- [10] Yan. B, Zhang. X, Gao. L, "Improvement on pupil positioning algorithm in eye tracking technique," in *Proc. of IEEE Int. Conf.on Information Engineering and Computer Science*, pp. 1- 4, Dec. 2009. [Article \(CrossRef Link\)](#).
- [11] Kasinski. A, Schmidt. A, "The architecture and performance of the face and eyes detection system based on the Haar cascade classifiers," *Pattern Analysis & Applications*, vol.13, no.2, pp. 197-211, May, 2010. [Article \(CrossRef Link\)](#).

- [12] Bacivarov. I, Ionita. M, Corcoran. P, "Statistical models of appearance for eye tracking and eye-blink detection and measurement," *IEEE Trans. on Consumer Electronics*, vol.54, no.3, pp.1312-1320, Aug. 2008. [Article \(CrossRef Link\)](#).
- [13] Akinlar C, Topal C, "ED-Circles: A real-time circle detector with a false detection control," *Pattern Recognition*, vol.46, no.3, pp.725-740, March, 2013. [Article \(CrossRef Link\)](#).
- [14] Li W, Jiao F, He C, "Face detection using ellipsoid skin model," *Emerging Technologies for Information Systems, Computing, and Management, Lecture Notes in Electrical Engineering*, vol 236, pp.513-521, 2013. [Article \(CrossRef Link\)](#).
- [15] Burt. P, Adelson. E, "The Laplacian pyramid as a compact image code," *IEEE Trans. on Communications*, vol.31, no.4, pp. 532-540, Apr. 1983. [Article \(CrossRef Link\)](#).
- [16] Viola. P, Jones. M, "Rapid object detection using aBoosted cascade of simple features," in *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 511-518, Dec. 2001. [Article \(CrossRef Link\)](#).
- [17] Akinlar C, Topal C, "EDPF: A real-time parameter-free edge segment detector with a false detection control," *Int. Journal of Pattern Recognition and Artificial Intelligence*, vol. 26, no.1, pp.1-22, May 2012. [Article \(CrossRef Link\)](#).
- [18] Akinlar C, Topal C, "EDLines: A real-time line segment detector with a false detection control," *Pattern Recognition Letters*, vol.32, no.13, pp.1633-1642, Oct. 2011. [Article \(CrossRef Link\)](#).
- [19] Nordstrøm M M, Larsen M, Sierakowski J, et al, "The IMM face database-an annotated dataset of 240 face images," *Technical University of Denmark, DTU Informatics, Building 321*, 2004. http://orbit.dtu.dk/fedora/objects/orbit:79272/datastreams/file_3069003/content
- [20] Jesorsky O, Kirchberg K J, Frischholz R W, "Robust face detection using the hausdorff distance," in *Proc. of 3th Int. Conf. Audio-and Video-based Biometric Person Authentication*, pp.90-95, Jun. 2001. [Article \(CrossRef Link\)](#).
- [21] Dalal N, Triggs B, "Histograms of oriented gradients for human detection," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 886-893, Jun. 2005. [Article \(CrossRef Link\)](#).
- [22] Li Z, Liu J, Zhu X, "Multi-modal multi-correlation person-centric news retrieval," in *Proc. of 19th ACM Int. Conf. on Information and Knowledge Management*, pp.179-188, Oct. 2010. [Article \(CrossRef Link\)](#).
- [23] Li Z, Liu J, Lu H, "Sparse constraint nearest neighbour selection in cross-media retrieval," in *Proc. of IEEE Int. Conference on Image Processing*, pp.1465-1468, Sep. 2010. [Article \(CrossRef Link\)](#).
- [24] Oyini Mbouna R, Kong S G, Chun M G, "Visual analysis of eye state and head pose for driver alertness monitoring," *IEEE Transactions on Intelligent Transportation Systems*, vol.14, no.3, pp. 1462-1469, Sep. 2013. [Article \(CrossRef Link\)](#).
- [25] Huang D, Storer M, De la Torre F, "Supervised local subspace learning for continuous head pose estimation," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921-2928, Jun. 2011. [Article \(CrossRef Link\)](#).
- [26] Fanelli G, Gall J, Van Gool L, "Real time head pose estimation with random regression forests," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 617-624, Jun. 2011. [Article \(CrossRef Link\)](#).
- [27] Yan Y, Subramanian R, Ricci, "Evaluating multi-task learning for multi-view head pose classification in interactive environments," in *Proc. of IEEE Int. Conference on Pattern Recognition*, pp.4182-4187, Aug. 2014. [Article \(CrossRef Link\)](#).
- [28] Valenti R, Sebe N, Gevers T, "Combining head pose and eye location information for gaze estimation," *IEEE Transactions on Image Processing*, vol.21, no.2, pp.802-815, Jan. 2012. [Article \(CrossRef Link\)](#).
- [29] Zhang W, Wang Z, Xu J, "A Method of gaze direction estimation considering head posture," *Int. Journal of Signal Processing, Image Processing and Pattern Recognition*, vol.6, no.2, pp.103-111, Apr. 2013. http://www.sersc.org/journals/IJSIP/vol6_no2/9.pdf
- [30] Zhang Y, Yue J, Dong Y, "Robust eye openness and head pose estimation based on multi modal camera," in *Proc. of IEEE Workshop on Advanced Research and Technology in Industry Applications*, pp.418-422, Sep. 2014. [Article \(CrossRef Link\)](#).



Prof. Zhe Chen received his B.S degree in electronics engineering, the M.S degree in signal and information processing, and the PhD. Degree in signal and information processing from Dalian university of Technology (DUT), Dalian, China, in 1996, 1999 and 2003 respectively. He joined the Department of electronic Engineering, DUT, as a lecturer in 2002, and become an Associate professor in 2006. His researches include speech processing, image processing and wideband wireless communication.



Bingbing Yang is a candidate of M.S degree at School of Communication and information engineering, Dalian University of Technology, Dalian, China. He received his B.S degree in electronics and information engineering from School of Communication and information engineering, Dalian University of Technology, Dalian, China. His main researches include image processing, pattern recognition and computer vision.



Prof. Fuliang Yin was born in Fushun city, Liaoning province, China, in 1962. He received B.S. degree in electronic engineering and M.S. degree in communications and electronic systems from Dalian University of Technology (DUT), Dalian, China, in 1984 and 1987, respectively. He joined the Department of Electronic Engineering, DUT, as a Lecturer in 1987 and became an Associate Professor in 1991. He has been a Professor at DUT since 1994, and the Dean of the School of Electronic and Information Engineering of DUT from 2000 to 2009. His research interests include digital signal processing, speech processing, image processing and broadband wireless communication.