Driver Gaze Tracker Using Deformable Template Matching

Wei Wang
Department of Optical-Electrical and Computer
University of Shanghai for Science and Technology
ShangHai, China
lvinni@126.com

Yingping Huang
Department of Optical-Electrical and Computer
University of Shanghai for Science and Technology
ShangHai, China
emlyghg@hotmail.com

Renjie Zhang
Department of Optical-Electrical and Computer
University of Shanghai for Science and Technology
ShangHai, China
zhangrj@usst.edu.cn

***Abstract*—In driver assistance system, human eye gaze direction is an important feature described some driver's situation such as distraction and fatigue. This paper proposes a method to track driver's gaze direction by using deformable template matching. The method is divided into three steps: first, identifying the face area. Second,localizing the eye area. Finally, combining the eye region model and sight algorithm to determine the driver's gaze direction. Experimental results show that this method can effectively identify human's eye and track gaze motion in a real-time running basis.**

Keywords-deformable template;face detection;eye detection; gaze direction; eye regional model

I. Introduction

In a variety of dynamic monitoring of the human eye, you can use electronic means and image processing means to tracking the changes of human eye. For electronic means, it can be divided into contact and non-contact [1]. Contact means need to wear appropriate receive device on driver's face like wearing glasses, and the equipment can turn driver's feel into some elements that we can analysis and deal with, such as voltage, current, etc. Non-contact means needn't direct contact with driver, using another way to measure eye movements, for example, we can use infrared light reflection for pupil to measure the degree of eye rotation. For image processing means, you can access eye image through camera, and track eye movement using a series of image processing method.

In general, the human eye for positioning and tracking algorithm is mainly divided into two categories: one is directly tracking the eye in a background image. Another is to find face region in background image firstly, and then find eye region in the face image previously we got [2, 3, 4]. This paper proposes a successive algorithm which combination of face detection, eye recognition using deformable template and a method what calculate line of sight angle of driver's eye. This paper will introduce this algorithm in next chapters, so we just use some simple word to explain it in this chapter. First, using AdaBoost cascade classifier to train Haar features and extract human face regions from initial image, and mark it [5, 6]. Then using deformable template method to get the precise location, shape of eye and get the position, size of pupil and other information. Further, according the information received after template match, to determine gaze direction using coordinate system and calculates formula this paper proposed.

II. Face Recognition Based On Adaboost Clissifier

A. AdaBoost Theory

Classifier is a mathematical model through dividing image feature points to constitute some set of points with simailar characteristics. AdaBoost is an adaptive classifier with multiple features; it is constantly updated by the sample weight to get the best clustering. In this paper, select Haar features of image as input feature points, after training and superposition can get a strong classifier. This classification has high precision and good versatility.

B. Implement face recognition using OpenCV

In this paper, using sophisticated computer vision library (OpenCV) for face recognition. Implementation is divided into two parts.

1) *Sample training:* Training needs the following work.

a) Sample creating. Here need two types of samples as input source:target sample image and non-target samples image, it can be downloaded from the MIT Laboratory.

b) Convert the sample format. OpenCV Trainer need input file format such as ".vec" format, so the next step is using program(CreateSample.exe) to convert file format.

c) Training. Call the function "haarTraining" training and save results to XML file.

2) *Face recogntion.* By calling the trained classifier to find the face, and then mark this region. The test results shown in figure 1. We can see that wearing glasses and not wearing glasses test results had no effect on the human face detection, Use this method of face detection have high accuracy and robustness.

For better observing examination effect, make a statistics to the number that the system read and handle images in a period of time and the rate of miss detection. Drive the result shown as table 1. It can be seen from the test data, using AdaBoost cascade classifier method to detect human face, have average detection rate of 6.6(frame)/s, recogition rate more than 95% . it a good detection result.

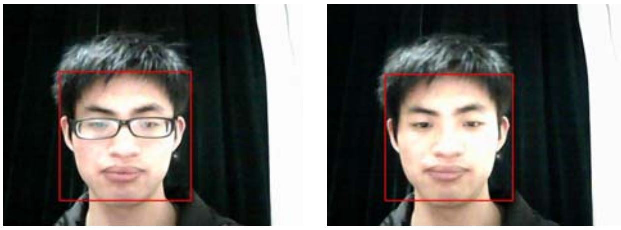

Figure 1. Example of face detection.The left image is wearing glasses, the right image is not wearing glasses.

TABLE I. TEST RESULT FOR FACE DETECTION

Number of Image frame	The rate of miss detection	Times（s）
80	2.5%	11.4
150	4.0%	25.0
180	4.2%	27.1
200	5.3%	29.7
212	4.8%	31.8

III. DEDORMABLE TEMPLATE

Deformable template is a dynamic method applied to image recognition, which describes some obvious characterstics in the image by defining the energy function, such as image peak, valley, edge, brightness and so on. The basic idea is very simple, that is changing the prototype to generate variants, and then choose the best variant to match the object be identified in one image. Yuille proposed a deformable eye template that combines the geometrical feature and gray feature [7], it be binded by energy formnula to implement the best match effect. The eye model is showned in figure 2.

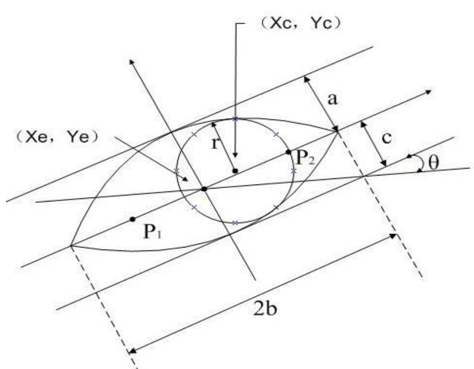

Figure 2. Defoemable template for a human eye.

As shown: (X_e, Y_e) is the center of the eye, (X_c, Y_c) is the center of the pupil, a and c is given by two parabolic sections representing the upper and lower parts of the boundary, P_1 and P_2 is the focus of the parabola. Template binding energy formula is:

$$E_c = E_v + E_e + E_i + E_p + E_{prior} \tag{1}$$

In this formula, E_v is the energy valley within the pupil, E_e is the edge energy for pupil and eye, E_i energy makes brightness of the pupil smallest and brightness between the edge of eye and the edge of pupil largest, E_p response the peak of image ,which are assessed by P_1 and P_2, E_{prior} response the image prior energy. To get the best results, it must be given a reasonable definition of the energy equation initial value. This paper presents the initial values as follows:

TABLE II. INITIALIZATION OF THE TEMPPLATE WEIGHTS

c_1	c_2	c_3	c_4	c_5	c_6	k_1	k_2	k_3	k_4
100	100	0	0	400	0	1	4	4	0.1

IV. EYE REGION MODEL AND GAZE TRACKING

A. Eye region model

Before detected in the line of sight, you need to know the gaze direction of each mark corresponds to what is meaningful or meaningless, for example: When the driver looked at the window in front of traffic, cars and rearview mirror et al. the line of sight is meaningful. When driver see the seat or the tp of car's window et al. the line of sight is meaningless. This viewing area will be divided into five meaningful regions: up, down, left, right, and front, expressed as a simple form in Figure 3.

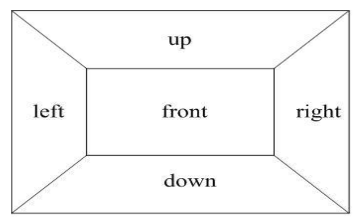

Figure 3. Simple division model.

Real life, a reasonable division of gaze region, it need different vehicles and people to calculate the angle of their gaze regional scope. Assuming d_1 is the vertical distance between driver's eye and car's window, d_2 is the shortest distance between driver's eye and the top of car's window, d_3 is the maximum distance between driver's eye and left rearview mirror, d_4 is the maximum distance between driver's eye and right rearview mirror, d_5, d_6 and d_7 is respectively the maximum distance between and oblique left, oblique right and just below for car's window. The model is shown in Figure 4.

As the model shows, the black spot corresponds to the driver's eyes, L1 and L2 is the farthest edge line in effective area (using the rearview mirror as baseline), L3 and L4 as the baseline in front of window. In order to correspond to the above regions, the line of sight can be described using eye's angle in horizontal direction and vertical direction. Establishment a coordinate system for the x-axis to shaft centerline, the y-axis through black spot and perpendicular to the axis, so:

- Region I: For the regional front, the model can be expressed as region surrounded by the shaft center line of car's windows, the window bottom line, L3, and L4.

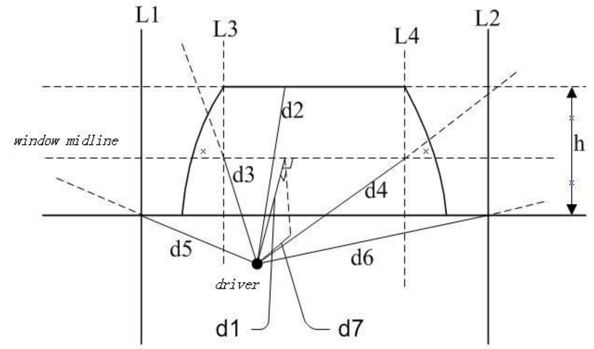

Figure 4. Eye region model for driver's car.

At this time, horizontal angle and vertical angle can be defined for the follow range:

$$\begin{cases} \theta_x \in [-arc\cos\dfrac{d_1}{d_3}, \arccos\dfrac{d_1}{d_4}] \\ \theta_y \in [-\arcsin\dfrac{h/2}{d_1}, 0] \end{cases} \tag{2}$$

- Region II: For the ramp above the regions, the model can be expressed as region surrounded by the shaft center line of car's windows, the window top line, L3, and L4.

$$\begin{cases} \theta_x \in [-arc\cos\dfrac{d_1}{d_3}, \arccos\dfrac{d_1}{d_4}] \\ \theta_y \in [0, \arcsin\dfrac{h/2}{d_1}] \end{cases} \tag{3}$$

- Region III: For the slope below the regions, the model can be expressed as region surrounded by the edge line of sight from the instrument panel where the line (d7 where the straight line), the window bottom line, L3, and L4.

$$\begin{cases} \theta_x \in [-arc\cos\dfrac{d_1}{d_3}, \arccos\dfrac{d_1}{d_4}] \\ \theta_y \in [-arccos\dfrac{d_1}{d_7}, -\arcsin\dfrac{h/2}{d_1}] \end{cases} \tag{4}$$

- Region IV: For the left of the regions, the model can be expressed as Fan-shape region constituted by the window's midline and bottom line, along the d3 and d5 direction.

$$\begin{cases} \theta_x \in [-arc\cos\dfrac{\sqrt{d_1^2 + (\frac{h}{2})^2}}{d_5}, -\arccos\dfrac{d_1}{d_3}] \\ \theta_y \in [-\arcsin\dfrac{h/2}{d_1}, 0] \end{cases} \tag{5}$$

- Region V: For the right of the regions, the model can be expressed as Fan-shape region constituted by the window's midline and bottom line, along the d4 and d6 direction.

$$\begin{cases} \theta_x \in [\arccos\dfrac{d_1}{d_4}, arc\cos\dfrac{\sqrt{d_1^2 + (\frac{h}{2})^2}}{d_6}] \\ \theta_y \in [-\arcsin\dfrac{h/2}{d_1}, 0] \end{cases} \tag{6}$$

B. *Gaze tracking*

When the deformable template matching, we can get some important information, such as size, shape, direction, eye center coordinate and puipil center coordinate. It is assumed that pupil center coordinate expressed as $\bar{x}_c = (x_c, y_c)$, eye center coordinate expressed as $\bar{x}_e = (x_e, y_e)$ and its direction of rotation: θ, then, if established as the x-axis center line of eyes, through the eye center and perpendicular to the center line for the y-axis of the eye coordinate system. We can convert coordinate relations using the following formula.

$$\begin{cases} \dfrac{x}{x'} = \sin\theta \\ \dfrac{y}{y'} = \cos\theta \end{cases} \tag{7}$$

In the new coordinate system, define the eye center coordinate and puipil center coordinate as $\overrightarrow{x'_e} = (x'_e, y'_e)$ and $\overrightarrow{x'_c} = (x'_c, y'_c)$, its deflection angle in the horizontal direction and vertical direction can be computed use formula (8).

$$\begin{cases} \theta_{x'} = \arctan\dfrac{y'_c - y'_e}{x'_c - x'_e} \\ \theta_{y'} = \arctan\dfrac{x'_c - x'_e}{y'_c - y'_e} \end{cases} \tag{8}$$

Finally, make the deflection angle comparison for the actual definition of region, if the angle value located in one of region we defined, marking out the area and output, if not located in any of the define regions, the angle is considered to be meaningless line of sight, that this case, driver could not concentrate on driving or in a state of fatigue.

V. EXPERMENTAL RESULTS AND ANALYSIS

Monitoring of eye gaze direction is the ultimate goal of this article, the final result be affected by localization of human eye. This paper proposed a deformable template matching method to locate driver's eye, the method is an effective solution to problem of dynamic tracing difficult, and providing precondition for the line of sight detection. This paper build experimental platform based on VC6.0 and OpenCV, during the programming processing, the system using the function of MFC class library and some image processing function of OpenCV's library. The interface of real-time detection system as follow Figure:

Figure 5. Effect of real-time monitoring and tracking.

In order to obtain more accurate data, the test results output in the form of frames, recording a total time 131.6s, the output image 507, the average output of 3.8 frames. Some records as shown below:

Figure 6. Iimage frame for non-wearing glasses.

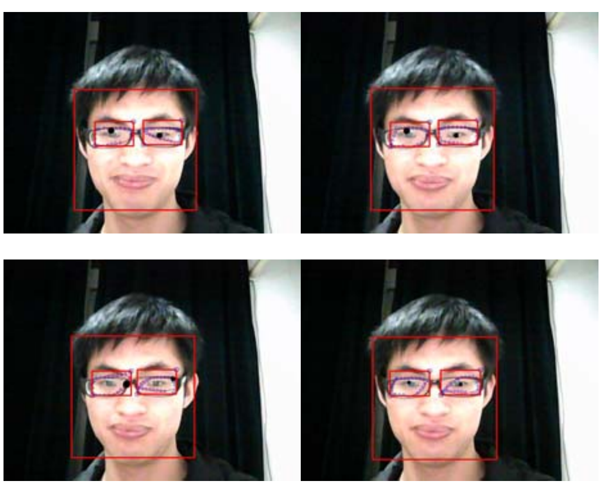

Figure 7. Image frame for wearing glasses

For better observing examination effect, make a statistics to the number that the system read and handle images in a period of time and the rate of miss detection. Drive the result shown as table 3.

TABLE III. TEST RESULT FOR GAZE TRACKING

Number of Image frame	The rate of miss detection	The rate of error detection	times（s）
50	0.0%	36.0%	13.5
134	4.5%	47.2%	34.4
239	3.3%	13.8%	62.9
345	2.6%	9.6%	90.8
465	2.8%	9.4%	116.3
507	2.6%	10.6%	131.6

VI. CONCLUSION

From the experimental results, the correct eye gaze detection rate of about 85%, of which about 100 frames before the image is wearing glasses tested. By analysis the test data, we know that when driver wearing glasses, the correct detection rate is only 53%. Without wearing glasses the correct detection rate is 85% or more. Therefore, it can be concluded that the system will more vulnerable impact by driver wear glasses. For the other case, the system robustness and immunity are better.

ACKNOWLEDGMENT

We would like to thank the M.I.T. Center for Intelligent Control provide download for face database; the image processing Lab provide good working environment.

REFERENCES

[1] Liu Tianjian. A Study of Eyes Tracking and Driver Drowsiness Detection[D].Zhejiang University,2005.

[2] Bala J, Dejong K, Huang J, Vafaie H, Wechsler H. Visual routine for eye dection using hybrid genetic architechures. In:Bacher E, Gelsema ES, eds. Proceedings of the 13th International Conference on Pattern Recofnition. Los Alamitos:IEEE CS Press,1996,p.606-610.

[3] Reinders MJT, Koch RWC, Gerbrands JJ. Location facial features in image sequences using neural networks. In: Essa I,ed. Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition. Los Alamitos:IEEE CS Press 1996,p.230-235.

[4] J.M.Vincent, J.B.Waite and D.J.Myers. Precise Location of Facial Features by a Hierarchical Assembly of Neural Nets. Proc.IEE Conf. onArtifical Neural Networks,1991,p.69-73.

[5] Viola.P, Jones.M. Rapid object dection using a boosted casade of simple features[A]. IEEE Computer Society Press, 2001:511-518.

[6] Viola.P, Jones.M. Robust Real-Time Face Detection[J]. International Journal of Computer Vision, 2004,57(2):137-154.

[7] A.L. Yuille, Peter W.Hallinan and David S.Cohen, Feature extraction from faces using deformable templates,International Journal of Computer Vision, 1992, 8(2): 99-111.