A Detection Method of Driver's Face Orientation Based on Visual Cues and SVM

Mingheng Zhang
State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian, China
Department of Vehicle Engineering, Dalian University of Technology, Dalian, China
Email: zhangmingheng@yahoo.com.cn

Longhui Gang
Navigation College, Dalian Maritime University, Dalian, China
Email: ganglh@yahoo.com.cn

***Abstract*—Face orientation can be used to derive vigilance cues about the vehicle driver. In this paper, a face pose estimation model based on visual features is presented. The face orientation is estimated through SVM with features detection such as eyes, mouth, facial outlines and so on. Innovative point of this research is the utilization of the features triangle, which composed with eyes and mouth locations, to obtain a robust face pose estimate. This method leads to the free face pose estimation and it is non-intrusive to the driver. Finally, the system is assessed with a commercial capture system. The evaluation results indicate that the system presented can be used in the driver assistant system.**

***Index Terms*—Machine Vision, Gaze Detection, SVM, SAD (Safety Assistant Driving)**

I. INTRODUCTION

Driver drowsiness and vigilance is a significant contributing factor to road related crashes worldwide. This would endanger lives of the driver and passengers could cause serious accidents along major roads [1]. Recent studies show that driver's drowsiness accounts for up to 20% of serious or fatal accidents on motorways and monotonous roads, which impair the drivers' judgment and their ability of controlling vehicles. If symptoms of drowsiness and spirit distraction could effectively warn drivers, corrective measures could be taken and disastrous outcomes would be prevented [2].

Researchers all over the world have taken great efforts in this field. In general, there are three types of pre-warning methods based on the driver's physiological parameters [3-5], the vehicle's running statues [6, 7] and the driver's spirit state [8-11].

In which, physiological parameters such as EOG is used to estimate drowsiness [3]. A large number of features can be extracted from the EOG such as blink duration, blink frequency, PERCLOS and other features. EOG is generally high sampled which allows features to be accurately calculated. But, unfortunately, EOG requires at least three electrodes placed on the driver skin, which is not really practical. The vehicle's running statues such as lane deviation, safe-range keeping and steering movements also are used to deduce the driver's vigilance statues. But, in most cases, its accuracy is depending on the type of the vehicle and the driver's habits. Among above different methods, ocular measures, such as eye blinking and eyelid closure, are considered as promising ways for monitoring alertness. At present, it has been a leading technique due to its hardware's characteristics which include small size, low cost and non-intrusiveness to drivers. In this research field, various visual parameters could be extracted when the driver's fatigue occurred.

At normal driving status, driver's face orientation should be the front and within a certain range. Otherwise, the driver's spirit status should be drowsy or distracted. Benoit [10] proposed a real time frequency method to estimate global rotation or translation and the corresponding direction of a moving head based on analysis of the image spectrum in the log polar domain. The method was used for interpretation of global head nods (up/down and right/left oscillations). Hansen [11] seeks to advance the state-of-the-art of 3D face capture and processing via novel Photometric Stereo (PS) hardware and algorithms. The presented method has a good precision, but it's very time consuming. Therefore, the model accuracy and algorithm complexity need to be considered in the development of a practical application.

This paper presents a face orientation estimate method based on SVM with facial visual features detection. Innovative point of this research is the utilization of the features triangle, which composed with eyes and mouth locations, to obtain a robust face pose estimate. This method leads to the free face pose estimation and it is non-intrusive to the driver.

Therefore, this paper is organized as follows. In the section below, firstly the face region is detected through a fusion method with adaboost detector and skin color model; secondly, the features such as eyes, mouth and face contour are obtained based on their visual character respectively; the features vector for SVM is acquired through a characteristic triangle. And then the SVM is reviewed for classification problems. Finally, experiment results are discussed and conclusions are derived respectively.

doi:10.4304/jsw.8.4.924-931

II. FACE FEATURE REGION LOCATION

2.1 Face Region Location

Face region located reliably is very important to locate and analyze other features. At present, methods for face detection can be divided into two types roughly: method based on knowledge [12,13] and the one based on statistical characteristics [14], which all have their own advantages and shortcomings: the former can detect face rapidly, but its precision is lower than the latter's. Therefore, this paper uses AdaBoost classifier based on knowledge to detect the possible face ROI (Region of Interest) in image. In the ROI, the skin color model based on statistical characteristics is adopted to locate face region accurately. In this way, the confliction between detection speed and precision is resolved perfectly. The flow chart of the method is shown as Fig. 1.

Figure 1. The flow chart of face detection

2.1.1 Adaboost detector

The principal of AdaBoost classifier is to collect a large number of simple characteristics (shown as Fig. 2) from target or non-target samples, to train these characteristics through AdaBoost algorithm, to obtain a series of weak classifiers with general categorizing ability and to utilize a certain method to make these weak classifiers form a strong classifier with good categorizing ability. Finally, a classifier cascade for detection is achieved through boosting several strong classifiers [8]. The training and detecting procedure is shown as Fig. 3(a). Fig. 3(b) shows the basic steps for training strong classifier. The result of Adaboost detector is shown as Fig. 7(a).

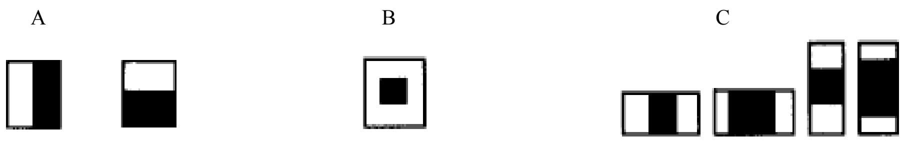

Figure 2. Basic Haar-like features, (A, edge features; B, central features and C, linear features)

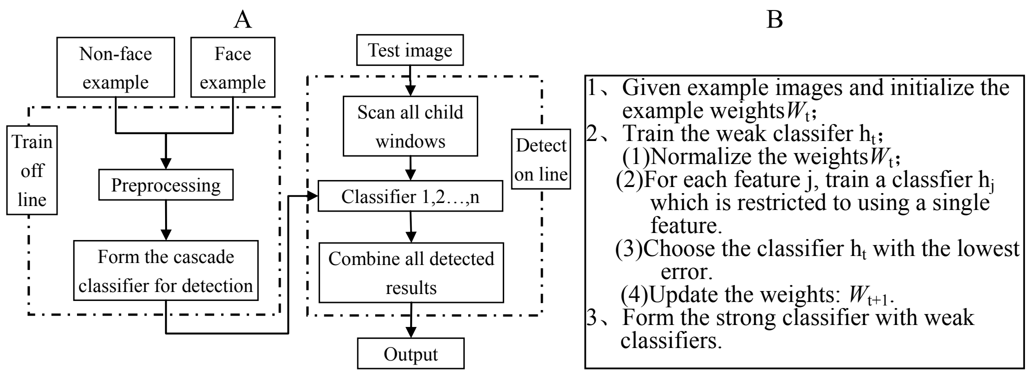

Figure 3. Detection and training procedure of AdaBoost classifier, (A, flow chart of face detection; B, strong classifier training procedure).

2.1.2 Skin color model

For face detection precisely, a skin color model which can be suitable for different skin colors and illuminations is needed. Experiment shows that the difference of Cr and Cb in YCrCb color space is not obvious compared with brightness of different persons. The Cr and Cb of different complexion have the same Gauss skin model [15], shown as Fig. 4.

With the skin color model, the probability of pixels belonging to skin is:

$$P(Cr, Cb) = \exp\{-0.5(x-m)^T C^{-1}(x-m)\} \qquad (1)$$

Here, m and C are the mean value and the covariance of complexion respectively.

With formula (1), the complexion probability image is obtained based on the pixels probability, while the face region is obtained based on the pixels probability, chroma similarity and spacing correlativity. The experiment result is shown as figure 7 (b).

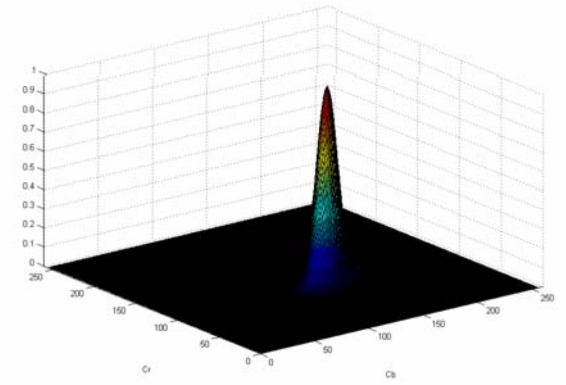

Figure 4. *Cr*, *Cb* distribution of different persons, (*x* axis represents the C_r value, *y* axis represents the C_b value, *z* axis represents the probability value versus C_r and C_b)

2.2 Eyes Detection

Eyes locate in the upper part of face and their gray values are smaller than that of the face skin. But it is influenced by outside illuminations to realize the image binary segmentation on simple gray feature. Analysis shows that the *H*, *Cr*, *Cb* components of face skin and eyes' have some certain aggregation characteristics in *HSV* and *YCrCb* color spaces, which is described in figure 5. Therefore, this paper adopts BP neural network to train a classifier for separating eyes from face region. During the training, pixel's *H*, *Cr*, *Cb* are used as input vectors and face skin (0), eye (1) as the output vectors. The training deviation curve is shown as figure 6.

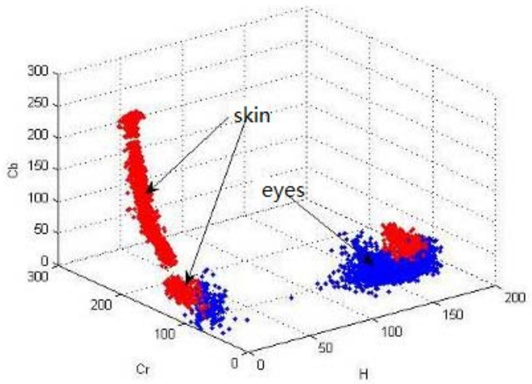

Figure 5. *H*, *Cr*, *Cb* distribution

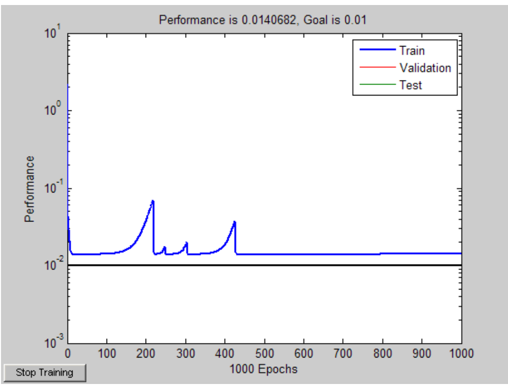

Figure 6. Training deviation

With the trained classifier, the binary image of eyes is obtained. Then, the eyes' region is located through the region index algorithm and eye's distribution characteristic in face region. The experiment result is shown as figure 7 (c).

2.3 Mouth Region Location Based on the Fisher Linear Classifier

Because the distinction on gray level between the lip and the skin is not clear, the gray information is easy influenced with light condition changing; face moving and face rotating. So the lip region edge is not easy to be detected. Analysis shows that the lip is redder than face skin and the unitized *RGB* color would not change under various light conditions, face movement or rotations. So the Fisher linear classifier is adopted to make the segmentation between skin color and lip color. The training steps of Fisher linear classifier as bellow.

1. Calculating the mean of skin color and lip color:

$$m_i = \frac{1}{N_i}\sum_{x\in\omega_i} x \qquad i = 1,2 \tag{2}$$

2. Calculating the congeneric scatter matrix S_i and the sum S_w:

$$S_i = \sum_{x\in\omega_i}(x-m_i)(x-m_i)^T \qquad i = 1,2$$

$$S_w = S_1 + S_2 \tag{3}$$

3. Calculating the most appropriate Fisher classify vector:

$$w^* = S_w^{-1}(m_1 - m_2) \tag{4}$$

After training, the projection axis w^* is obtained. With it we can calculate every color pixel's (skin and lip color) Fisher projection point. Finally, the mouth position is achieved through the location and shape features.

The test results of above algorithm are shown as figure7 (d).

2.4 Facial Contour Detection

The facial contour's detection accurately is important for posture estimation. Some researches about face detection apply the constraint that facial contour is similar to ellipse to validate whether the region determined is face or not [16]. So, this paper conducts a series of image preprocessing such as morphology, region identification, hole filling, LOG edge detection, morphological thinning and edge chain code tracking. Then the edge points which represent the facial contour can be acquired through "specific" chain codes screening which describe the facial contour.

The image preprocessing procedure experiments are shown as figure 8. The experiments indicate that the algorithm can not only remove the "non-specific" facial contour points, but also retain the features of facial contour.

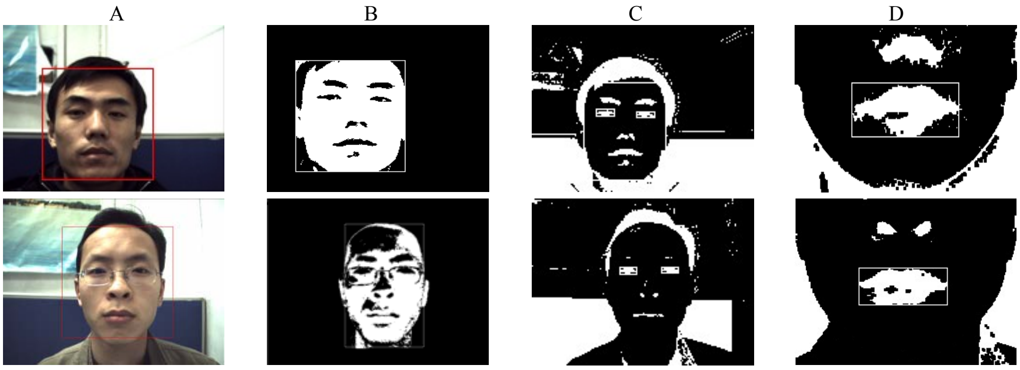

Figure 7. Results of face and features detection, (A, adaboost detection results; B, skin color model detection result; C, eyes detection; D, mouth detection)

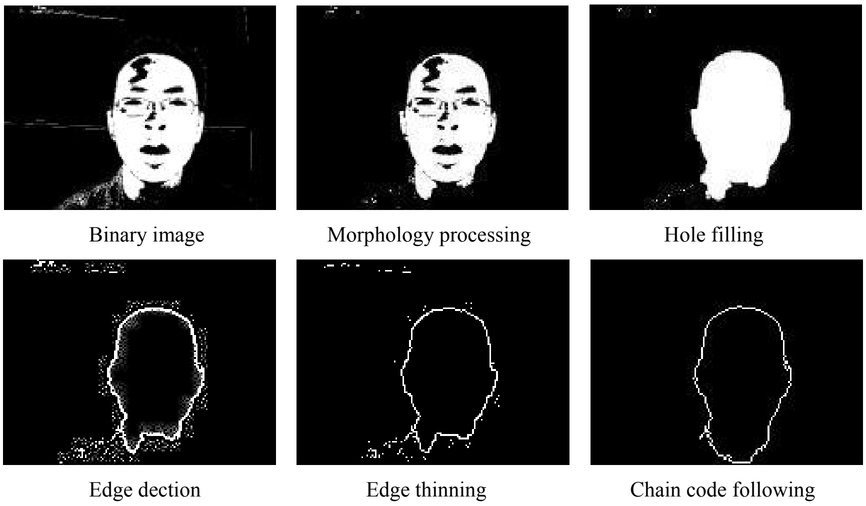

Figure 8. Preprocessing of the facial contour

After the preprocessing with above constraints, most of the selected edge points should be located on the facial contour. Then the least-squares fitting operation is applied to the contour according to the elliptical general expression.

The elliptical general analytical expression is:

$$x^2 + C_0xy + C_1y^2 + C_2x + C_3y + C_4 = 0 \tag{5}$$

All contour points coordinates selected are substituted into the formula, then the five parameters are obtained through the following equation:

$$\begin{bmatrix} x_0y_0 & y_0^2 & x_0 & y_0 & 1 \\ x_1y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n-1}y_{n-1} & y_{n-1}^2 & x_{n-1} & y_{n-1} & 1 \\ x_ny_n & y_n^2 & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} C_0 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} = \begin{bmatrix} -x_0^2 \\ -x_1^2 \\ -x_2^2 \\ -x_3^2 \\ -x_4^2 \end{bmatrix} \tag{6}$$

Let the elliptical center be (x_c, y_c), the major axis be a, the minor axis be b, and the angle between major axis and x-axis positive direction be φ. Then the relations among $\boldsymbol{C_0}$, $\boldsymbol{C_1}$, $\boldsymbol{C_2}$, $\boldsymbol{C_3}$, $\boldsymbol{C_4}$, (x_c, y_c), a, b and the angle φ are described as follows:

$$x_c = (-2C_1C_2 + C_0C_3)/(4C_1 - C_0^2) \tag{7}$$

$$y_c = (-2C_3 + C_0C_2)/(4C_1 - C_0^2) \tag{8}$$

$$\phi = \frac{1}{2}(\tan^{-1}(\frac{C_0}{1-C_1})) \tag{9}$$

$$a = \sqrt{G(\cos^2\varphi - \sin^2\varphi)/(C_1\sin^2\varphi - \cos^2\varphi)} \tag{10}$$

$$b = \sqrt{G(\cos^2\varphi - \sin^2\varphi)/(\sin^2\varphi - C_1\cos^2\varphi)} \tag{11}$$

Here, $G = C_4 + \frac{C_2x_0 + C_3y_0}{2}$.

With formula (7) to (11), the parameters which describe facial contour curve can be calculated. In order to verify efficiency of the method, some facial contour fitting experiments are conducted shown as figure 9. Experiments indicate that the facial contour curve can be detected through the algorithm adopted with the front and lateral orientation state.

Binary image   Freeman chain code   Fit Ellipse   Ellipse detection

Figure 9. Test of facial contour's fitting

III. The Detection Of Face Orientation Based On Svm

3.1 Support Vector Machine

Support Vector Machine (SVM), a novel machine learning algorithm, has been recently proved that is a promising tool for both data classification and pattern recognition. SVM is a particular learning system that is based on the margin-maximization principle (figure 10). This learning strategy introduced by Vapnik in 1995 is a powerful method which has already been successfully received a wide variety of applications ranging from particle identification, face identification and text categorization to intrusion detection, bioinformatics and database marketing in the few years since its introduction [17]. Here, we only give a very brief introduction to SVM, more detailed descriptions can be found in Burges, Campbell, Cristiannini and ShaweTaylor.

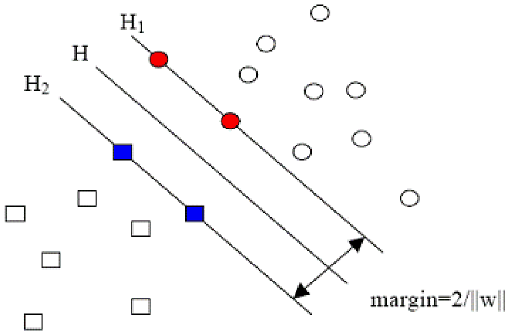

Figure 10. A maximal margin hyperplane

Given a training set of instance-label pairs (x_i, y_i); i=1,…,l, where $x_i \in R^n$ and $y \in \{1,-1\}^l$, the SVM requires solution of the following optimization problem:

$$\begin{aligned} &\min_{\omega,b,\xi} \quad \frac{1}{2} w^T \cdot w + C\sum_{i=1}^{l} \xi_i \\ &s.t. \quad y_i[w^T \cdot \varphi(x_i) + b] \ge 1 - \xi_i \\ &\xi_i \ge 0, i = 1,2,\cdots,l \end{aligned} \tag{12}$$

Here training vectors x_i is mapped into a higher (maybe infinite) dimensional space by the function Φ. SVM finds a linear separating hyper plane with the maximal margin in this higher dimensional space. $C > 0$ is the penalty parameter of error term. Furthermore, $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is called the kernel function. Now, new kernels are being proposed by researchers, such as the following four basic kernels:

- Linear: $K(x_i, x_j) = x_i^T x_j$
- Polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- Radial basis function (RBF): $K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2} \quad \gamma > 0$
- Sigmoid: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)^d$

Here, γ, r and d are kernel parameters.

The dual optimization problem is obtained with Lagrange multiplier ∂_i, shown as following [18]:

$$\begin{aligned} &\max_{\partial} \quad W = \sum_{i=1}^{l} \partial_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{i=1}^{l} \partial_i \partial_j y_i y_j K(x_i, x_j) \\ &s.t. \quad \sum_{i=1}^{l} \partial_i y_i = 0, 0 \le \partial_i \le C, i = 1,2,\cdots,l \end{aligned} \tag{13}$$

Then, the SVM decision function is :

$$f(x) = \text{sgn}(\sum_{i=1}^{l} \partial_i k(x_i, x) + b) \tag{14}$$

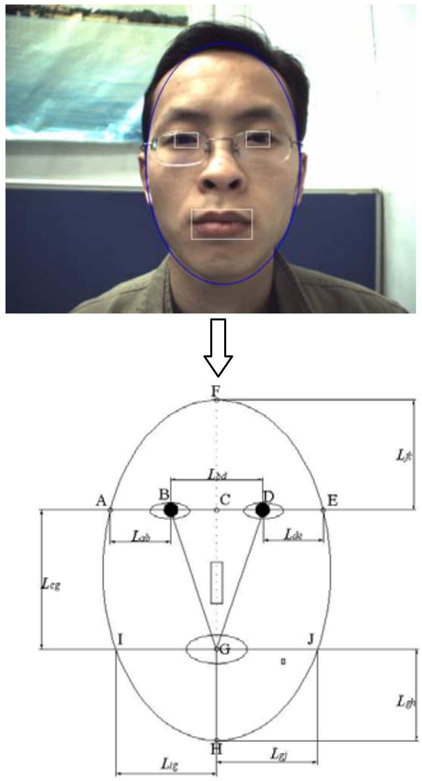

Figure 11. Sketch map of eye, mouth and face, (the features triangle composed with the eyes, mouth central point location)

3.2 Features Dataset Construction

In order to estimate the face's orientation, a feature vector which can reflect the face direction changing is important. Analysis shows that the position relationship changes obviously between the face outline and the triangle composed of eyes and mouth when the head occurs deflecting and pitching. Therefore, we can fit an ellipse based on the face edge detection. The features vector is obtained with relationship between the triangle and the ellipse, shown as figure 11.

The following information can be drawn from geometrical relationship between the triangle and the ellipse, which include the center coordinates of two eyes (x_{el}, y_{el}) and (x_{er}, y_{er}), the mouth center coordinate (x_m, y_m), other point coordinate such as (x_a, y_a), (x_e, y_e), (x_f, y_f), (x_h, y_h), (x_i, y_i) and (x_j, y_j).

Then, we can get the features vector used for constructing the dataset for SVM:

$$Z = \left(E_1, E_2, E_3, E_4\right)^T \tag{15}$$

Here,

$$\begin{cases} E_1 = \dfrac{L_{ab}}{L_{de}} = \sqrt{\dfrac{(x_{er}-x_a)^2+(y_{er}-y_a)^2}{(x_{el}-x_e)^2+(y_{el}-y_e)^2}} \\ E_2 = \dfrac{L_{bd}}{L_{cg}} = \sqrt{\dfrac{(x_{er}-x_{el})^2+(y_{er}-y_{el})^2}{(\frac{x_{er}+x_{el}}{2}-x_m)^2+(\frac{y_{er}+y_{el}}{2}-y_m)^2}} \\ E_3 = \dfrac{L_{fc}}{L_{gh}} = \sqrt{\dfrac{(\frac{x_{er}+x_{el}}{2}-x_f)^2+(\frac{y_{er}+y_{el}}{2}-y_f)^2}{(x_m-x_h)^2+(y_m-y_h)^2}} \\ E_4 = \dfrac{L_{ig}}{L_{gj}} = \sqrt{\dfrac{(x_m-x_i)^2+(y_m-y_i)^2}{(x_j-x_m)^2+(y_j-y_m)^2}} \end{cases} \tag{16}$$

In order to verify the above analysis, a statistic test of E_1~E_4 changing with face posture is conducted, which is shown as figure 12. Here, we select 50 data from the five types of face posture which include the normal straight, the left and right deflection and the up and down pitching.

Some important information can be drowned from the statistics. Firstly, the features E_1 and E_4 changes obviously when the head occurs left or right deflection, but not with the up or down pitching (figure 12 (a) and (d)). Secondly, the feature E_3 changes obviously when the head occurs up or down pitching, but not with the left or right deflection (figure 12 (c)). Therefore, the face postures changing can be reflected from the feature vector *Z*.

3.3 Experiments and Discussion

In this paper, the SVM model constructed will be trained with the features vector dataset extracted from image, and then it will be used to classify the face orientation with the validation dataset. Multi-class implementation of Libsvm, which uses the one-against-one approach, is utilized for this purpose. In this experiment, RBF kernel functions are selected. After conducting the grid search on the training dataset (including 150 groups), the optimal (C, γ) is (128, 0.5), and a cross validation on the dataset (including 50 groups), the accuracy is 98.9%.

Table 1 shows the classifying results of the experiments. As we can see, a better recognition effects is achieved for the face orientation. The tracking experiment results are shown as figure 13.

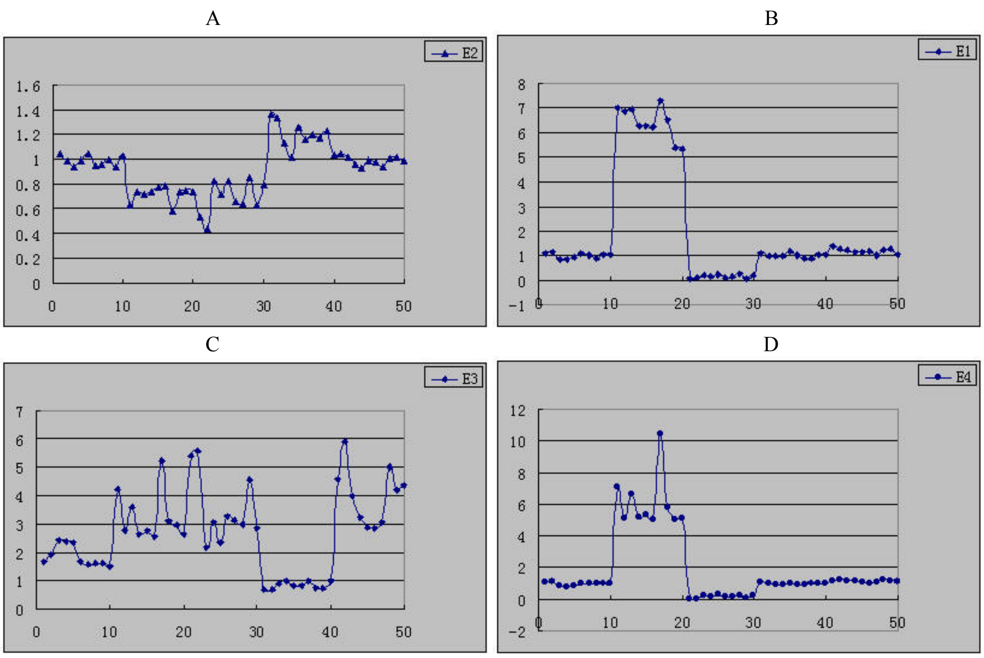

Figure 12. Statistics of the features changing with different posture, (A, feature E_1 changing with the different posture; B, feature E_2 changing with the different posture; C, feature E_3 changing with the different posture, and D, feature E_4 changing with the different posture). Horizontal ordinate is the data index from 50 selection samples, which represent five state including normal straight, the left and right deflection and the up and down pitching , with 10 data per group. Vertical ordinate represents the feature value versus the posture state.

TABLE I.
EXPERIMENT RESULTS

	E_1	E_2	E_3	E_4	training set	validate set	accuracy
Normal	1.107856 1.151562 0.856226	1.046769 1.120824 0.876202	1.040198 0.98291 0.945798	1.658779 1.933162 2.408722	50	20	98%
Deflecting	7.00904 6.832366 6.945105	7.052528 5.069548 6.616059	0.625644 0.741283 0.712245	4.231647 2.762607 3.576928	50	15	98%
Pitching	0.962093 0.963685 1.198012	0.927706 0.965739 0.999314	1.126263 1.018819 1.259405	0.870745 0.989419 0.825243	50	15	97%
Amount to	——	——	——	——	150	50	97.5%

A B

C D

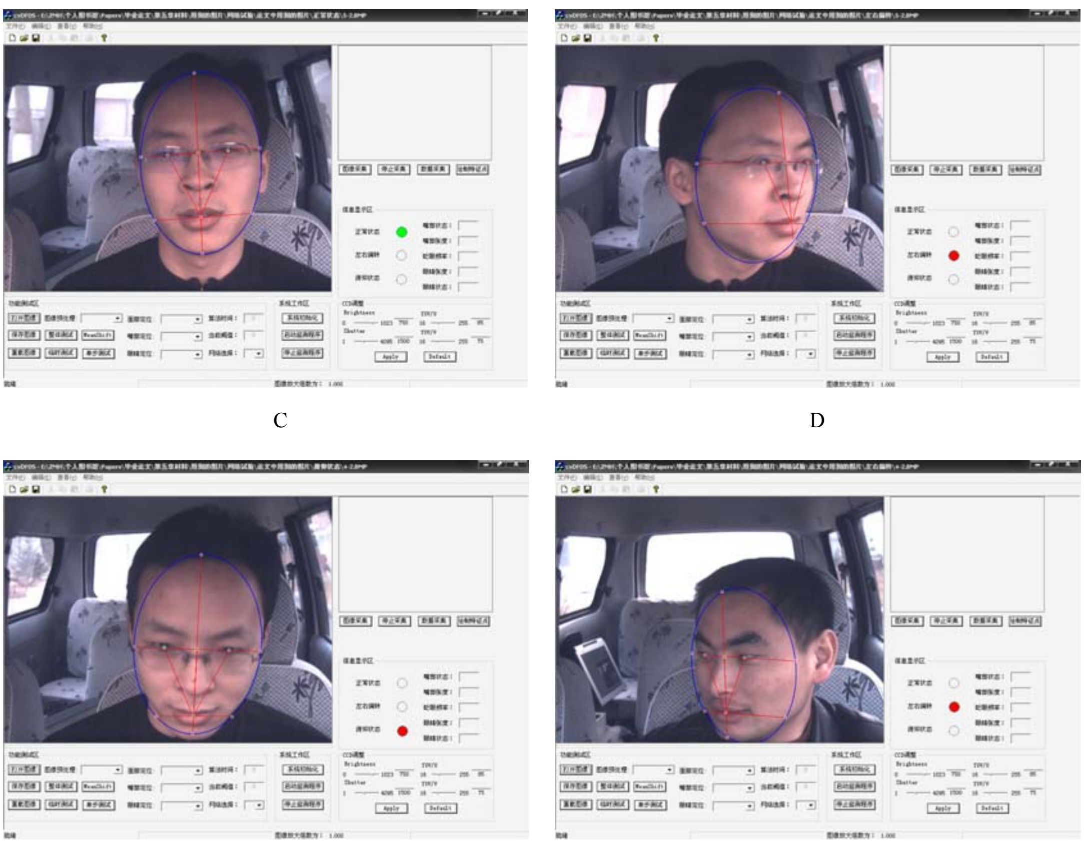

Figure 13. Results of the tracking experiments, (A, normal state; B, left deflecting state; C, pitching state, and D, right deflecting state). Normal, pitching and deflecting state detection are colored with green and red lamp, respectively.

IV. CONCLUSION AND FUTURE WORK

This paper proposes a number of methods to detect face quickly by using combination of AdaBoost detector and skin color model, to identify eye region from skin based on the BP neural network, and to separate the mouth in face region with the Fisher linear classifier. Based on the eye, mouth and face ellipse contour detection, the driver's face orientation can be determined with SVM. Test results show that the system can monitor the face features effectively and realize warning based on specified rules.

However, there are some unresolved issues to be discussed in the future work. First, there are many kinds of different face orientations, how to decide which kind they belong to is an important issue for practical application. In addition, face features segmentation in actual environment is a difficult task in this research area.

ACKNOWLEDGMENT

This work is partially supported by grants from the Humanities and Social Sciences Foundation of the Ministry of Education in China (Project No.12YJCZH280) and the Ph.D. Programs Foundation of Ministry of Education of China (20112125120004).

REFERENCES

[1] Lal, S.K.L., Craig, A., Boord, P , "Development of an algorithm for an EEG-based driver fatigue countermeasure", Journal of Safety Research, Vol.34, 2003, pp.321-328.

[2] Yihu Wu, Jiajia Li, Dan Yu, Haitao Cheng , Research on Quantitative Method about Driver Reliability, Journal of Software, Vol 6, No. 6, June, 2011, pp.1110-1116.

[3] Bruno Jammes, Hassan Sharabty, Daniel Esteve, "Automatic EOG analysis: A first step toward automatic drowsiness scoring during wake-sleep transitions", Somnologie, vol.12, 2008, pp.227−232.

[4] M. Yeo, X. Li, K. Shen, E. Wildersmith, "Can SVM be used for automatic EEG detection of drowsiness during car driving?", Safety Science, vol.47, 2009, pp.115-124.

[5] Rami N. Khushaba, Sarath Kodagoda, Sara Lal, and Gamini Dissanayake, "Driver Drowsiness Classification Using Fuzzy Wavelet-Packet-Based Feature-Extraction Algorithm", IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, vol.58, 2011, pp.121-131.

[6] Jyun-Guo Wang, Cheng-Jian Lin, Shyi-Ming Chen, "Applying fuzzy method to vision-based lane detection and departure warning system", Expert Systems with Applications, vol.37, 2010, pp.113-126.

[7] T.H. Yu, R.B. Wang, L. Guo, "A Study on Recognition Methods of Linear Lane Markings in Different Illumination Conditions", Automotive Engineering, vol.27 , 2005, pp.510-513.

[8] Antoine Picot, Sylvie Char-bonnier and Alice Caplier, "Drowsiness detection based on visual signs blinking analysis based on high frame rate video", Instrumentation and Measurement Technology Conference, 2010, pp.801-804.

[9] I. G. Damousis, D. Tzovaras, M.G. Strintzis, "A fuzzy expert system for the early warning of accidents due to driver hypo-vigilance", Pers Ubiquit Comput, vol.13, 2009, pp.43-49.

[10] A. Benoit, A. Caplier, Motion Estimator Inspired From Biological Model For Head Motion Interpretation, IEEE Transactions on Circuits and Systems: Special Issue on Bio-Inspired Processors and CNNs for Vision, vol.46 , 2009, pp.269-280.

[11] Mark F. Hansen, Gary A. Atkinson, Lyndon N. Smith, Melvyn L. Smith, 3D face reconstructions from photometric stereo using near infrared and visible light, Computer Vision and Image Understanding, vol.114, 2010, pp.942-951.

[12] P. Viola，M.J. Jones, Robust Real-Time Face Detection, International Journal of Computer Vision, vol.57, 2004,pp.137-154.

[13] J.G. Wang, T.N. Tan, A new face detection method based on shape information, Pattern Recognition Letters, vol.21 , 2000, pp.463-471.

[14] Chuanxu Wang, Face Segmentation Based on Skin Color in Complicated Background and Its Sex Recognition, Journal of Software, Vol 6, No. 7, July, 2011, pp.1209-1216.

[15] R.B. Wang, K.Y. Guo, A Monitoring Method of Driver Fatigue behavior Based on Machine Vision, IEEE Intelligent Coference, 2003.

[16] Y. C. Fang, Y. H. Wang, and T. N. Tan, Improving Face Detection through Fusion of Contour and Region Information, Chinese Journal of Computers, vol.27, 2004, pp.482-491.

[17] C. Cortes and V. Vapnik, Support-vector network, Machine Learning, vol.20, 1995, pp.273–297.

[18] S.Y. Hu and G.T. Zheng, Driver drowsiness detection with eyelid related parameters by Support Vector Machine, Expert Systems with Applications, vol.36, 2009, pp.7651-7658.

Mingheng Zhang was born in 1977.7, he was received M.S.and Ph.D. degrees from the Jilin University in 2004 and 2007 respectively. He has published more than 30 papers. Now he is a lecturer in Dalian University of Technology. His research interests include intelligent vehicle, safety assistant driving and image processing.

Longhui Gang was born in 1979. 12, she was received M.S.and Ph.D. degrees from the Jilin University in 2004 and 2007 respectively. Now she is a lecturer in Dalian Maritime University. Her research interests include traffic safety and traffic information proceeding.