2004 IEEE Intelligent Vehicles Symposium
University of Parma
Parma, Italy • June 14-17, 2004

A Monitoring Method of Driver Mouth Behavior Based on Machine Vision

Chu jiangwei Jin Lisheng Tong Bingliang Shi Shuming Wang rongben
Transportation College, JiLin University, Changchun 130025, China
Ph: 86 431 5705461 Fax: 86 431 5705461
e-mail: cjw_62@jlu.edu.cn bull_jin@sohu.com

Abstract—**When we use the computer vision to inspect the driver's driving behavior, the identifying of the mouth state is one of the key technologies. In fact, when a driver drives in a normal, talking or dozing state, his/her mouth opening degree will quite different. According to this fact, this paper uses the Fisher classifier to extract the mouth shape and position, then uses the mouth region's geometry character as the feature value, and put all of these features together to make up an eigenvector as the input of a three-level Bp network, then we get the output among three different spirit states. The experiment results show that this new method can inspect the driver's mouth region state accurately and quickly.**

I. INTRODUCTION

As we all know that a person will doze when he/she is very tired and his/her mouth is very bigger sometimes. As a man speaks his mouth' shape will change continuously. All of these phenomenons are forbidden to drivers under normal driving conditions. Some data shows that more than 50 percent speedway traffic accidents occur because of the driver's tiredness results from their long time driving or scatterbrained state when the drivers talk.

When these happen, the shape of the driver's mouth changes obviously. So we can monitor the driver's behavior by judging his/her mouth state, and offers the driver with a necessary assistant.

II. THE LOCATION OF PEOPLE'S FACE AND MOUTH

As we all know, there are three sorts of skin colors according to different people: black, white and yellow. From the analysis about people face's RGB pixels, we know that the R and G portions follow planar Gauss distribution[1][2][3]. So we can use these two portions to locate the driver's face. In the conference of the IEEE intelligent vehicles 2003, we have introduced the locating method of driver's face and eyes[4]. According to this, we can find the probable position of the mouth because the mouth is always at the lower part on the face, show in Figure 1.

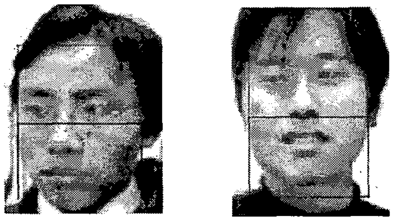

Fig.1 The location of driver's face and mouse

Red line represents face area

Blue line represents the mouse area

III. THE DETECTION OF THE MOUTH SHAPE AND POSITION

Because the distinction on gray level between the lip and the skin is small and the gray information is easy influenced when the light condition changes, the face moves or the face rotates, so the lip region edge is not easy to detect. Especially when the shadow and teeth appears by turn, the lip edge becomes much fainter[5]. So we can't get credible result by the detection on the gray level and the edge information of the lip and the face.

Analysis shows that, compared with the skin color, the lip's main character is redder and the unitized RGB color is not change under various light conditions, the face moves or rotates[6]. So we can overcome the intrinsic shortcomings of the gray image by using the color information and the mode classifying technology to detect the lip area. We know that the Fisher linear classifier can divide two classes at the most degree, so it can distinguish

0-7803-8310-9/04/$20.00 © 2004 IEEE

the color parts. Besides this, the light conditions have no effect on Fisher classifier. Since its training is offline, it can save more time. On all these merits we select Fisher classifier to divide the lip region.

A. Fisher linear classifier

During the classification of two classes, like the lip and the skin, Fisher linear classifier can find a projection axes W*, and the W* can make the value of the rule formula $J_F(w)=\frac{w^T S_b w}{w^T S_w w}$ to be maximum, which makes the distinction between the two classes bigger and makes the distinction in the same class as small as possible. The training steps of the Fisher linear classifier are bellow.

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

$$m_k = \frac{1}{n_k}\sum_{x \in xk} x \quad \text{k=1,2} \tag{1}$$

2) Calculating the congeneric scatter matrix S_k and the sum S_w:

$$S_k = \sum_{x \in xk}(x-m_k)(x-m_k)^T \tag{2}$$

3) Calculating the most appropriate Fisher classify vector:

$$w^* = S_w^{-1}(m_1 - m_2) \tag{3}$$

In this paper, R G and B portion of each pixel constitute the vector x.

$$x=(r,g,b)^T \tag{4}$$

$$r=\frac{R}{R+G+B} \tag{5}$$

$$g=\frac{G}{R+G+B} \tag{6}$$

$$b=\frac{B}{R+G+B} \tag{7}$$

In order to ensure the adaptability of the Fisher classifier, we use 30 different persons' face image data include different light conditions and different speaking states as its training data. We demarcate the lip color and the skin color pixel (excluding the beard and the teeth that are not belonging to the skin and the lip color pixel) and put them into their training lists, then start to train the Fisher classifier through the steps given above.

After training, we get projection axis w*. With it we can calculate every color pixel's (skin color, lip color) Fisher projection points through equation (8), then we get the Fisher transformer projection Figure 2.

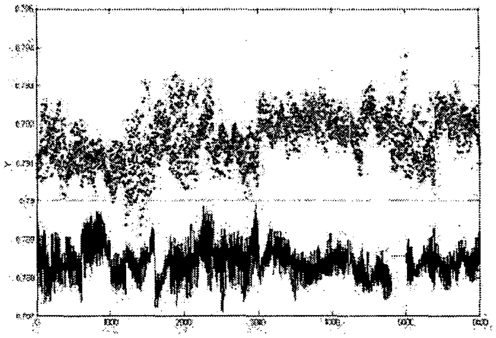

Fig.2 Fisher transformer chart of the stylebook

Red represents the lip; blue represents the skin.

$$y = w^{*T} x \tag{8}$$

From Figure 2, we can see that if we could find an appropriate threshold y0. We will classify the skin color and the lip color easily by compare the entire y with y0. In this paper, we determine y0 as equation (9).

$$y_0 = \frac{N_1 y_1 + N_2 y_2}{N_1 + N_2} \tag{9}$$

N1, N2 represents the sum of two categories stylebook, and y1, y2 represents the mean of two stylebooks projection value. Figure 3 shows the images after the projection using the Fisher linear classifier.

B. Locating the mouth profile.

According to the geometrical character of human face and physical location of the eyes, we can establish the Area of Interest of the mouth. Through Fisher transformation, a series of continuous areas can be obtained. These areas include a skin-color area and a background area whose color is close to lip color. In complex color environment, it is possible to locate a lot of similar color area to lip color such as the areas of nostril and facial skin.

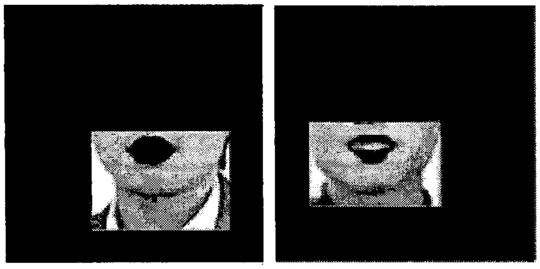

Fig.3 Fisher projection transformer

Red line represents the skin region, blue represents the lip region, and Green line represents mouth area

To eliminate the noises and locate the target area, we carry out a preliminary image processing. We mark the continuous areas and partition them. Eventually, we analyze the size the width and the height of the partitioned areas. The location and the profile of the mouth can be obtained through a horizontal and vertical projection. See Figure 4.

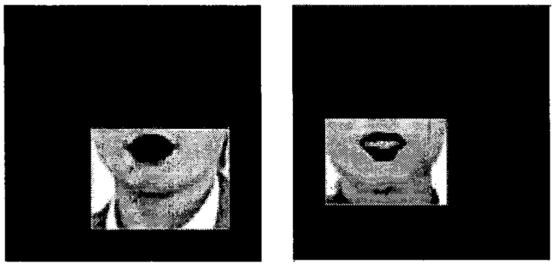

Fig. 4 Sample images

IV. THE EXTRACTION OF MOUTH CHARACTER AND BP NET CLASSIFIER

A. *The extraction of mouth character*

Compared with mouth normal state, talking and dozing is quite different in mouth shape. Talking, dozing in particular, the mouth is wide open while at normal condition the mouth is hardly open. So, mouth states can be obtained with these characters.

In research, we find that the maximum width ($W_{\max}$) and maximum height of the mouth ($H_{\max}$) can indicate the different openness. Thus we put it as character in this method. The height (H_m) between top lip and the bottom lip varies greatly when one is talking and keeping mouth closed. It is reasonable to make it the feature value. A group of vectors can be formed with the three index ($W_{\max}$, $H_{\max}$, H_m). Then the mouth geometrical characters can be obtained. See Figure 5.

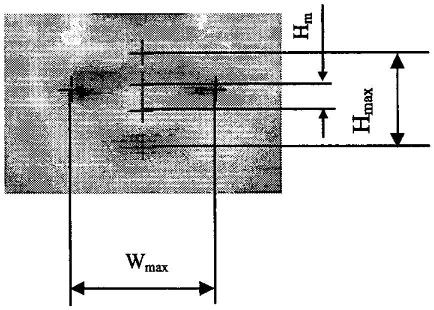

Fig.5 Mouth character

Thus, parameters describing the driver's mouth states can be obtained. A character vector Z is composed of these parameters which are later used as input vectors in order to get the driver's mental states.

B. *BP net classifier*

BP net is the most popular artificial nerve net. At present, 80% of net in use is BP net. It has been developing into a classical model in this field[7]. Up to now, it has been used in such area as automatic controlling, image processing, mode recognizing; though the training process takes time, in practice, we can take measures to improve its performance by optimizing net structure, training offline. In this article, we adopt BP net to identify the driver's mouth states.

Based on regional character, BP net is a 3-layer structure. A 3-node input layer, which represent different characters of mouth region.

A 10-node hiding layer and a 3-node output layer which is relative to the three different mouth characters. In the hiding layer, the transfer function is a sigmoid one.

There are three output vectors. They are Y1=[1,0,0],Y2=[0,1,0] and Y3=[0,0,1]. Y1, Y2 and Y3 refer to the different mouth states of dozing, talking and shut-up while driving respectively. The structure is shown in Figure 6:

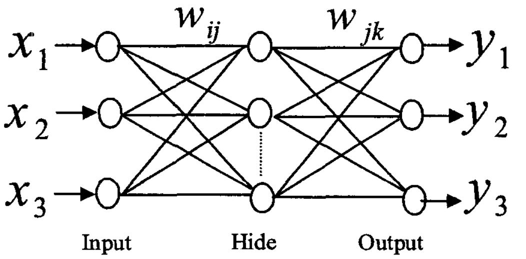

Fig.6 BP net Structure

C. *Iterative method of the BP*

We adopt improved BP method to train the net. The samples are$(p_1, p_2, ..., p_n; \quad t_{kp})$. The errors function of the sample is shown in equation (10).

$$e_p = \frac{1}{2}\sum_{k=1}^{l}(t_{k,p} - y_{k,p)}) \qquad (10)$$

Where, $t_{k,p}$refers to the anticipate output value of the net and yk,p refers to the actual output value of the net.

W_{ij} is the weights which connects the hiding and the input layers. The threshold of the hiding layer isθ_i. In the reverse direction of the error function's grads, the value of W can be adjusted to ΔW. So ΔW is defined as follows.

$$\Delta W^{(n)} = -\eta\frac{\partial e_p}{\partial W} + \alpha\Delta W^{(n-1)} \qquad (11)$$

In equation (11), $\Delta W^{(n)}$ is the adjusted value of W through n times iteration.

Correspondingly, $\Delta W^{(n-1)}$is the value through (n-1) times iteration.

Momentum factor α=0.92; learning rate η=0.3. Then, W can be adjusted as follows:

$$W = W + \Delta W \qquad (12)$$

D. *Training the BP*

The samples consist of driver's three different mouth states. The three different samples are dozing (wide open), talking (moderate open) and shut-up. The numbers of the three in sequence are 30, 25 and 35. During the procedure of training, The parameters used are as follows:

Max training times: 10000; SQ Error: 0.02; Error Index: 0.02; adapting velocity: 0.01.

By training several times, we get the convergence result which in the error range of 0.001. The graphs in Figure 7 indicate the training results.

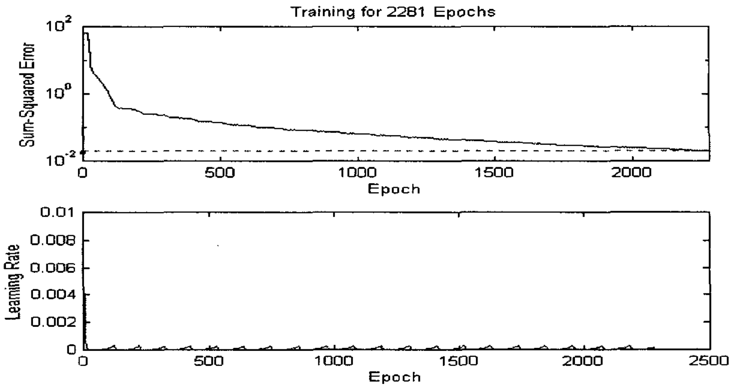

Fig.7 The training results of different mouth characters

The training weights W and threshold matrix B list as follows:

$$W1=\begin{bmatrix} -0.0936 & -0.0209 & -0.0114 \\ 0.0231 & -0.1463 & -0.1904 \\ -0.0248 & 0.0117 & 0.0576 \\ -0.0403 & -0.0758 & -0.1105 \\ 0.0119 & -0.0679 & 0.0109 \\ -0.0931 & -0.0763 & 0.3920 \\ -0.0327 & -0.0465 & -0.0660 \\ 0.0882 & 0.0084 & 0.0102 \\ 0.0243 & 0.0208 & 0.0231 \\ 0.0571 & -0.0616 & -0.0380 \end{bmatrix} \quad B1=\begin{bmatrix} 9.3864 \\ 5.8444 \\ 0.1452 \\ 9.6058 \\ 3.6545 \\ 0.4564 \\ 8.8527 \\ -7.9506 \\ -3.3594 \\ -5.9235 \end{bmatrix} \quad B2=\begin{bmatrix} 0.4519 \\ 0.9424 \\ -0.1631 \end{bmatrix}$$

$$W2=\begin{bmatrix} 0.8973 & 0.7337 & 0.3480 & 0.4261 & -0.3276 & 0.0255 & 0.8206 & 0.5990 & 0.4458 & 0.7473 \\ -0.5991 & -0.6721 & -0.1620 & -0.7044 & 0.3638 & -0.7973 & -0.2936 & -0.4074 & -0.4767 & 0.7659 \\ 0.2190 & -0.0311 & 0.0337 & -0.8430 & 0.8946 & 0.0746 & -0.3185 & -0.6841 & -0.0694 & 0.3705 \end{bmatrix}$$

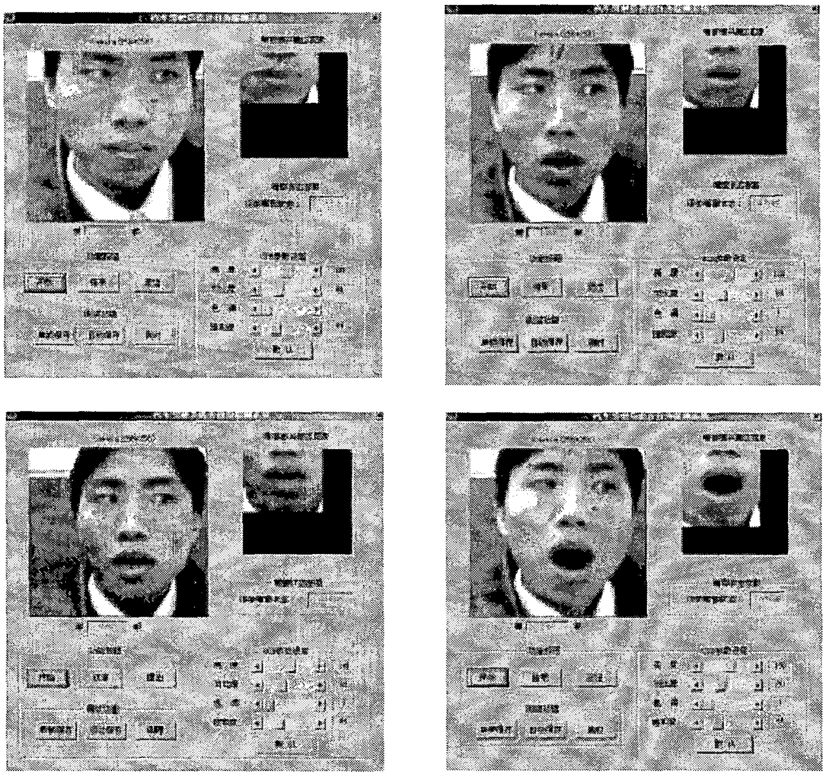

Fig.8 Surveillance system experimental results of driver fatigue behavior based on mouth states

V. Experiments

After the training session, we pick up 124 sample images of different mouth states to carry out a simulation experiment. More than 87% of the samples have been identified correctly. In addition, this method is tested in practice in real time. The experimental results of monitoring driver's fatigue states are shown in Figure 8. Unless the driver swerves his head dramatically, the method mentioned in this article can give a quick accurate trace to the driver's mouth. The tracing frequency can reach as high as 12 frames per second. In experiment, while part of the mouth is out of the machine vision, the method fails to trace the mouth. But when the mouth reenters the vision, this method will gain tracing again.

VI. Conclusion

The traffic accidents keep yearly increasing because of the driver's artificial factors. In this article we put forward a new method based on machine vision to monitor the driver's mouth states. By constantly monitoring and tracing the driver's mouth states, we can send an alarming message while the driver is dozing or talking with other. The method proves to be of high accuracy and efficiency. Thus it offers a valid monitoring method to Safe Assistance Driving.

References

[1] Yang J,Lu W,Waibe A. Shin-color modeling and adaptation[R].Pittsburgh: CMU-CS, 1997.97-146.

[2] Y Gong, M Sakauchi. Detection of regions matching specified chromatic features[J].Computer Vision and Image Understanding,1995,61(2):263-269.

[3] M.A.Turk and A.Pentland.Face recognition using eigenfaces.Proc.IEEE Conf.on Computer Vision and Pattern Recoginition, pp. 586-591, Maui, HI, USA,1991.

[4] Wang rongben, A Monitoring Method of Driver Fatigue behavior based on Machine Vision, in: Proceedings of the IEEE Intelligent Vehicles Symposium'03, Columbus, Ohio, 2003.

[5] R. Kaucic and A. Blake. Accurate, real-time, unadorned lip tracking. Proc. 6th Int.Conf. Computer Vision, Bombay, India, pages 370–375, 1998.

[6] P. Duchnowski, M, Hunke, D. Bushing, U.Meier, and A. Waibel. Toward movement invariant automatic lip-reading and speech recognition. In Proc. Int. Conf. On Acoust., Speech, Signal Processing, Detroit, USA, 1995.

[7] Fukuda Toshio, Shibata Takanori. Theory and Applications of Neural Networks for Industrial Control Sytems. IEEE Transctions on Industrial Electronics, December 1992, 39(6):472–489.