Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007

YAWNING DETECTION FOR MONITORING DRIVER FATIGUE

XIAO FAN, BAO-CAI YIN, YAN-FENG SUN

Beijing Key Laboratory of Multimedia and Intelligent Software, College of Computer Science and Technology, Beijing University of Technology, Beijing 100022, China
E-MAIL: wolonghongni@emails.bjut.edu.cn

Abstract:

Fatigue driving is an important reason of traffic accidents. Yawning is an evidence of driver fatigue. This paper proposes to locate and track a driver's mouth movement using a CCD camera to study on monitoring and recognizing a driver's yawning. Firstly detecting drivers' faces uses Gravity-Center template, then detecting drivers' left and right mouth corners by grey projection, and extracting texture features of drivers' mouth corners (left and right) using Gabor wavelets. Finally LDA is applied to classify feature vectors to detect yawning. The method is tested on 400 images from twenty videos. In contrast, yawning is also detected by the ratio of mouth height and width. The experiment results show that Gabor coefficients are more powerful than geometric features to detect yawning and the average recognition rate is 95% which has more than 20% improvement.

Keywords:

Computer vision; Human fatigue; Gabor wavelet; Discriminant analysis

1. Introduction

Driver fatigue is one of the main reasons causing traffic accidents. According to the National Highway Traffic Safety Administration (NHTSA) [1] estimates, 100 000 police-reported crashes are directly caused by driver fatigue each year, which results in an estimated 1550 deaths, 71 000 injuries, and $12.5 billion losses. In 2002, the National Sleep Foundation (NSF) [2] reported that 51% adult drivers had driven a vehicle while feeling drowsy and 17% had actually fallen asleep.

If drivers get a caution one second earlier than accidents, about 90% accidents can be prevented. Therefore, developing technologies for monitoring the driver fatigue is essential to accident prevention. When a driver fatigues, he will take many special features on his face and body, and yawning is an important evidence of fatigue.

When yawning, mouths open wide and the geometric features of the mouth change obviously. So most previous research about yawning detection focuses their methods on geometric features of the mouth. For example, Rongben Wang et al. [3] took the mouth region's geometric features to make up an eigenvector as the input of a BP ANN, and they acquired the BP ANN output of three different mouth states that represent normal, yawning or talking state respectively. Qiang Ji et al. [4] detected yawning by the openness of the mouth represented by the ratio of mouth height and width. Tiesheng Wang et al. [5] represented the openness of the mouth by the ratio of mouth height and width similarly, and detected yawning if the ratio is above 0.5 in more than 20 frames. More research on monitoring driver fatigue can be found in [6], [7].

There are some disadvantages in yawning detection using geometric features of the mouth. First, left and right mouth corners are obvious feature points, but the lip positions are difficult to detect precisely. At the same time, lips move more acutely, which makes the lip detection more difficult. Third, geometric features are liable to pose and have more difference for individual.

We propose a method using Gabor wavelets of mouth corners (left and right) and LDA to detect yawing. Firstly, detecting faces uses Gravity-Center template, then detecting drivers' mouth corners by grey projection, and extracting texture features of drivers' mouth corners using Gabor wavelets. Finally we use LDA to classify feature vectors to detect Yawning.

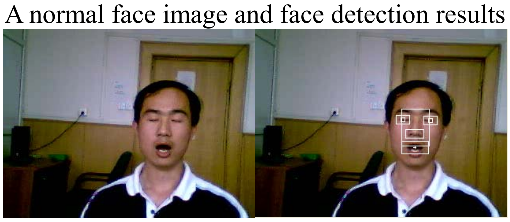
The paper is organized as follows. Section 2 introduces the face detection and mouth corner detection methods that we adopted. In section 3, mouth corner feature extraction and how to classify feature vectors and detect yawning are presented. Finally, experiment results and discussion are presented.

2. Face detection and mouth corner detection

2.1. Face detection

Face detection is the base of yawning detection. We use Gravity-Center template [8] to detect human faces. The method can significantly save the time consumed in rough detection of human faces in a mosaic image, and is able to detect rather slanted faces ($-25^\circ \sim 25^\circ$) and faces with much horizontal rotating angles ($-45^\circ \sim 45^\circ$) and vertical rotating angles ($-30^\circ \sim 30^\circ$), which are in various sizes and at unknown locations in an unconstrained background. Figure 1 illustrates the original images and the results of face detection.

A normal face image and face detection results



A yawning face image and face detection results

Figure 1. Original images and face detection results

2.2. Mouth corner detection

Face detection based on Gravity-Center template will produce the location information about faces and their organs such as eyebrows, eyes, noses and mouths, which is necessary for further facial feature detection. Once a face in an image is located, simultaneously, its facial organs can be also located roughly. Figure 1 illustrates the face detection results and the localization of face organs. The detail description on detecting procedures can be found in reference [8].

Before mouth corner (left and right) detection, we rotate images based on the positions of pupil centers to keep the line between the pupil centers horizontal. This will help the mouth corner detection because in most cases the line between the pupil centers and the line between the two mouth corners are parallel. Figure 2 illustrates the images before and after rotation based on the pupil centers.

Images before rotation

Images after rotation

Figure 2. Face images before and after rotation

The vertical position of the mouth corners can be determined by horizontal grey projection in the mouth area. The parameters x_1 and x_2 are the horizontal rang of the mouth area, and I(x, y) represents the grey value at position (x, y). We can get the vertical position of mouth corners by searching the minimum position of H(y).

$$H(y) = \sum_{x=x1}^{x2} I(x, y) \tag{1}$$

$$H(x) = \sum_{y=y1}^{y2} I(x, y) \tag{2}$$

Similarly, we can find the horizontal positions of mouth corners by vertical projection in the mouth area. Searching from the middle of x to the two ends, we get the two jumping positions of H(x) as the horizontal positions of mouth corners. Mouth corner detection results are illustrated in figure 3.

Figure 3. Mouth corner detection results (Mouth, horizontal projection, vertical projection, vertical projection gradient, mouth corners in yawning and normal states)

3. Mouth corner feature extraction and classification

In addition to the geometric change, the texture of the mouth corners has obvious change when yawning. We use Gabor wavelets to extract texture feature, and we take LDA as the classifier to detect yawning.

3.1. Two-dimensional Gabor wavelet representation

Gabor wavelet [9] is suitable for analyzing and processing texture images for better resolving capacity in space field and frequency field, so we select Gabor wavelet to extract drivers' mouth image features. The two dimensional Gabor function is shown as follows:

$$\varphi_j\left(\vec{x}\right) = k_j^2 \exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right)\exp\left(i\vec{k_j}\,\vec{x}\right) \quad (3)$$

$$\vec{k_j} = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos\phi_u \\ k_v \sin\phi_u \end{pmatrix} \quad (4)$$

$$k_v = 2^{-\frac{v+2}{2}}\pi \quad (5)$$

$$\phi_u = u\frac{\pi}{K} \quad (6)$$

Where, v determines the frequencies of the Gabor kernel, u determines the orientations of the Gabor kernel. For an image with the grey of $L\left(\vec{x}\right)$, at a given point $\vec{x} = (x, y)$, the Gabor transformation is as follows:

$$J_j\left(\vec{x}\right) = \int L\left(\vec{x}'\right)\varphi_j\left(\vec{x} - \vec{x}'\right)d^2\vec{x}' \quad (7)$$

In our system, we use a discrete set of Gabor kernels which comprise 5 spatial frequencies, i.e., scales, and 8 distinct orientations. We set K=8, v=0, 1, 2, 3, 4, u=0, 1, 2, 3, 4, 5, 6, 7. Each image is convolved with both even and odd Gabor kernels at the location of moth corners. We have therefore 40 complex Gabor wavelet coefficients at each mouth corner. In our study, only the magnitudes are used, because they vary slowly with the position while the phases are very sensitive. In summary, with Gabor wavelet coefficients, each mouth image is represented by a vector of 80 (40*2) elements.

3.2. Image classification based on LDA

The examples we consider can be treated using two class linear discriminant analysis [10], e.g., normal or yawning. Linear discriminant analysis seeks a single projection optimally separating the labeled classes in the training set, while minimizing variance within each projected class.

Given a train set $\{x_1, x_2, x_3, \cdots, x_N\}$ which belongs to c classes Ω_i, $i = 1, 2, \cdots, c$, where c is the number of classes. The mean of each class is defined as:

$$\mu_i = \frac{1}{N_i}\sum_{x_k \in \Omega_i} x_k, \quad i = 1, 2, \cdots, c \quad (8)$$

where N_i is the number of samples in the Ω_i class. The mean of the projected samples of the train set is defined as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (9)$$

The within-class covariance is defined as:

$$\Sigma_i = \frac{1}{N_i}\sum_{x_k \in \Omega_i}(x_k - \mu_i)(x_k - \mu_i)^T \quad (10)$$

Scatter within-class and between-class, are defined, respectively, as formula (11) and (12), where $\Pr(\Omega_i)$ is the priori probability of Ω_i class.

$$S_w = \sum_{i=1}^{c}\Pr(\Omega_i)\Sigma_i \quad (11)$$

$$S_b = \sum_{i=1}^{c}\Pr(\Omega_i)(\mu_i - \mu)(\mu_i - \mu)^T \quad (12)$$

LDA seeks a projection direction, defined by vector *W*, along which the ratio of the between-class scatter to the within-class scatter defined as formula (13) is maximized.

$$J(W) = \frac{\left|W^T S_b W\right|}{\left|W^T S_w W\right|} \quad (13)$$

The optimal projection matrix is:

$$W_{opt} = \arg\max_W \frac{\left|W^T S_b W\right|}{\left|W^T S_w W\right|} \quad (14)$$

To classify an input feature vector, it is projected along the corresponding discriminant matrix W_{opt} calculated from training examples. The distance to each class center or sample is calculated, and the input sample is classified as a member of the nearest class or belongs to the same class with the nearest example.

4. Experiments

4.1. Face detection results

We test our face detection method on twenty videos, and there are 4512 faces in them. The results are illustrated at table 1. We can see that our method not only can detect faces almost in a real time, but also the accuracy is very high at about 94.7%. In our experiment circumstance, very few non-face objects are detected as faces, and the FAR is 3.4%. The face detection method is suitable for our research.

Table 1. Face detection results (FAR: False Acceptance Rate, FRR: False Rejection Rate)

face number	FAR	FRR	correct rate
4512	3.4%	5.3%	94.7%

4.2. Yawning detection results

From the twenty videos, we select 200 yawning images and 200 normal images as the data set. From each video, we select 10 yawning images and 10 normal images. We test our method twice, and at each time we select half of the images as probe set and the other half as training set for LDA. We extract the Gabor coefficients of the left and right moth corners as mouth features. Nearest mean classifier is adopted to detect yawning, and we get an average recognition rate of 95%. The correct rate of normal images is 94%, and that of yawning images is 96%. The results of classification are presented in table 2.

In contrast, we also detect yawning by ratios of mouth height and width. First we detect the four corners (left, right, top and bottom) of the mouth as how we detect the left and right mouth corners, and compute the ratios. Second, we get a threshold of ratios from half of the data set. At last, if the ratio of the mouth is above the threshold in a novel image from the other half, we judge that the object in the image is yawning. Then, we change the probe set and training set, and make another test. The mouth corner detection results are showed in figure 4. The results of classification are also presented in table 2.The recognition rate of normal state is 74.5%, and that of yawning state is 69.5%. In total, the average recognition rate is 72%.

Figure 4. Mouth corner detection results

Table 2. Yawning detection results

method	state	normal	yawning	correct rate
geometric feature	normal	149	51	74.5%
	yawning	61	139	69.5%
Gabor feature	normal	188	12	94%
	yawning	8	192	96%

5. Conclusions

We propose a method to locate and track a driver's mouth corners using Gravity-Center template face detection and grey projection, and extract texture features of drivers' mouth corners using Gabor wavelets for yawning detection. The feature points, left and right mouth corners, are easier to detect and the features we extract are more robust to pose and individual. Finally, we use LDA to classify the feature vectors to detect yawning. The experiment results show that Gabor coefficients are more powerful than geometric features and get more than 20% improvement in the recognition rate. The Gabor features of mouth corners can represent yawning correctly, and our method will help to make drive safer.

In future, we will capture more videos to train and test our method. Our goal is to develop a system to fuse more features including mouth features, eye features and image sequence features to monitor driver fatigue.

Acknowledgements

This research is partially sponsored by National Natural Science Foundation of China (No.60533030), Beijing Natural Science Foundation (No.4061001), PHR (IHLB).

References

[1] NHTSA, "Drowsy driver detection and warning system for commercial vehicle drivers: Field proportional test design, analysis, and progress", National Highway Traffic Safety Administration, Washington, DC, [Online], Available: http://www.nhtsa.dot.gov/.

[2] NSF, "Sleep facts and stats", National Sleep Foundation, Washington, DC, [Online], Available: http://www.sleepfoundation.org/.

[3] Rongben Wang, Lie Guo, Bingliang Tong, Lisheng Jin, "Monitoring mouth movement for driver fatigue or distraction with one camera", The 7th International IEEE Conference on Intelligent Transportation Systems, pp. 314-319, Oct. 2004.

[4] Qiang Ji, Zhiwei Zhu, P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue", IEEE Transactions on Vehicular Technology, Vol 53, No. 4, pp. 1052-1068, July 2004.

[5] Tiesheng Wang, Pengfei Shi, "Yawning detection for determining driver drowsiness", Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology, pp.373-376, May 2005.

[6] A. Williamson, T. Chamberlain, “Review of on-road driver fatigue monitoring devices”, NSW Injury Risk Management Research Centre, University of New South Wales, [Online], Available: http://www.maa.nsw.gov.au/, April, 2005.

[7] Qiang Ji, P. Lan, C. Looney, “A probabilistic framework for modeling and real-time monitoring human fatigue”, IEEE Transactions on Systems, Man and Cybernetics, Part A, Vol 36, No. 5, pp. 862–875, Sept. 2006.

[8] Jun Miao, Baocai Yin, Kongqiao Wang, Lansun Shen, Xuecun Chen, “A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template”, Pattern Recognition, Vol 32,No. 7, pp. 1237-1248, 1999.

[9] L. Wiskott, J.M. Fellous, N. Kuiger, C. von der Malsburg, “Face recognition by elastic bunch graph matching”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 19, No. 7, pp. 775–779, July 1997.

[10] A.M. Martinez, A.C. Kak, “PCA versus LDA”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 23, No. 2, pp. 228–233, Feb. 2001.