Online Vigilance Analysis Combining Video and Electrooculography Features

Ruo-Fei Du[1], Ren-Jie Liu[1], Tian-Xiang Wu[1] and Bao-Liang Lu[1,2,3,4,⋆]

[1] Center for Brain-like Computing and Machine Intelligence
Department of Computer Science and Engineering
[2] MOE-Microsoft Key Lab. for Intelligent Computing and Intelligent Systems
[3] Shanghai Key Laboratory of Scalable Computing and Systems
[4] MOE Key Laboratory of Systems Biomedicine
Shanghai Jiao Tong University
800 Dongchuan Road, Shanghai 200240, China
bllu@sjtu.edu.cn

Abstract. In this paper, we propose a novel system to analyze vigilance level combining both video and Electrooculography (EOG) features. For one thing, the video features extracted from an infrared camera include percentage of closure (PERCLOS) and eye blinks, slow eye movement (SEM), rapid eye movement (REM) are also extracted from EOG signals. For another, other features like yawn frequency, body posture and face orientation are extracted from the video by using Active Shape Model (ASM). The results of our experiments indicate that our approach outperforms the existing approaches based on either video or EOG merely. In addition, the prediction offered by our model is in close proximity to the actual error rate of the subject. We firmly believe that this method can be widely applied to prevent accidents like fatigued driving in the future.

Keywords: Vigilance Analysis, Fatigue Detection, Active Shape Model, Electrooculography, Support Vector Machine.

1 Introduction

There is no doubt that vigilance plays a crucial role in our daily life. It is essential to ensure the vigilance level of the drivers since a large number of accidents are resulted from fatigued driving, which has been investigated by [1] and [2]. Meanwhile, vigilance is an indispensable characteristic for other occupations such as policemen, soldiers and operators who have to deal with hazardous equipments. Thus, an efficient and accurate system is badly in need in order to prevent accidents by warning the users in advance.

In the last two decades, extensive researches have been conducted regarding vigilance analysis [3]. These studies can be broadly classified into 3 categories based on the techniques which are adopted during the procedure of feature extraction: (infrared) video, electrooculography (EOG) and electroencephalography (EEG).

Firstly, video is the most convenient approach. Compared with EOG and EEG based systems, in which the subjects have to interact directly with the equipments, cameras

⋆ corresponding author.

T. Huang et al. (Eds.): ICONIP 2012, Part V, LNCS 7667, pp. 447–454, 2012.

are much less intrusive. Moreover, not only eye movement can be measured through video, but yawn state and facial orientation can be estimated [4][5][6]. Unfortunately, there are 3 apparent drawbacks for video approach: the accuracy would decrease due to various luminance; the range of the video is limited to the horizon of the cameras; the recognition usually fails on account of the appearance of the subjects such as wearing sunglasses.

Secondly, EOG signals is a moderate method since it is irrelevant to the environment and easier to analyze. In order to estimate the vigilance level, features like eye blinks, blink duration, slow eye movement (SEM), rapid eye movement (REM) are utilized, which has been proved to be accurate according to [7].

Finally, vigilance analysis based on EEG signals gains a good accuracy. It has been shown that both delta waves (0-4 Hz), which are relevant to slow wave sleep (SWS), and theta waves (4-7 Hz), which are relevant to drowsiness of older children and adults play significant roles in vigilance analysis [3]. Nevertheless, it is ineffective as a result of its poor user experience of the interaction.

From our perspective, in order to utilize comprehensive information like yawn state and improve the accuracy of vigilance analysis, we construct an online vigilance analysis system combining both video and EOG features. On one hand, we extract the displacements of feature points around eyes and mouth based on the well-known Active Shape Model (ASM) and the location of the eyelids based on a binarization approach. Afterwards, we calculated the average height, area and PERCLOS of eyes and mouth. Finally we extract features of blinks and eye movements for vigilance analysis from the information we have. On the other hand, we preprocess the EOG signals using a low-pass filter with the frequency of 10Hz and a normalization procedure. After that, features of blinks and eye movements are extracted according to the algorithm proposed in [8]. In our experiments, based on the actual error rate signals, we employ SVM regression to offer the prediction of the vigilance level. Compared with the actual error rate, our algorithm is proved to work rather well for the online vigilance estimation. In addition, although the accuracy provided by the video features is not so good as that of EOG features, the combination of both video and EOG features outperform both models running alone.

In Section 2, our online vigilance analysis system is introduced in details. In section 3, the preprocess in our system is illustrated. We describe the approaches of features extraction in Section 4. Experiments and analysis are conducted in section 5, followed by the conclusion and further discussions in section 6.

2 Vigilance Analysis System

2.1 Equipments

In our vigilance analysis system, the video signals are extracted from infrared cameras in the front. Meanwhile, the EOG signals are extracted by 8 electrodes and recorded by the NeuroScan system [1].

[1] Neuroscan Inc, Herndon, VA, USA.

There are 2 vertical and 2 horizontal electrodes on the forehead. Different traffic signs in four colors (red, green, blue and yellow) are displayed on the screen every 6 seconds and each sign lasts for only 0.5 seconds. Meanwhile, the subject is ought to press the button with the same color as the signs flashed on the screen. Therefore, not only the actions of the subjects could be recorded but the actual error rate could be calculated for further analysis.

2.2 Overall Process

The overall process of our vigilance analysis system is illustrated in Fig. 1. In summary, four procedures including preprocessing, feature extraction, model training and prediction of vigilance are followed for both video and EOG signals.

3 Preprocess

The video signals are captured by infrared cameras which is in front of the subject. The recorded video has the resolution of 640×480 pixels with 30 frames per second in RGB color. In order to extract features based on video, we need to preprocess the images in each frame. For a single test of 67 minutes, the raw data is approximately 70 GB. For each frame, we apply the Active Shape Model [9]. After acquiring landmarks of eyelids, eyeballs and mouth, we employ an extraction algorithm to get blink, eye movement and yawn information.

3.1 Face Recognition

Firstly, we employ a cascade Adaboost classifier based on Haar-like features, which is proposed by [10], for the face recognition procedure. If multiple faces are successfully recognized in the frame, the one with the largest area is regarded as the subject's face.

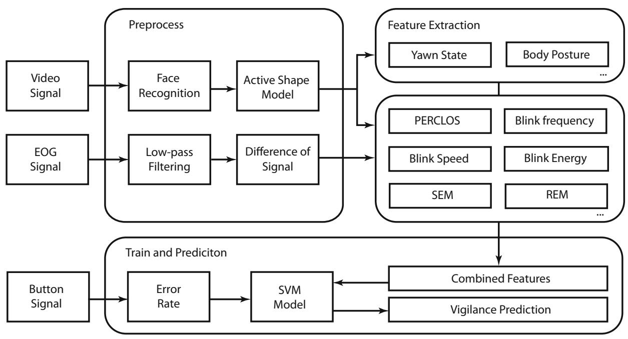

Fig. 1. The proposed vigilance analysis system by combining video and EOG features

3.2 Active Shape Model

After getting the approximate location of the subject in the image, we adopt the Active Shape Model (ASM), which is a statistical model to recognize the shape of deformable object, to get the displacements of both eyes and mouths [9]. Generally speaking, ASM is used to locates landmarks on the image by fitting the parameters of the shape using optimization techniques like gradient descent. The shape S provided by ASM is consisted of n landmarks, explicitly $S = [x_1, y_1, x_2, y_2, ..., x_n, y_n]$, where x_i, y_i are the coordinates of the i_{th} landmark. In our vigilance system, there are totally 68 landmarks for the ASM Model. The distance $D(i, j)$ between landmarks i and j is defined as $D(i, j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. Furthermore, we calculate the current average height of both eyes and that of the mouth according to: $H_e = \frac{D(35,33)-D(34,32)}{2}$ and $H_m = D(64, 57)$. Similarly, the approximate areas A_e, A_m are calculated according to the polynomial shape of the eyes and mouth.

3.3 Preprocess of EOG Signals

Firstly, in order to extract blink features from the EOG signals, we filter the vertical EOG signal by a low-pass filter with a frequency of 10Hz using EEGLab [11]. Afterwards, a multiplier is used to adjust the amplitude of the signals. In order to obtain the variance ratio from the EOG signals, we compute the difference of signals for blink feature extraction. Denote D as the difference signal, V as the signal and R as the sampling rate, we have $D(i) = (V(i + 1) - V(i)) \times R$.

4 Feature Extraction

In this section, we will briefly introduce how we extract the features based on video and EOG signals. Both kinds of features are extracted based on a time window of 8 seconds from a 67-minutes test.

4.1 Video Features

The main features we extracted from the video include PERCLOS, blink frequency, eye movement and yawn frequency.

- PERCLOS of eyes
  The PERCLOS of eyes, which refers to the percentage of eye closure, is an efficient feature to estimate vigilance [12] defined as follows:

$$PERCLOS_e = \frac{\overline{H}_e - H_e}{\overline{H}_e}$$

  where $\overline{H}_e$ indicates the average open eye height above a certain threshold. Another PERCLOS feature is calculated according to the areas of eyes. Finally, the proportions of fully-closed and half-closed eyes and mouth during a certain time window are also regarded as PERCLOS features.

- Blink frequency
  The frequency of blinks also has a strong relationship to vigilance. We setup four thresholds H_{c1}, H_{c2}, H_{o1}, H_{o2}, , which indicate the relative height when eyes are about to close, already closed, about to open and fully open. This procedure suggests a complete blink. The times of eye blinks during a time window is calculated as vigilance features.
- Eye Movement
  The relative position of the pupil can be recognized by ASM. Thus the moving frequency of eyes are recorded as an important feature. During each time window, we calculate the movement of eye pupils and its the amplitude. The speed of the eye movement is also calculated as a feature.

- Yawn frequency
  Since an action of yawn suggest significantly that the subject has already been fatigued. The window size w for yawn state should be large enough such as 16 seconds $\times$ 30 frames. Denote the average of the least k heights of mouth as H_m^k

$$Y_i = \frac{\sum_{j=i-w}^{i} (H_j / H_m^k) > C}{w}$$

  Here C is a threshold and indicates the ratio between open mouth height and normal mouth height when the subject is about to yawn.

- Body Posture
  We estimate the posture of body by locating the relative position of eyes, nose and mouth on the face. Denote orientation of the face as α, θ and β, corresponding to different reference of eyes, nose and mouth. This degree can be calculated as follows:

$$\alpha = \frac{D(67,2)}{D(67,12)}; \theta = \frac{D(31,0)}{D(36,14)}; \beta = \frac{D(66,3)}{D(66,11)}$$

  where points 67, 66, 31, 36 denote the center of the nose, mouth, left and right pupil separately while the others indicate the left and right side of the face horizontally and correspondingly.

4.2 EOG Features

- Blink Features
  After the preprocess of EOG signals, every blink is marked at four time points $c1, c2, o1$ and $o2$, which indicate the time when the eye is to close, closed, to open and opened. Denote V as the signal, D as the difference of signal, we have the following features:

$$T_{blink} = T_{o2} - T_{c1}; \quad T_{close} = T_{c2} - T_{c1}$$

$$T_{open} = T_{o2} - T_{o1}; \quad T_{closed} = T_{o2} - T_{c2}$$

$$S_{close} = \frac{\sum_{i=T_{c1}}^{T_{c2}} D_i}{T_{close}}; \quad S_{open} = \frac{\sum_{i=T_{o1}}^{T_{o2}} D_i}{T_{open}}; \quad E_{blink} = \sum_{i=T_{c1}}^{T_{o2}} V_i^2$$

where T indicates the time during a window size, S indicates the speed, and E indicates the energy of blinks.

– Eye Movements
  Two kinds of eye movements, Slow Eye Movement (SEM) and Rapid Eye Movement (REM) are extracted, according to different kinds of time threshold in [13]. In order to get these features more accurately, two methods of Fourier transformation and wavelet transformation are used. In the Fourier transformation method, we use a band-pass filter with frequency 0.5Hz and 2Hz to process the horizontal EOG signal. The sampling rate is 125Hz and the period is 8 seconds.

4.3 Linear Dynamic System

Considering the fact that both video and EOG features introduce much noise, we adopt the linear dynamical system, which is proposed in [14] to process them. As an unsupervised learning method, LDS can increase the main component of the features and reduce the noise, leading to a higher correlation with vigilance.

Finally, all the features are normalized between 0 and 1. Afterwards, the features are eventually used for training and prediction.

5 Experiments

There are totally five healthy subjects for our experiments, including four men and 1 woman, all of whom are around 23 years old. Particularly, we ensure that none of the subjects is color-blind. Each subject is asked to have sufficient sleep at night and get up early in the morning. Each experiment is conducted after lunch and lasts for 67 minutes so that the subject behaves sober at first and sleepy after a period of about half an hour in the experiment. The room is in silence and the light is soft.

As is stated in Section 2, the subject is asked to press the button with the same color as the traffic sign displayed on the screen. With the help of NeuroScan Stim Software, we can collect error rate data and EOG signals, as well as video images from the infrared camera.

We use LibSVM [15] to train and test data and use the square correlation coefficient and mean squared error to evaluate the model. The parameters of the SVM Model are selected as

$$s = 3(\epsilon - \text{SVR}), \quad t = 2(\text{RBF kernel}), \quad c = 8, \quad g = 1/64, \quad p = 1/1024$$

Data for each subject is 400 points long. It is divided into 2 parts with the same length. After the data is divided, the first part is used as the testing set while the other one is used as the training set. Fig. 2 indicates the example of prediction result for the second part of the subject 1.

The correlation and squared error are displayed in Table 1.

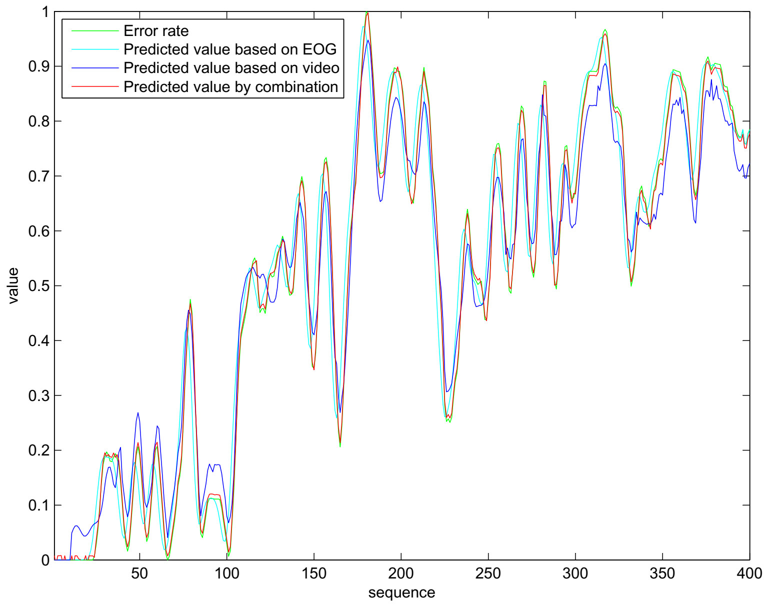

Fig. 2. Example of comparison between error rate and different prediction methods

Table 1. Squared correlation coefficient and Mean squared error of regression result

Subject	Video-based	EOG-based	Combinination
1	0.731/0.0256	0.843/0.0136	**0.852/0.0117**
2	0.778/0.0129	0.892/0.0064	**0.919/0.0170**
3	0.750/0.0151	0.866/0.0148	**0.882/0.0111**
4	0.750/0.0175	0.929/0.0091	**0.937/0.0045**
5	0.756/0.0170	0.809/0.0051	**0.921/0.0072**
Average	0.752/0.0882	0.88/0.0098	**0.898/0.0089**

6 Conclusions and Future Work

In this paper, we have proposed a novel system for vigilance analysis based on both video and EOG features. From the experimental results, we can arrive at the conclusion that our new system offers a good prediction of the actual vigilance level. Moreover, this method outperforms the existing approaches using either video features or EOG features alone, since our proposed method utilizes both the accuracy of EOG signals but the yawn state and body postures provided by video as well. In the future, we plan to utilize comprehensive features including depth information and grip power to get a better performance. Besides, more experiments will be conducted and the stability and robustness of the algorithms are expected to be improved.

Acknowledgments. This work was partially supported by the National Basic Research Program of China (Grant No. 2009CB320901).

References

1. May, J., Baldwin, C.: Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. Transportation Research Part F: Traffic Psychology and Behaviour 12(3), 218–224 (2009)
2. Stutts, J., Wilkins, J., Vaughn, B.: Why do people have drowsy driving crashes. A Foundation for Traffic Safety, 202-638 (1999)
3. Wang, Q., Yang, J., Ren, M., Zheng, Y.: Driver fatigue detection: a survey. In: The Sixth World Congress on Intelligent Control and Automation, WCICA 2006, vol. 2, pp. 8587–8591 (2006)
4. Ji, Q., Yang, X.: Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. Real-Time Imaging 8(5), 357–377 (2002)
5. Dinges, D., Mallis, M., Maislin, G., Powell, I., et al.: Final report: Evaluation of techniques for ocular measurement as an index of fatigue and as the basis for alertness management. National Highway Traffic Safety Administration (1998)
6. Fletcher, L., Apostoloff, N., Petersson, L., Zelinsky, A.: Vision in and out of vehicles. IEEE Transaction on Intelligent Transportation Systems 18(3), 12–17 (2003)
7. Ma, J., Shi, L., Lu, B.: Vigilance estimation by using electrooculographic features. In: Proceedings of 32nd International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6591–6594 (2010)
8. Wei, Z., Lu, B.: Online vigilance anaysis based on electrooculography. In: The 2012 International Joint Conference on Neural Networks, pp. 1–7 (2012)
9. Cootes, T., Taylor, C., Cooper, D., Graham, J., et al.: Active shape models - their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
10. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of Computer Vision and Pattern Recognition (2001)
11. Delorme, A., Makeig, S.: Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. Journal of Neuroscience Methods 134(1), 9–21 (2004)
12. Dinges, D., Grace, R.: Perclos: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. Federal Highway Administration. Office of Motor Carriers, Tech. Rep. MCRT-98-006 (1998)
13. Bulling, A., Ward, J., Gellersen, H., Troster, G.: Eye movement analysis for activity recognition using electrooculography. IEEE Transactions on Pattern Analysis and Machine Intelligence (99), 1–1 (2011)
14. Shi, L., Lu, B.: Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning. In: Proceedings of 32nd International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6587–6590 (2010)
15. Chang, C., Lin, C.: Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2 (2011)