Detecting Driving Fatigue with Multimodal Deep Learning

Li-Huan Du, Wei Liu, Wei-Long Zheng and Bao-Liang Lu* *Senior Member,IEEE*

***Abstract*— Physiological signals such as EEG and EOG have been successfully applied to detect driving fatigue in single modality. In this paper, we propose a multimodal approach by combining partial EEG and forehead EOG to enhance driving fatigue detection. We investigate the key brain area where we collect the EEG to combine with forehead EOG. Our experiment results demonstrate that the temporal EEG signals from six-channel have the best performance when combining with forehead EOG to extract shared features. Furthermore, we propose a novel multimodal fusion strategy using deep autoencoder model to learn a better shared representation. We assess our approach with other fusion strategies on 21 subjects. Our multimodal approach achieves the best performance that the average COR and RMSE are 0.85 and 0.09, respectively. The experiment results demonstrate that our multimodal approach could learn an efficient shared representation between modalities and could significantly improve the performance of detecting driving fatigue.**

I. INTRODUCTION

Over the past decades, physiological signals have been widely used to detect the complex states of human beings [1]. Compared with traditional video and voice signals, the performance of physiological signals is more accurate and stable. Various methods have been proposed to detect the level of driving fatigue [2], [3], [4].

Among the various methods, electroencephalogram (EEG) is considered to be a promising one [5]. In recent years, some EEG-based methods have been proposed to detect vigilance [3]. Electrooculogram (EOG) is another widely used promising physiological signals. EOG signals are recorded by four electrodes around eyes [6]. Compared with EEG, the amplitude of EOG is relatively higher, which makes it more robust to noise. Many previous work shows the features extracted from EOG have a significant correlation with the fatigue status [5]. Some specific features have been found as available predictor to estimate the level of fatigue, such as slow eye movements (SEM), blinks and other specific behaviours of eye and eyelid [2]. When using EOG-based method to estimate the level of driving fatigue, the traditional electrode placements of EOG are closely around eyes, which may lead to uncomfortable and inconvenient for practical applications. Forehead EOG is a novel approach and has been demonstrated to be effective [7].

Since fatigue is a complex state, it is difficult to build an effective and robust system to detect driving fatigue with only one kind of physiological signals. Previous work points out that combining multiple electrophysiological signals to detect driving fatigue or drowsiness is a more efficient approach [8]. Some fusion strategies have been proposed to combine EEG and EOG, such as decision level fusion (DLF), feature level fusion (FLF) [9]. Other approach incorporates the temporal dependency of vigilance into model training [10].

In this paper, we adopt a multimodal deep autoencoder model using EEG and EOG signals to detect driving fatigue [11]. This approach has been successfully applied in EEG-based emotion recognition [12]. In this work, we capture EEG signals from different brain areas: temporal and posterior. We then discuss the performance of each brain area when combining with forehead EOG to estimate driving fatigue. Our results demonstrate that multimodal learning method could significantly improve the accuracy of driving fatigue detection. Fusion of temporal EEG with forehead EOG could extract a better shared representation between different modalities.

II. METHODS

A. Data Preprocessing

The subjects were asked to drive a car in our simulated driving environment. The experiment simulated complex traffic situations. The subjects were required to drive for at least two hours to make sure them fall into driving fatigue state. We used NeuroScan system to record forehead EOG and EEG signals. The signals were recorded by the NeuroScan system at a sampling rate of 1000 Hz. We then adopted a 1-75 bandpass filter for EEG signals to reduce the drift and artifacts. To reduce the computational complexity, EEG signals were down-sampled to 200 Hz. For EOG, we down-sampled the signals to 125 Hz and used a 0-30 bandpass filter to reduce the noise signals.

B. Feature Extraction

EEG Signals: According to our previous work [3], 12-channel EEG signals from posterior site (CP1, CPZ, CP2, P1, PZ, P2, PO3, POZ, PO4, O1, OZ, O2) and 6-channel from temporal site (FT7, T7, TP7, FT8, T8, TP8) are critical brain areas when detecting fatigue status, which are shown in Fig.1:

This work was supported in part by grants from the National Natural Science Foundation of China (Grant Nos.61272248 and 61673266), the National Basic Research Program of China (Grant No.2013CB329401), the Major Basic Research Program of Shanghai Science and Technology Committee (Grant No.15JC1400103), and the Technology Research and Development Program of China Railway Corporation (Grant No.2016Z003-B).

Li-Huan Du, Wei Liu, Wei-Long Zheng and Bao-Liang Lu are with the Center for Brain-Like Computing and Machine Intelligence, Department of Computer Science and Engineering, the Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering and the Brain Science and Technology Research Center, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China.

*Corresponding author (bllu@sjtu.edu.cn)

Power spectral density (PSD) and differential entropy (DE) are two commonly used features extracted from EEG. PSD is a widely accepted features when analysing EEG signals which shows high relationship with fatigue and vigilance [4]. When the length of EEG is fixed, DE is considered as a better feature compared to the PSD [13]. The frequency band of EEG signals varies from 1 Hz to 50 Hz and is divided into five frequency bands: δ (1-4 Hz), θ (4-8 Hz), α (8-14 Hz), β (14-31 Hz) and γ (31-50 Hz). To get more specific details, we divide the EEG signals into 25 continuous bands with a 2 Hz frequency resolution. Short-term Fourier transform (STFT) is adopted to compute the different entropy features with a 8 s non-overlapping window.

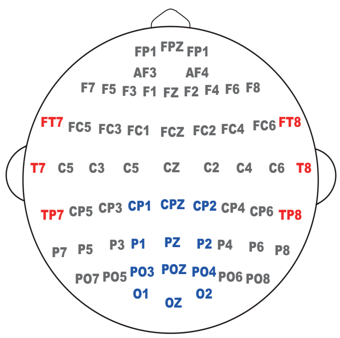

Fig. 1. 12-channel EEG signals from posterior site (in blue) and 6-channel EEG signals from temporal site (in red)

EOG Signals: EOG signals are demonstrated to have statistical significance [5], so we apply a fixed 8 s non-overlapping window to get the features. In traditional EOG-based experiments, signals are recorded from two pairs of electrodes closely around eyes numbered one to four in Fig.2(a). In this work, all EOG features are extracted from forehead EOG electrodes [7] numbered five to eight in Fig.2(b).

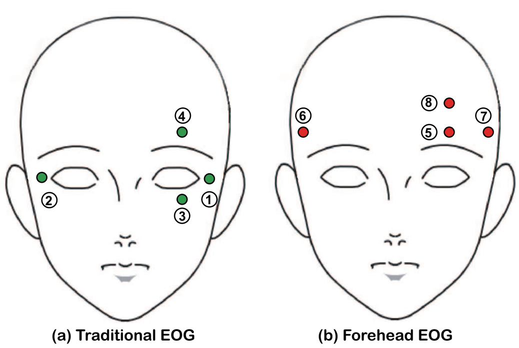

Fig. 2. Traditional and forehead EOG electrodes placements

For traditional EOG, we calculate the absolute value of substraction by the two pairs electrodes to get vertical electrooculogram (VEO) and horizon electrooculogram (HEO). However, for forehead EOG, a blind source separation method called independent component analysis (ICA) is proved to be more efficient to extract forehead VEO while subtraction is still used to calculate forehead HEO [7]. We utilize Mexican mother wavelet to compute the continuous wavelet to extract more accuracy forehead EOG features. Because of the sensitive of singularity, blink and saccade features are extracted more clearly by wavelet transform comparing with other derivative methods. We then use peak detection algorithm to extract saccade and blink features on the continuous wavelet. To avoid unnecessary computation, we detect blink features from forehead VEO while saccade features were extracted from forehead HEO. When we already obtain saccade, blink and fixation (the duration of blink or saccade), we then calculate useful components of the three features. While some features show a insignificant correlation with the fatigue states, we do not use them. The 36 features we used are shown in Table 1. We then use EOG to represent forehead EOG rather than traditional EOG in the following paper. All the EEG and EOG features are normalized to the range of [0, 1].

TABLE I
TOTAL 36 FEATURES EXTRACTED FROM EOG

Group	Features
blink	maximum/mean of blink rate variance/amplitude variance maximum/mean/sum of blink rate blink number maximum/minimum/mean of blink amplitude power/mean power of blink amplitude
saccade	maximum/mean of saccade rate variance/saccade amplitude variance maximum/minimum/mean of saccade rate/saccade amplitude power/mean power of saccade amplitude saccade number
fixation	mean/maximum of blink duration variance/saccade duration variance maximum/minimum/mean of blink duration/saccade duration

C. Fatigue Labeling

How to label the level of fatigue is still a serious problem when constructing driving fatigue detection model using supervised learning paradigm. Because the ground truth of mental status varies from time to time. PERCLOS is a widely used method to annotate the vigilance. In traditional fatigue detection method, PERCLOS is calculated with facial video, which may severely affected by environment change. A novel approach use eye tracking glasses to record detailed eye movement data. The SMI glasses we used could capture the expected eye status, i.e. fixation, blink and saccade. CLOS is another statu defined as duration of long-time eyelid closure and slow closures. PERCLOS index can be calculated as follows:

$$PERCLOS = \frac{blink + CLOS}{interval} \tag{1}$$

$$interval = blink + saccade + fixation + CLOS \tag{2}$$

D. Fusion Strategies

To utilize different modalities, some fusion strategies are widely used [9], such as decision level fusion (DLF) and feature level fusion (FLF). For DLF, data from different modalities are trained in different regression models and the final value is the maximum one picked from all the results calculated by regression models. For FLF, feature vectors from different modalities are concentrated into a larger feature vector. These fusion strategies are demonstrated to be efficient. However, the traditional strategies is often difficult to learning the shared representation between modalities.

We introduce the Deep Autoencoder model [11] as a multimodal approach to extract the shared representation:

RBM: The Restricted Boltzmann Machine (RBM) is a network with two layers connected by a symmetrical weighted matrix. The first layer is the set of visible binary units(v) while hidden binary units (h) in second layer. The energy of this RBM can be calculated by the joint configuration (v, h):

$$E(v,h) = -\sum_{i \in v} a_i v_i - \sum_{j \in h} b_j h_j - \sum_{i,j} v_i h_j w_{ij} \quad (3)$$

We can calculate the probability of connection between a hidden unit and a visible by the definition:

$$p(v,h) = \frac{1}{Z} e^{-E(v,h)} \quad (4)$$

Let Z represents the sum of all possible visible-hidden unit pairs:

$$Z = \sum_{v,h} e^{-E(v,h)} \quad (5)$$

Contrastive divergence algorithm is widely used to train RBMs which uses some tricks to speed up the sampling process compared with Gibbs Sampling [14].

Deep Autoencoder: When extracting multimodal features from EEG and EOG, a direct approach is to train a RBM using the concatenated EEG and EOG features. However, this approach doesn't work well for detecting the driving fatigue. Because EEG and EOG signals share highly non-linear correlations which make it hard for a shallow RBM model to learn the features mapped to across modality. If using single RBM to train multimodal data, we can find that only visible units and hidden units which from the same modalities have strong connections.

To get a better multimodal representation, we utilize some pre-trained RBMs with single modalities and then use the hidden units as input to train a RBM which represents high-dimension correlations between modalities:

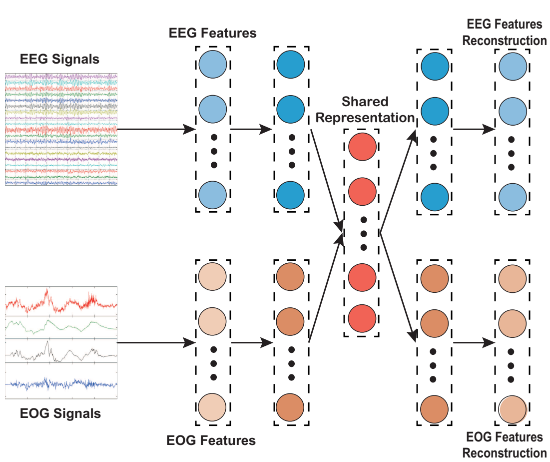

Fig. 3. The procedure of multimodal deep autoencoder model

In this multimodal Deep Autoencoder model (DAEabbreviated), the posterior of the hidden units of EEG and EOG are:

$$p(h_j|v) = sigmoid(\frac{1}{\sigma^2}(b_j + w_j^T v)) \quad (6)$$

The distribution of the EEG-based and EOG-based posteriors are inputted into the second layer as training data. We use contrastive divergence algorithms to fine-tune the symmetrically weighted connections w. For the decoding process, we unfold the two-layer RBMs to reconstruct EEG and EOG features. The weighted matrix is assigned to be the transposed form of encoding process.

III. Experiment Setup

We evaluated our methods on experimental results of 21 subjects including 12 men and 9 women. We recorded 6-channel temporal EEG, 12-channel posterior EEG and 4-channel forehead EOG simultaneously by the NeuroScan System. At the same time, PERCLOS was calculated with the eye movement data recorded by SMI ETG glasses.

In this paper, we put more attention on the performance of feature fusion strategies, Thus Support Vector Regression (SVR) with Radial Basis Function (RBF) kernels is used as a regression model. We then combine EOG with temporal, posterior and all EEG signals separately using DAE model which are abbreviated as T + EOG, P + EOG and ALL + EOG, respectively. We utilize the three types of shared representation to detect driving fatigue with PERCLOS label.

To demonstrate the efficiency of our fusion strategy, we train two unimodal models with EEG and EOG and use the results as baseline. Then we apply DLF, FLF and DAE as fusion strategies to get the shared features of EEG and EOG. Finally, SVR is used to calculate the level of fatigue. To avoid over fitting, we adopt cross validation with 5-fold to separate the training data and test data.

IV. Results and Discussion

A. Feature Fusion with EOG and Partial EEG

In general, to detect the driving fatigue is a regression problem, so we use Correlation Coefficient (COR) and Root Mean Square Error (RMSE) calculated by our prediction values and PERCLOS label as metrics.

If we only use EEG to detect fatigue status, posterior area was demonstrated to have the most obvious relationship with fatigue level [3]. However, when combining with EOG, it shows totally different results:

TABLE II
The average and standard deviations of COR and RMSE for the Predictions with Different Features

Multimodal Features		P + EOG	T + EOG	ALL + EOG
COR	Ave.	0.836	**0.852**	0.845
	Std.	0.079	0.064	0.068
RMSE	Ave.	0.098	**0.094**	0.096
	Std.	0.018	0.017	0.018

The result shown in Table II illustrates that fusion of temporal EEG with EOG achieves better performance than other areas. It only uses 6-channel electrodes to get relatively accurate predictions. The COR and RMSE of the best result are 0.85 and 0.09. This approach is practical in real applications because we could collect more signals with other sensors simultaneously to build a more robust system.

B. *Performance of Fusion Strategies*

In this section, we compare the performance of unimodal features extracted only from EEG or EOG. We also show the performance of DLF, FLF and DAE fusion strategies.

In Table III, we note that between the two single modalites, EOG shows a better performance than EEG with a higher COR (0.78) and a lower RMSE (0.12). It means EOG is a promising physiological signals to detect driving fatigue.

While using EOG features alone performs relatively well for detecting driving fatigue, fusing EOG with EEG shows a better performance, especially when the fusion features are learned by DAE model. In our experiment, using fusion strategies has been proved to be an effective way to enhance the accuracy of driving fatigue detection. For DLF, the prediction values are calculated by the outputs of single modalities. The training process only utilizes unimodal data and thus can improve the performance slightly.

TABLE III

PERFORMANCES OF UNIMODAL STRATEGIES AND MULTIMODAL FUSION STRATEGIES

Features		Unimodal		Multimodal		
		EEG	EOG	DLF	FLF	DAE
COR	Ave.	0.701	0.778	0.782	0.803	**0.852**
	Std.	0.126	0.115	0.096	0.085	0.064
RMSE	Ave.	0.133	0.118	0.114	0.116	**0.094**
	Std.	0.025	0.025	0.023	0.020	0.017

While FLF and DAE are both fusion on feature level, we note that DAE performs better with higher accuracies and lower standard deviations. The COR and RMSE of DAE are 0.85 and 0.09, respectively. The data from different modalities have non-linear correlations, thus concatenating features directly may lose the hierarchical information between modalities. DAE approaches uses RBMs to learn a new representation of raw data with hidden units. Thus the transformed features share a normalized form and more connections between different modalities can be extracted.

In this work, we focus on the feature fusion strategy. However, the status of driving fatigue is a dynamic process which is changing over time. To incorporate the temporal dependency into driving fatigue detection, continuous conditional random field (CCNF) has been applied to detect driving fatigue [10]. In our future work, we could combine the fusion strategy on feature level and other optimization methods on regression models to enhance the accuracy of driving fatigue detection.

V. CONCLUSION

In this paper, we have discussed the key brain areas where we combine EEG signals with forehead EOG to detect driving fatigue. While posterior site achieves better performance for single EEG, our experiment results demonstrate that temporal EEG and EOG (T + EOG) fusion features perform better than other features. EOG is considered to be a promising unimodal physiological signals to detect driving fatigue. However, the multimodal fusion strategies can dramatically improve the performance. We also presents a novel multimodal Deep Autoencoder model which is widely applicable in across modality combination. The COR and RMSE of unimodal physiological signals are 0.78 and 0.12, respectively, whereas the modality fusion with DAE can significantly enhance the performance with values of 0.85 and 0.09, respectively. The DAE method is more robust and can learn a batter shared representation. Since the correlation between multimodal features are highly non-linear, only using linear stage is difficult to extract the relationships. Thus the normalized features learned from modalities can be well suited to detect driving fatigue.

REFERENCES

[1] S. K. D'mello and J. Kory, "A review and meta-analysis of multimodal affect detection systems," *ACM Comput. Surv.*, vol. 47, no. 3, pp. 43:1–43:36, 2015.

[2] T. Morris and J. Miller, "Electrooculographic and performance indices of fatigue during simulated flight," *Biological Psychology*, vol. 42, no. 3, pp. 343 – 360, 1996.

[3] L.-C. Shi and B.-L. Lu, "EEG-based vigilance estimation using extreme learning machines," *Neurocomputing*, vol. 102, pp. 135–143, 2013.

[4] C.-T. Lin, R.-C. Wu, S.-F. Liang, W.-H. Chao, Y.-J. Chen, and T.-P. Jung, "EEG-based drowsiness estimation for safety driving using independent component analysis," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 12, pp. 2726–2738, 2005.

[5] J. X. Ma, L. C. Shi, and B. L. Lu, "Vigilance estimation by using electrooculographic features," in *Annual International Conference of the IEEE Engineering in Medicine and Biology*, 2010, pp. 6591–6594.

[6] A. Bulling, J. A. Ward, H. Gellersen, and G. Troster, "Eye movement analysis for activity recognition using electrooculography," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 741–753, 2011.

[7] Y. F. Zhang, X. Y. Gao, J. Y. Zhu, W. L. Zheng, and B. L. Lu, "A novel approach to driving fatigue detection using forehead EOG," in *7th International IEEE/EMBS Conference on Neural Engineering*, 2015.

[8] M. Golz, D. Sommer, M. Chen, U. Trutschel, and D. Mandic, "Feature fusion for the detection of microsleep events," *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, vol. 49, no. 2, pp. 329–342, 2007.

[9] Y. Lu, W.-L. Zheng, B. Li, and B.-L. Lu, "Combining eye movements and EEG to enhance emotion recognition," in *Proceedings of the 24th International Conference on Artificial Intelligence*, 2015, pp. 1170–1176.

[10] W.-L. Zheng and B.-L. Lu, "A multimodal approach to estimating vigilance using EEG and forehead EOG," *Journal of Neural Engineering*, 2017.

[11] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Ng, "Multimodal deep learning," in *Proceedings of the 28th International Conference on Machine Learning*, 2011, pp. 689–696.

[12] L. Wei, Z. Wei-Long, and L. Bao-Liang, "Emotion recognition using multimodal deep learning," in *Neural Information Processing: 23rd International Conference*, 2016, pp. 521–529,.

[13] R. N. Duan, J. Y. Zhu, and B. L. Lu, "Differential entropy feature for EEG-based emotion classification," in *6th International IEEE/EMBS Conference on Neural Engineering*, 2013, pp. 81–84.

[14] G. E. Hinton, *A Practical Guide to Training Restricted Boltzmann Machines*. Springer Berlin Heidelberg, 2012, pp. 599–619.