EOG-based drowsiness detection:
Comparison between a fuzzy system and
two supervised learning classifiers

Antoine Picot [*] Sylvie Charbonnier [**] Alice Caplier [**]

[*] *Sleep, Metabolism and Health Center, The University of Chicago,
Chicago, IL, USA (e-mail: picota@uchicago.edu).*
[**] *Gipsa-Lab, Grenoble, France (e-mail:
name.surname@gipsa-lab.grenoble-inp.fr).*

Abstract: Drowsiness is a serious problem, which causes a large number of car crashes every year. This paper presents an original drowsiness detection method based on the fuzzy merging of several eye blinking features extracted from an electrooculogram (EOG). These features are computed each second using a sliding window. This method is compared to two supervised learning classifiers: a prototype nearest neighbours and a multilayer perceptron. The comparison has been carried out on a substantial database containing 60 hours of driving data from 20 different drivers. The method proposed reaches very good performances with 82% of true detections and 13% of false alarms on 20 different drivers without tuning any parameters. The best results obtained by the supervised learning classification methods are only 72% of true detection and 26% of false alarms, which is far worse than the fuzzy method. It is shown that the fuzzy method overtakes the other methods because it is able to take into account the fact that drowsiness symptoms occur simultaneously and in a repetitive way on the different features during the epoch to classify, which is of importance in the drowsiness decision-making process.

Keywords: Fuzzy expert systems, Supervised learning, Drowsiness detection, Monitoring systems, EOG

1. INTRODUCTION

Drowsiness is a serious concern for drivers. The National Highway Traffic Safety Administration (NHTSA) has indeed enlightened that driver drowsiness is responsible for about 100,000 car crashes every year. This is the reason why more and more researches are made to build automatic detectors of this dangerous state.

It has been shown (Caffier et al. (2003); Galley et al. (2004)) that several features can be extracted from eye-blinking analysis in order to evaluate drowsiness. Electrooculogram (EOG) is the measurement of the eye electrical activity using a set of electrodes placed on the skin. It is the most reliable technique to study blinking activity and evaluate drowsiness. Indeed, expert doctors use EOG to manually evaluate drowsiness by slices of $20s$ (Gillberg et al. (1996); Muzet et al. (2003)). However, for obvious reasons, the research community has also focused on using video cameras to monitor drowsiness through the study of blinking (Ji and Yang (2002); Bergasa et al. (2006)).

Several multivariable techniques have been proposed to monitor drowsiness in an automatic way. These techniques attempt to copy the expert doctor decisions mechanism by merging several blinking features on a fixed time-window. Thus, techniques such as Hidden Markov Model (Noguchi et al. (2007)), clustering k-mean (Ohsuga et al. (2007)), multiple regression analysis (Omi et al. (2008)), Fuzzy Expert System (Damousis et al. (2009)) and Support Vec-

tor Machines (Hu and Zheng (2009)) have been recently explored to monitor drowsiness.

In this paper, a method to monitor drowsiness using EOG blinking features is presented. It consists in merging a selection of blinking features using fuzzy logic on a sliding window. A comparison between this technique and two other supervised classification methods on a large dataset of 60 hours recorded on drowsy drivers is carried out. The purpose of this comparison is to study the relevancy of the fuzzy approach in the case of drowsiness monitoring.

The outline of this paper is the following. The dataset used for the comparison, the EOG features and the fuzzy method are described in section 2. In section 3, the results obtained on the same database by the fuzzy method and by a prototype nearest neighbours and a multilayer perceptron are presented. These results are compared and discussed in section 4, where a comparison with the results from the literature review is also achieved.

2. MATERIAL AND METHOD

2.1 Material

The database used in this study includes 60 hours of driving data from 20 different drivers. Each subject was recorded while driving on a simulator for 90 minutes, a first time perfectly rested and a second time suffering from sleep deprivation (the subject had slept for 4 hours only).

Each recording includes four EEG channels and one EOG channel. Only the EOG channel is used in this paper. The EOG is recorded at 250Hz. Data acquisition was performed by the CEPA (*Centre d'Études de Physiologie Appliquée*), Strasbourg, FR.

Objective sleepiness was evaluated every 20s (named epoch) on each recording by an expert doctor using the scale described in Muzet et al. (2003). The level of drowsiness is assessed depending on the number of long blinking events present during the epoch. This scale has been converted into a two-levels scale $[awake; drowsy]$. This represents 6461 epochs classified as "awake" and 1096 as "drowsy".This database is randomly split in two databases in order to have a similar number of "awake" and "drowsy" epochs in both databases. The first half (*dataset 1*) is used to design the fuzzy functions and to train the supervised methods. The other half (*dataset 2*) is used to test the different methods.

2.2 Feature extraction

Several features can be extracted from blinking detected on EOG. A recent study of Picot et al. (2009) has shown that some blinking features can be extracted with a similar accuracy from a high frame-rate video analysis than from EOG analysis. These features are the duration at 50% (*D50*)), measured by the time between the half rise amplitude to half fall amplitude, the percentage of eye closure at 80% (*P80*)), which is the percentage of time where the eyes were closed at least 80%, the blinking frequency (*F*) and the ratio between the amplitude of the blinking and the peak closing velocity, which is the maximum speed during the closing period calculated on the same blinking, (*A/PCV*). As the video solution is more practical for the driver, this paper focuses on these particular features.

All these features are depicted in Fig. 1. It shows an example of EOG signal measured when a blink is occurring and an illustration of the EOG derivative (velocity).
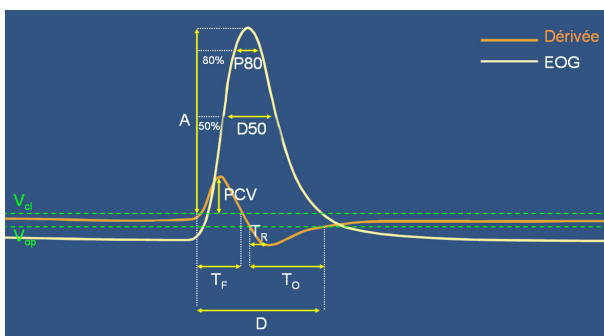

Figure 1. EOG signal during a blink, its corresponding derivative and the features considered.

These blinking features were semi-automatically detected on EOG using the detection algorithm described in Jammes et al. (2008). The mean value of each feature is computed on a sliding window at each second i. Several lengths of sliding windows were tested on *dataset 1*: 5s, 10s, 20s, 40s and 60s. The mean value of D50, P80, F and A/PCV at time i are respectively noted $D50(i)$, $P80(i)$, $F(i)$ and $A/PCV(i)$.

Blinking features generally increase with drowsiness. When a driver is drowsy, he tends to blink more frequently and the duration of each blinking increases. Thus, each feature can be seen as a drowsiness indicator where the driver is considered as "drowsy" if the value of the feature is larger than a given threshold. For each feature, the ability of various threshold values to discriminate between drowsy and not drowsy states is evaluated on *dataset 1*. The threshold obtaining the best results is selected and named s_F. The results are presented in section 3.1.

2.3 Fuzzy drowsiness detector

To increase the relevancy of the drowsiness detection, the different features are first converted into fuzzy drowsiness indicators then merged into a global fuzzy indicator. Each feature is transformed into a fuzzy variable whose value is between 0 and 1. The fuzzy variable is defined as the membership degree to the "drowsy" state: the closer to 1 the value of the variable, the higher the likelihood of being drowsy. The features are transformed into fuzzy variables thanks to a membership function described by equation (1).

Coefficients a and b are defined by equations (2) depending on the optimal threshold s_F presented in section 3.1. The values -0.25 and $+0.25$ were chosen to centre the membership function around the considered threshold s_F.

$$\mu(x) = \begin{cases} 0 & if & x \leq a \\ \dfrac{x-a}{b-a} & if & a \leq x \leq b \\ 1 & if & x \geq b \end{cases} \quad (1)$$

$$a = s_F - 0,25 \cdot s_F \quad and \quad b = s_F + 0,25 \cdot s_F \quad (2)$$

Then, the fuzzy variables are merged at each time i using equation (3).

$$\mu_f(i) = \frac{1}{4} \sum_F \mu(F(i)), \; F \in \{D50, P80, F, A/PCV\} \quad (3)$$

μ_f is a global drowsiness indicator. Its value varies between 0 and 1: the closer to 1 μ_f, the more drowsy the observed driver. Actually, equation (3) makes a compromise between the drowsy indicators provided by the different features. To make the decision crisp, the driver is finally considered as drowsy if μ_f is larger than 0,5. Else, he is awake. A crisp decision is then made every second.

The evaluation of drowsiness by the expert doctor is made on epochs of 20s. The final drowsiness decision of the system must therefore be made every 20s, to be compared to the evaluation made by the expert doctor. So, the driver is considered as "drowsy" on the epoch of 20s if the decision "drowsy" is made during at least 10s in the epoch. The decision is noted $D(ep(j))$ where $ep(j)$ corresponds to the jth epoch. It is computed according to equation (4).

$$D(ep(j)) = \begin{cases} 1 \; if \; card_{i \in ep(j)}\{\mu_f(i) > 0,5\} \geq 10 \\ 0 \; otherwise \end{cases} \quad (4)$$

2.4 Supervised learning algorithms

In order to compare the results obtained with the method proposed, several supervised learning methods algorithm

were tested on the database: a Prototype Nearest Neighbor (*PNN*) algorithm and a Multi-Layer Perceptron (*MLP*).

A PNN is processed by first building kernels (prototypes) off-line using a clustering algorithm. A class membership is then affected to each kernel. Any new instance is affected to the class of its nearest prototype (Hastie et al. (2001)). In this study, the number of kernels was equal to ten. The Mahalanobis distance is used to compute the distance to each kernel. A MLP is a feedforward artificial neural network model that maps a set of input data to a set of appropriate outputs. Here, the MLP has 15 neurons in the hidden layer with four input data (the four features) and only one output which is the drowsy decision. The activation function used is a sigmoid function. Both algorithms were trained and tuned on *dataset 1*, using two sub-sets : a learning set and a validation set, to select the right number of kernels and the architecture of the MLP. The results presented below are obtained on *dataset 2*. Both algorithms were built using the free data mining software *TANAGRA* (Rakotomalala (2005))

Since it is not possible for supervised method to extract the blinking features directly with a sliding window as in the method proposed, several strategies have been implemented to avoid this problem.

The first one is to compute the average of each feature on each epoch of 20s. Blinking are detected during the 20 seconds corresponding to the epoch analysed, D50, P80, F and A/PCV are calculated for each blink and the average value are calculated. The inputs of the two classifiers are then the average value of each four features calculated on a fixed window of 20 seconds corresponding to the epoch. This strategy is named *snapshot*.

The other strategy consists in computing the average of each feature on a 20s sliding window every second, as for the fuzzy method proposed. The features extracted are D50(i), P80(i) F(i) and A/PCV(i) presented in section 2.2. The features are then summed up during the 20 seconds epoch analysed in one single value by computing either the corresponding mean, median or the 90th percentile (90th pct). The 90th pct is the value below which 90 percent of the observations may be found. It provides an estimation of the maximal value, robust to the presence of outliers.

A strategy using a temporal granularity of 1s is also tested. In this case, a decision is taken depending on the values of each feature every second.

3. RESULTS

The results are compared using Receiver Operating Characteristic (ROC) curves, plotting true positive rate (TP_{rate}) in function of the false positive rate (FP_{rate}). The TP_{rate} is calculated as the percentage of true detections *ie* the percentage of "drowsy" epochs correctly classified as "drowsy". The FP_{rate} is computed as the percentage of false alarms *ie* the percentage of "awake" epochs wrongly classified as "drowsy" by the system.

3.1 Fuzzy detector

The results obtained on *dataset 1* using different sliding window lengths and different drowsiness thresholds, as

proposed in section 2.2, are shown in Fig. 2. In this figure, the threshold values are varying in a coherent range for each feature. For a selected feature, the results on the upper right are obtained with lower threshold values (more true detections but more false alarms) and the ones on the lower left are obtained with higher threshold values (less false alarms but less true detections).
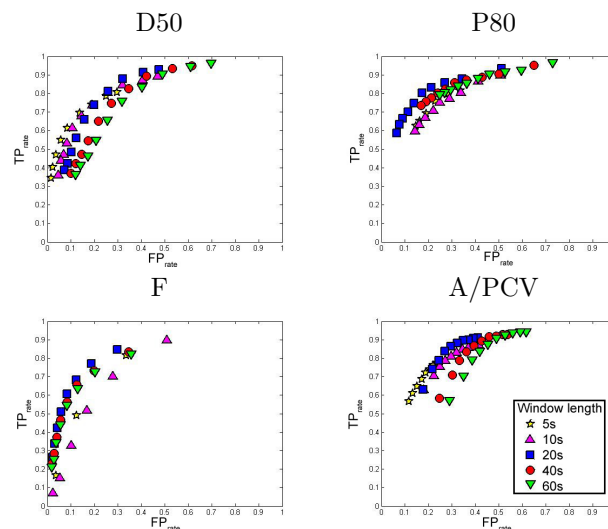

Figure 2. Results obtained with the different features

It can be seen on Fig. 2 that the number of true detections increases with the length of the sliding window. In the same time, the number of false alarms increases too. It can also be seen on this figure that too short a window length ($\leq 10s$) may increase the number of false alarms. A window length of 20s seems then the best compromise to have a large number (around 80%) of true detections while limiting the number of false alarms according to the visual inspection of Fig. 2. Table 1 shows the results optained with what we considered as the optimal threshold and a sliding window of 20s for each feature. We considered the results as optimal when they represent a good compromise between a large number of true detections (around 80%) and a low number of false detections (around 20%).

Table 1. The best results and the corresponding threshold obtained for each feature

Feature	Results	Threshold
D50	$TP_{rate} = 81,4\%$ $FP_{rate} = 24,9\%$	$s_F = 130ms$
P80	$TP_{rate} = 84,2\%$ $FP_{rate} = 22,3\%$	$s_F = 0,75\%$
F	$TP_{rate} = 78,3\%$ $FP_{rate} = 18,6\%$	$s_F = 0,1Hz$
A/PCV	$TP_{rate} = 78,5\%$ $FP_{rate} = 25,3\%$	$s_F = 17,25$

The optimal thresholds obtained on *dataset 1* are used to design the membership functions for each feature as described in section 2.3. The results obtained on *dataset 2* with the fuzzy method proposed are presented in table 2. The results obtained by each feature individually (monovariable detection) are also presented in the table.

The results obtained on *dataset 2* using one feature only are very similar to those obtained on *dataset 1*,

which proves the robustness of the selected threshold. The method is indeed independent of the driver since the same thresholds give good results on 20 different drivers. The impact of each driver is indeed really low in the learning dataset since the data from a high number of drivers are mixed together. So, even if the datasets used do not fit the criteria "no data from the same people in the learning and the testing dataset", the results are representative of future performance. Meanwhile, a cross-validation could be considered to confirm these results.

Moreover, the FP_{rate} significantly decreases when features are merged using fuzzy logic while the TP_{rate} is equivalent to the best one obtained with one feature only. This means that the false alarms obtained with the different features do not occur at the same time. Thus, when the different features are merged, the false alarms produced by one of the features are cancelled by the others which explains the decrease of the FP_{rate}. On the contrary, true detections occur concomitantly on most features and are not eliminated during the fusion process.

3.2 Supervised learning algorithms

In order to compare the results presented in section 3.1, the results obtained using a PNN algorithm and those obtained with a MLP algorithm are presented in this section.

Classification results using a temporal granularity of 1 second The features used are D50, P80,F and A/PCV, computed each second on a $20s$ sliding window, in the same way as in section 2.2. The data were labelled by the expert doctor by epochs of $20s$. To compare the results of the supervised algorithms with the expert classification, each second is labelled as the epoch to which it belongs. The results obtained on *dataset 2* with PNN and MPL methods are depicted in table 3.

It is obvious from table 3 that results are worse than those obtain with the fuzzy method (cf. table 2). It can be explained by the fact that if a $20s$ epoch is labelled as "drowsy", it does not mean that every single second from this epoch can be labelled as "drowsy". Indeed, expert classification is made depending on the blinking features the expert observes during the epoch analysed (frequency, blinking length). Blinking are very short (less than $1s$). If each second is labelled as the $20s$ epoch it belongs to,

Table 2. Results of the drowsiness detection

	Fuzzy logic	D50	P80	F	A/PCV
TP_{rate}	81,7%	80,7%	82,8%	77,0%	78,6%
FP_{rate}	13,4%	25,1%	21,4%	19,3%	23,9%

Table 3. Results of supervised method on features computed each second

	PNN	MLP
TP_{rate}	67,5%	38,1%
FP_{rate}	42,2%	5,2%

the scatters of points forming the two classes are likely to overlap. This is what happens for the results in table 3. Numerous true detections are made with PNN (67.5%) but with a very high rate of false alarms (42,2%). MLP produces only a few false alarms (5,2%) but it is not accurate to detect drowsiness (only 38,1% true detections). It is indeed difficult to separate the two classes in both cases because the classes are dispersed.

Classification results using a temporal granularity of 1 epoch The two supervised algorithms were trained on *dataset 1*. The results obtained on *dataset 2* with both supervised methods are shown in table 4. The algorithms are tested for each feature computing strategy presented in section 2.4. In order to study the impact of the way the features are computed, the results obtained by the fuzzy method proposed with the different features computing strategies are also displayed in table 4. For each feature, each computed value (snapshot, mean, median, or 90th pct) is fuzzified using equation (1) and the four corresponding fuzzy indicators are merged using equation (3). The decision between "drowsy" and "awake" is made by thresholding the resulting μ_f. The same optimal thresholds s_F presented in table 1 are used.

Table 4. Results obtained with the supervised learning classifiers and the fuzzy classifier

	Fuzzy logic	PNN	MLP
snapshot	$TP_{rate} = 68\%$	$TP_{rate} = 61\%$	$TP_{rate} = 52\%$
	$FP_{rate} = 27\%$	$FP_{rate} = 23\%$	$FP_{rate} = 11\%$
average	$TP_{rate} = 72\%$	$TP_{rate} = 65\%$	$TP_{rate} = 65\%$
	$FP_{rate} = 26\%$	$FP_{rate} = 22\%$	$FP_{rate} = 16\%$
median	$TP_{rate} = 72\%$	$TP_{rate} = 64\%$	$TP_{rate} = 52\%$
	$FP_{rate} = 27\%$	$FP_{rate} = 22\%$	$FP_{rate} = 11\%$
90th pct	$TP_{rate} = 80\%$	$TP_{rate} = 64\%$	$TP_{rate} = 53\%$
	$FP_{rate} = 39\%$	$FP_{rate} = 21\%$	$FP_{rate} = 12\%$

4. DISCUSSION

4.1 Comparison between the supervised algorithms and the fuzzy detector

Firstly, a comparison between table 3 and table 4 shows that better results are obtained when only one value per epoch is used for training PNN and MLP. It can be explained by the fact that when only one value by epoch is used for each feature, there is no classes dispersion during the learning. The classification obtained is then more accurate.

Secondly, it can be observed from table 4 that the fuzzy approach, combined to the same computing features strategies, give similar results than PNN or MLP. Results obtained with MLP are slightly less good than the other two techniques. It also appears that the fuzzy method proposed in section 2.3 obtains better results ($TP_{rate} = 81,7\%$ and $FP_{rate} = 13,4\%$, see table 2) than those obtained with the supervised techniques. This tends to prove that the way features are computed plays an important part in the quality of the results.

Finally, the way the different features values are summed up during the epoch analysed seems not to have any influence since using the mean, the median or the 90th

pct provide similar results. So, the better results obtained by the proposed method seem to be due the fact that it is possible to process the features second by second without dispersing the classes with the fuzzy method.

The results presented in section 3.2 have enlightened some problems linked to the way information is processed on the $20s$ epoch. An epoch labelled as "drowsy" means that some symptoms of drowsiness appear during the "20s". It does not mean that symptoms of drowsiness are present in every single second of the epoch. When using supervised techniques such as PNN or MLP, two solutions are possible: to process the information second by second which leads to a dispersion in the classes or to process the information globally which leads to a loss in the localisation of the information.

When the information is processed every second in order to keep a precise temporal granularity, the classes are dispersed during the learning phase. An accurate drowsiness detection is then not possible as shown in table 3. It is therefore necessary to make a preliminary processing on each feature (mean, median...) in order to have only one value per feature on each epoch before starting the learning process. In this way, two major pieces of information are lost. The first one is whether symptoms of drowsiness occur several times during the epoch. The second one is whether drowsiness symptoms occur simultaneously on the different features.

The fuzzy method proposed makes it possible to process information with a fine temporal granularity. The fact that drowsiness symptoms occur concomitantly on the different features can thus be taken into account in the decision. This is illustrated by the drop in the number of false alarms when merging the different features with the method proposed while the number of good detection remains high (cf. section 3.1). As decisions are made every second, it is possible to strengthen detections where the same decisions are made at the same time on the different features. It is also possible to decrease the number of false alarms as they do not occur at the same time on the different features. When a false alarm is produced by one feature, it is suppressed by the other features.

The fact that relevant drowsiness symptoms appear several times during the epoch is also taken into account in the decision. Indeed, drowsiness is detected if at least 10 seconds out of 20 are classified as drowsy. This is similar to the way the expert reaches his decision. He looks for drowsiness symptoms on the EOG signal and make the decision "drowsy" if their number is significant. This is not possible to do with supervised techniques where information has to be summed up into one single value for the whole epoch before making any decision.

Finally, results of section 4.1 have shown that the fuzzy classification method, which may seem "naive" because possible correlations between the features are not taken into account, gives as good results as some more elaborate supervised techniques, in this context of drowsiness detection. Moreover, the technique proposed, based on a pretreatment of the features using a sliding window followed by a fuzzyfication of the features and a fuzzy merging phase, gives much better results than those obtained with the two supervised methods carried out, mainly because of

its ability to take into account the repetition of symptoms during one epoch and the simultaneity of their occurrence on the different features. Another advantage of the method is that it is easy to use. It was trained once to find the appropriate thresholds to design membership functions. These thresholds can now be used on any new driver. This is shown by the fact that very good results were obtained on the validation set composed of 20 different drivers

4.2 Literature comparison

Some studies have focused on the possibility of merging different features to increase the relevancy of drowsiness detection. Omi et al. (2008) got good performances ($TP_{rate} = 84\%$ and $FP_{rate} = 9\%$) by using a multiple regression model on a dozen blinking features. However, their system was tested on only five drivers, whereas our method was tested on twenty different drivers. Similar results have been obtained by Hu and Zheng (2009) who merged the features with a Support Vector Machine (SVM) algorithm. They obtained 81% correct detections and 17% false alarms. Recently, a fuzzy expert system was proposed by Damousis et al. (2009) to monitor drowsiness through the duration, the frequency and the intervals of driver blinking. They obtained 90% of correct detections which is slightly better than our results. Nevertheless, they also had 30% of false alarms which is worse than the FP_{rate} obtained by the proposed method. This means that the features used in this article for drowsiness detection are more relevant. All these results are summarized in table 5.

Table 5. Synthesis of the different results for drowsiness detection from visual signs.

Authors	$\mathbf{TP}_{rate}$ $\mathbf{FP}_{rate}$	Technique
Picot et al., 2010	81,4% 13,1%	Fuzzy fusion of D50, P80, F and A/PCV features
Omi et al., 2008	84% 9%	Multiple regression on about ten features
Hu & Zheng, 2009	81% 17%	SVM on about ten features
Damousis et al., 2009	90% 30%	Fuzzy fusion duration, frequency and blink intervals

The results presented in this paper are very good compared to those found in the literature. However, it is necessary to remain somewhat reserved, since all these results were not obtained on the same database. Yet, the database used in this study is large enough (60 hours of driving time from 20 different drivers) to make a correct assessment of the performance of the system. In addition to the good performances obtained, the method proposed does not need to be tuned for each driver.

The advantage of the method proposed is that it is independent from the driver. The use of fuzzy function on the different features makes the system robust to inter-individual differences as the good performance has been obtained on 20 different drivers without tuning any parameters. Merging the different features allows a decrease in the number of false alarms which occur when only one feature is used. Even the earliest stages of drowsiness classified by the medical expert were detected by the system.

5. CONCLUSION

An original method of drowsiness detection using eye blink features extracted from EOG has been presented in this paper. This method is based on the fuzzy merging of several blinking features computed every second on a $20s$ sliding window. This method has been tested on a substantial database of 60 hours of driving EOG data from 20 different drivers and reaches good performances with 81,7% correct detections and 13,1% false alarms.

The comparison of this method to supervised learning methods such as a prototype nearest neighbour classifier and a multilayer perceptron algorithm on the same database allowed the advantages of the method to be enlightened. This comparison has pointed out the importance to make a decision with a fine temporal granularity. Indeed, the fact that drowsiness symptoms appear simultaneously and in a repetitive way on the different features is of importance in the drowsiness decision. The use of fuzzy logic makes it possible to process the features second by second and to obtain good classification performances, which is not possible to do with the two classification methods tested. The best results obtained by the supervised learning classification methods are 72% of true detection with 26% of false alarms, which is worse than the results obtained with the fuzzy method. Moreover, the fuzzy method proposed reaches good results compared to literature. One major advantage of the method is that the method is independent from the driver. It reached very good performances on 20 different drivers without tuning any parameters. So, the use of fuzzy logic to merge different features in order to monitor drowsiness seems to be a good solution as it is independent from the driver and it allows a good temporal granularity of the features which make the drowsiness detection more reliable.

Finally, one advantage of the method is that it uses eye blinking features that can be extracted with the same accuracy from a high frame video that from EOG. This was shown in (Picot et al. (2009)). The next step of this work would be the design and validation of a video-based drowiness detector.

ACKNOWLEDGEMENTS

REFERENCES

Bergasa, L., Nuevo, J., Sotelo, M., Barea, R., and Lòpez, M.E. (2006). Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, 7, 63–77.

Caffier, P., Erdmann, U., and Ullsperger, P. (2003). Experimental evaluation of eyeblink parameters as a drowsiness measure. *European Journal of Applied Physiology*, 89, 319–325.

Damousis, I., Tzovaras, D., and Strintzis, M. (2009). A fuzzy expert system for the early warning of accidents due to driver hypo-vigilance. *Personal and Ubiquitous Computing*, 13(1), 43–49.

Galley, N., Schleicher, R., and Galley, L. (2004). Blink parameter as indicators of driver's sleepiness - possibilities and limitations. *Vision in Vehicles*, 10, 189–196.

Gillberg, M., Kecklund, G., and Akerstedt, T. (1996). Sleepiness and performance of professional drivers in a truck simulator - comparisons between day and night driving. *Journal of Sleep Research*, 5, 12–15.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The elements of statistical learning - Data mining, Inference & Prediction.* Springer Verlag.

Hu, S. and Zheng, G. (2009). Driver drowsiness detection with eyelid related parameters by support vector machine. *Expert Systems with Applications*, 36(4), 7651–7658.

Jammes, B., Sharabty, H., and Esteve, D. (2008). Automatic eog analysis: a first step toward automatic drowsiness scoring during wake sleep transitions. *Somnologie*, 12(3), 227–232.

Ji, Q. and Yang, X. (2002). Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging*, 8, 357–377.

Muzet, A., Pébayle, T., Langrognet, J., and Otmani, S. (2003). AWAKE pilot study no.2: Testing steering grip sensor measures. Technical Report IST-2000-28062, CEPA.

Noguchi, Y., Nopsuwanchai, R., Ohsuga, M., and Kamakura, Y. (2007). Classification of blink waveforms towards the assessment of drivers arousal level - an approach for HMM based classification from blinking video sequence. In *Proc. of the Symposium on Biological and Physiological Engineering.* Harbin, China.

Ohsuga, M., Kamakura, Y., Inoue, Y., Noguchi, Y., and Nopsuwanchai, R. (2007). Classification of blink waveforms toward the assessment of driver's arousal levels - an EOG approach and the correlation with physiological measures. In *Proc. of the Symposium on Biological and Physiological Engineering.* Harbin, China.

Omi, T., Nagai, F., and Komura, T. (2008). Driver drowsiness detection focused on eyelid behaviour. In *Proc. of the 34th Congress on Science and Technology of Thailand.* Bangkok, Thailand.

Picot, A., Caplier, A., and Charbonier, S. (2009). Comparison between EOG and high frame rate camera for drowsiness detection. In *Proc. of the IEEE Workshop on Applications of Computer Vision.* Snowbird, USA.

Rakotomalala, R. (2005). Tanagra : un logiciel gratuit pour l'enseignement et la recherche. *Actes de EGC'05*, 2, 697–702.