

Multi-sensor driver drowsiness monitoring

P Boyraz^{1*}, M Acar², and D Kerr²

¹Erik Jonsson School of Engineering and Computer Science, Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX, USA

²Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, UK

The manuscript was received on 6 January 2007 and was accepted after revision for publication on 23 July 2008.

DOI: 10.1243/09544070JAUTO513

Abstract: A system for driver drowsiness monitoring is proposed, using multi-sensor data acquisition and investigating two decision-making algorithms, namely a fuzzy inference system (FIS) and an artificial neural network (ANN), to predict the drowsiness level of the driver. Drowsiness indicator signals are selected allowing non-intrusive measurements. The experimental set-up of a driver-drowsiness-monitoring system is designed on the basis of the sought-after indicator signals. These selected signals are the eye closure via pupil area measurement, gaze vector and head motion acquired by a monocular computer vision system, steering wheel angle, vehicle speed, and force applied to the steering wheel by the driver. It is believed that, by fusing these signals, driver drowsiness can be detected and drowsiness level can be predicted. For validation of this hypothesis, 30 subjects, in normal and sleep-deprived conditions, are involved in a standard highway simulation for 1.5 h, giving a data set of 30 pairs. For designing a feature space to be used in decision making, several metrics are derived using histograms and entropies of the signals. An FIS and an ANN are used for decision making on the drowsiness level. To construct the rule base of the FIS, two different methods are employed and compared in terms of performance: first, linguistic rules from experimental studies in literature and, second, mathematically extracted rules by fuzzy subtractive clustering. The drowsiness levels belonging to each session are determined by the participants before and after the experiment, and videos of their faces are assessed to obtain the ground truth output for training the systems. The FIS is able to predict correctly 98 per cent of determined drowsiness states (training set) and 89 per cent of previously unknown test set states, while the ANN has a correct classification rate of 90 per cent for the test data. No significant difference is observed between the FIS and the ANN; however, the FIS might be considered better since the rule base can be improved on the basis of new observations.

Keywords: driver monitoring, driver vigilance, signal modelling, fuzzy inference systems

1 INTRODUCTION

Monitoring the drowsiness of drivers is currently the subject of active and preventative safety research. The aim for monitoring is to provide a solution to the drowsiness problem on roads from an engineering perspective. Several studies have been conducted in order to explain the phenomenon or to produce a reliable measure for on-road driver monitoring technologies [1–6]. One pragmatic study produced a

successful metric known as the percentage closure of the eyelid (PERCLOS), obtainable by computer vision (CV) techniques [7]. In order to employ PERCLOS for monitoring in a non-intrusive manner, real-time CV systems based on bright-pupil techniques have been widely investigated [8]. However, these systems are not able to infer the drowsiness level of the driver all the time; they are prone to error due to sensor failure, changing illumination conditions, and limited field of view as addressed by Bergasa *et al.* [9].

In addition to PERCLOS, methods based on different metrics such as steering-wheel movements have been widely investigated. Pilutti and Ulsoy [10]

*Corresponding author: Erik Jonsson School of Engineering and Computer Science, Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX, 75080, USA. email: pinar.boyraz@utdallas.edu; boyraz.pinar@gmail.com

have studied lane-keeping performance of drivers by taking the steering-wheel movements as the input and the lane deviation as the output. Swingler and Smith [11] and Sayed and Eskandarian [12] have developed methods to monitor driver alertness by measuring the steering actions and designing a neural network to classify them as representing a drowsy or alert driver state.

Although these metrics and signals themselves indicate drowsiness with high performance, monitoring systems relying on any single measure cannot be considered as sufficiently robust. For this reason, various researchers have conducted studies to combine and correlate the signals to overcome the deficiency of single-input systems. Bittner *et al.* [2] correlated the symptoms from three different signals, namely steering-wheel adjustments, vertical electro-oculography for eye movements, and several parameters by independent assessors obtained by watching the video of the driving simulations. In a similar study, Brookhuis and De Waard [3] correlated physiological signals such as electroencephalography with vehicle-related performance measures, namely car-following performance and number of steering-wheel reversals. Different combinations of these signals to obtain a more robust system have been established by Gress-Hernandez *et al.* [13], Apostoloff and Zelinsky [14] and Victor [15].

In this study, first, a reconsideration of previously suggested metrics is offered as well as suggesting new metrics. Second, a time window analysis is performed on these metrics to assess how narrow the processing time windows could be with a trade-off between the update rate of the prediction and the accuracy of the output. Finally, two different pattern recognition methods are compared. In summary, there are three different sources affecting the monitoring system performance:

- (a) metric definition and how well they can capture or extract the information in the signal;
- (b) the time window over which the metrics are calculated;
- (c) the decision-making system using the metrics as feature vectors.

All these three aspects are considered here as a framework and first step for obtaining robust drowsiness-monitoring systems.

2 EXPERIMENTAL TECHNIQUES

30 subjects participating in the experiments drive the same simulated highway scenario for 1.5 h under

'normal' and 'sleep-deprived' conditions. For the latter, subjects are asked to sleep at least 2 h less to induce sleepiness and the 'sleep-deprived' data collection is arranged to take place when a subject's circadian rhythm is in a low state (in the afternoon between 2 and 4 pm). The data collection is performed in alternate orders for sleep-deprived and normal conditions in order to avoid undesirable trends in data (i.e. the sleep-deprived session did not always follow the normal session, for half of the subjects the order was sleep deprived first and normal second, and for the other half vice versa). The highway scenario includes no events and is deliberately designed to be monotonous, but a wind gust effect is added to force the subjects to counter the drift caused by this disturbance. For each experiment, a sample from data is cropped to form three segments, each of 12 min duration: start, midterm, and final. These segments are labelled to indicate the corresponding time interval, for observing the temporal effect on task performance. Hence, for each subject, there are six different sessions in total; the first three sessions correspond to the start, midterm, and final sections for the 'normal' driver and the last three correspond to start, midterm, and final sections for the 'sleep-deprived' driver. Although 'normal' condition data batches are observed to be significantly different from 'sleep-deprived' data, gradual deterioration is also expected within a 'normal' condition data batch when its start, midterm, and final sections are compared.

The vehicle simulator is equipped with a CV system for eye tracking, strain gauges for measuring applied force on the steering wheel, and two encoders: one for measuring the longitudinal speed of the car via the throttle angle and the other for measuring the steering-wheel angle (SWA) (indicating lateral control for lane keeping). All the signals obtained from this multi-sensor system are intuitively grouped under either visual or non-visual channels, eye tracking comprising the visual channel and the rest of the signals constituting the non-visual channel. The multi-sensor system and derived metrics are explained in greater detail in section 3. However, the experimental geometry and overall view of the system are given here to develop insight into reasoning behind sensor selection and how signals from these sensors are interpreted. Fig. 1 depicts the experimental geometry and the highway image projected on to a screen at the front of the vehicle simulator.

The CV system is placed just in front of the driver to capture the eye gaze and closure. The measured

eye gaze vector is directly related to where the driver's attention is focused on the road scene or the speedometer. The x coordinate of the gaze vector is related to left-right movement of the eyes, whereas the y coordinate is related to vertical eye movement (checking different regions of the road). If the driving task is thought of as a combination of lateral and longitudinal control of the car position (called here lane keeping and speed control), the measured gaze vector can be related to these task depending on where the attention of the driver is. The x coordinate of the gaze vector is expected to indicate whether a driver's attention is fixed on the left or the right, whereas the y coordinate of the gaze vector is expected to indicate whether the driver is checking the speedometer (down) or road scene (up). The aim is to separate these two different tasks as much as possible in order to track changes in gaze vector distributions during the experiment. Driver models in the literature support the assumption here that in a simplistic way the driving task can be separated into longitudinal and lateral control of the vehicle [16]. Following the modelling approaches, the driver subjects are instructed to keep a constant speed for the longitudinal task and to keep the vehicle in the correct lane position for the lateral task as much as possible, allowing a separate examination of the two important driving tasks and their related metrics. Table 1 shows the separation between tasks and related signals for each of them.

3 SUBSYSTEMS OF THE VIGILANCE MONITORING UNIT AND INDICATIVE SIGNALS

This study proposes a new approach that produces a single reliable measure of drowsiness level by deriving and combining several metrics with an emphasis on trade-off between prediction accuracy and update rate of system output. The monitoring system consists of two integrated subsystems: first, the visual channel, i.e. a CV system to measure eye closure, gaze vector, and head motion, and, second, a non-visual channel, i.e. a driver input measurement unit consisting of strain gauges to measure indirectly the applied force on the steering-wheel column (SWC) and encoders to give the steering angle and throttle angle indicating the longitudinal speed. As these two channels are examined, the features and metrics derived from the signals and their indicative properties of metrics for drowsiness are discussed.

Table 1 Separation between tasks and related signals

Task	Related signal
Longitudinal task = speed regulation	Gaze vector y component, speed
Lateral task = lane keeping	Gaze vector x , y components, SWA, force on steering wheel

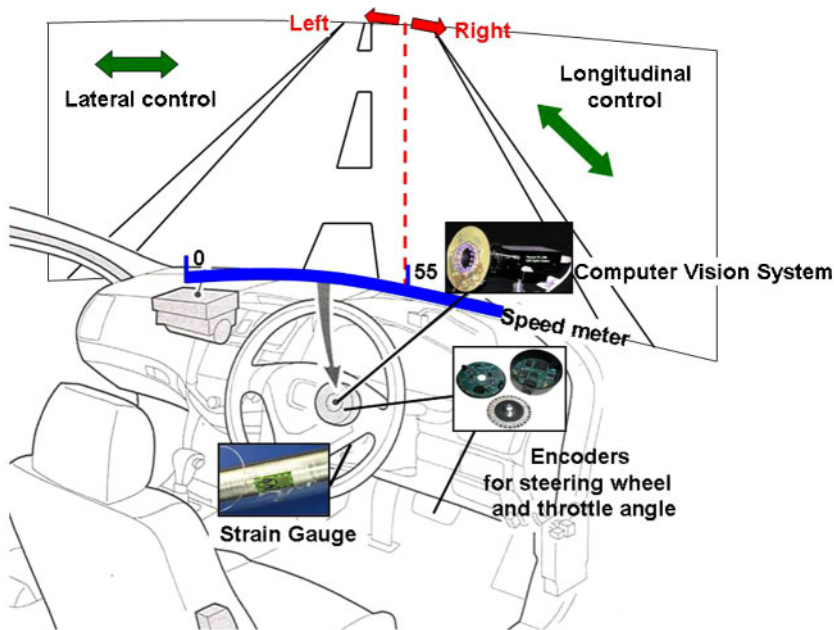


Fig. 1 Experiment geometry and system overview

3.1 Computer vision system (visual information channel)

The CV system is based on a bright-pupil technique to facilitate eye detection and tracking. In this technique a near-infrared (NIR) light source is placed coaxially with the optical axis of the camera. The reflected light from the eye retina of the human subject is caught by a camera as a result of this alignment. In order to obtain this bright-pupil effect, a monocular complementary metal–oxide–semiconductor (CMOS) camera is used, with illumination from a ring of light-emitting diodes aligned coaxially with the camera lens and emitting NIR light at a wavelength of 880 nm. A video stream of frontal face images including eye regions with the bright-pupil effect is continuously obtained and processed frame by frame. The details of the hardware and algorithms developed for the CV system is given in Appendix 1.

The CV system can output five different measurements: pupil area, and the vector components gaze x , gaze y , head x , and head y as defined below. The resultant segmented region of the driver's eye, with the associated NIR corneal reflection or 'glint' can be seen in Fig. 2(a). The pupil area is measured to obtain eye closure and number of blinks as functions of time, which is a direct indicator of drowsiness. In addition to eye closure, the attention of the driver as a function of time can be obtained by measuring the eye movement based on a vector between the centres of glint and pupil in the segmented image. The x and y components of this vector are called here gaze x and gaze y . In addition to this, plotting the coordinates of the centre of the pupil or glint versus time gives the head motion in two dimensions, called here head x and head y , representing the horizontal and vertical movements respectively. A sample of these signals and their interpretation are given in Figs 2(a) and (b). All the signals are normalized here, but in different intervals. A zero value of gaze x represents the line between left and right sections of the screen aligned with the optical axis of the camera, and hence a 'straight-ahead' gaze physically. For the gaze y vector, a zero value represents the vertical level of the speedometer and negative values correspond to different vertical levels on the screen. The negative sign occurs owing to the use of a vector description, although the important aspect is the magnitude of the measured gaze y value, rather than its sign.

The raw signals from the CV system can be summarized as a vector according to

CV raw

$$= [\text{gaze } x \quad \text{gaze } y \quad \text{head } x \quad \text{head } y \quad \text{pupil_area}] \quad (1)$$

The CV system signals are treated as ground truth, i.e. the most representative signals for the actual drowsiness level. A drowsy and an alert subject can be clearly distinguished by the measurements from the CV system, as shown in Fig. 3, which depicts the histograms of the measurements arranged in three successive time intervals for an 'alert' and a 'drowsy' driver.

The three rows show gaze y , gaze x , and pupil area histograms for start, midterm, and final parts of the driving sessions, each of 12 min duration. Quantitatively, a significant change is observed in the overall gaze y distribution as well as decay of gaze y in the interval $[-0.5, -1]$, indicating lack of attention to the lane-keeping task as time passes.

To utilize the full potential of the CV system's output, three metrics are defined, based on the observations from histograms of measurements for six sessions for 30 subjects, based on a 180-session database of observations of the same phenomena; they are the attention division ratio (ADR) defined by

$$\begin{aligned} \text{ADR} &= \frac{\text{longitudinal task}}{\text{lateral task}} \\ &= \frac{\text{speed checking}}{\text{road checking}} \\ &= \frac{N_{\text{gaze } y > -0.5}}{N_{\text{gaze } y < -0.5}} \end{aligned} \quad (2)$$

the first metric of eye closure (ECM1) defined by

$$\begin{aligned} \text{ECM1} &= \frac{\text{pupil fully closed}}{\text{pupil wide open}} \\ &= \frac{\sum_1^n N_{\text{pupil closed}}(i)}{\sum_1^n N_{\text{pupil wide open}}(i)} \end{aligned} \quad (3)$$

and the second metric of eye closure (ECM2) defined by

$$\text{ECM2} = \frac{> \text{pupil area} \times 0.9}{< \text{pupil area} \times 0.9} \quad (4)$$

The metrics are observed to be powerful in predicting the driver drowsiness level and they are believed to be more enriched than an eye closure metric alone (see Appendix 2). The ADR, for instance, gives extra information about driver attention level by

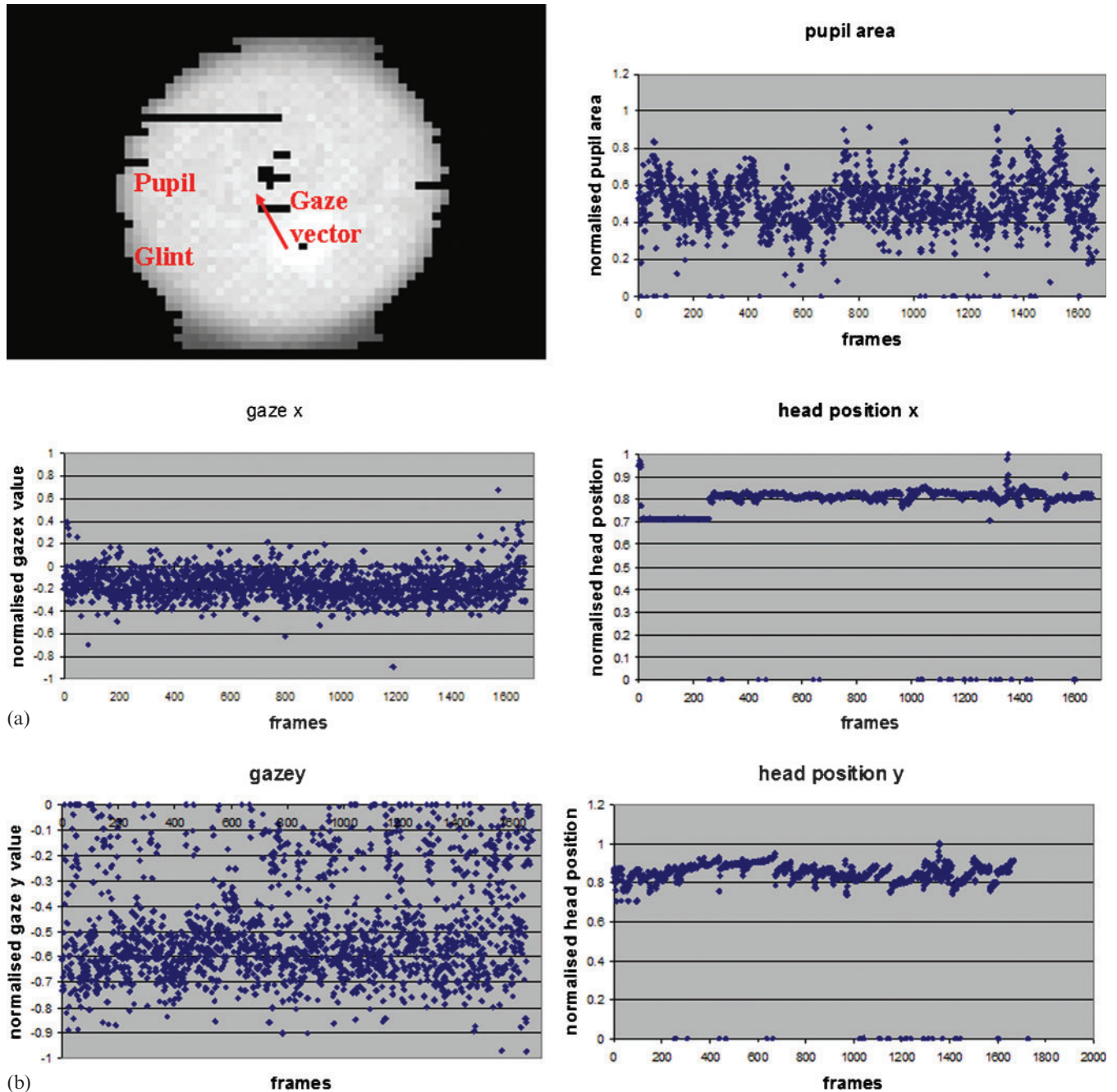


Fig. 2 (a) Sample outputs from the CV system: top left, low-level image features segmented; pupil, glint, and gaze vector defined by the pixel distance between pupil and glint centres; top right, pupil area; bottom left, a typical gaze x graph of a typical alert driver through time; bottom right, head position through time. (b) Sample outputs from the CV system: left, a typical gaze y graph of a drowsy subject, $0 > \text{gaze } y > -0.5$ indicates speed checking and $-0.5 > \text{gaze } y > -1$ indicates eye gaze for lane-keeping scene leading to a new metric defined as the ADR; right, head motion in the y coordinate, i.e. head y

including the ratio of the speed-checking eye gaze movements to the road-side-checking eye gaze movements. The other two metrics ECM1 and ECM2 exploit cumulative functions of eye closure with respect to time. ECM1 emphasizes the number

of blinks by taking the proportion of fully closed to fully open eye cases, while ignoring the cases in between. ECM2 represents partially closed cases as well as fully closed eye cases in the denominator of the ratio and inversely related to drowsiness level. A

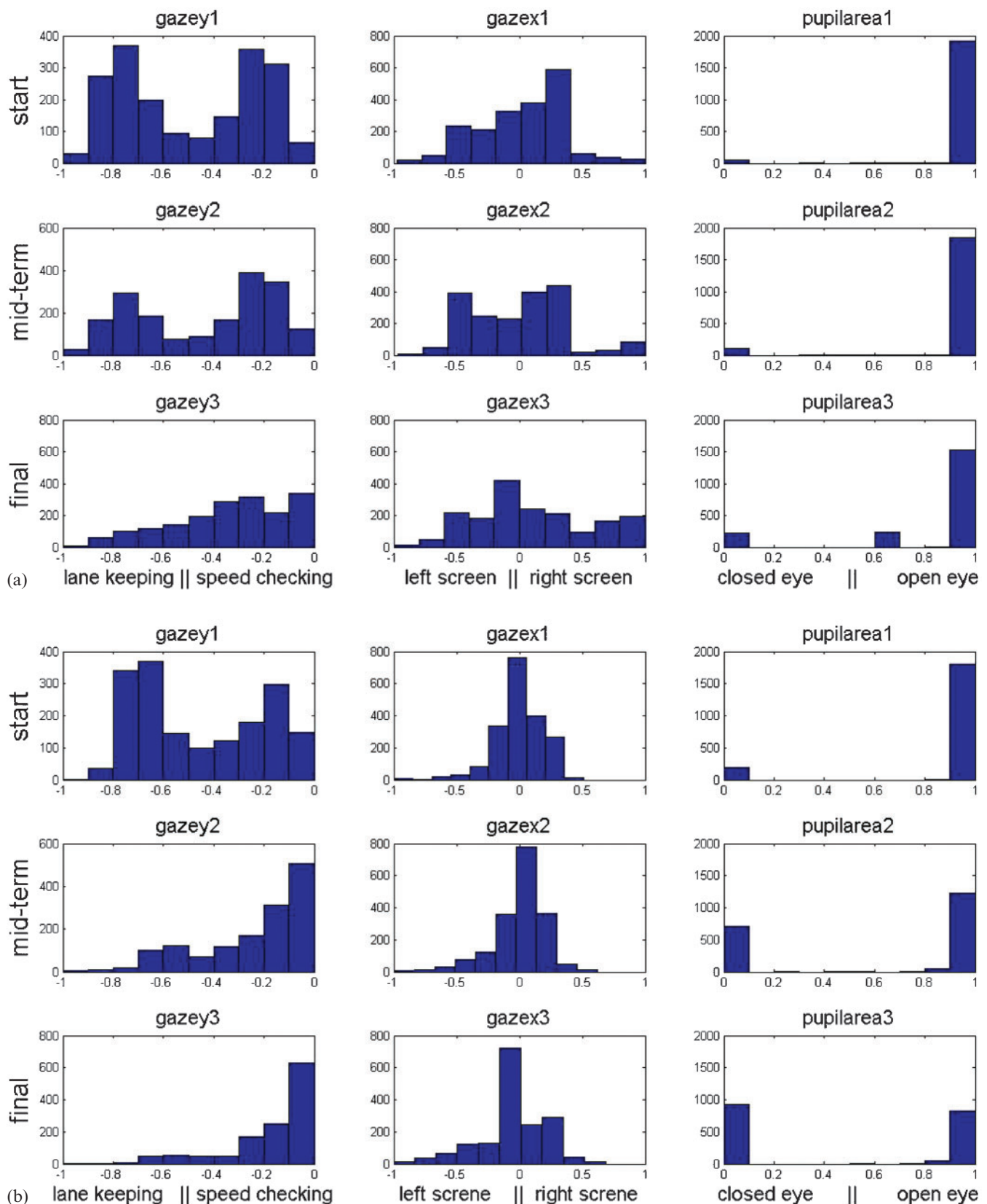


Fig. 3 Typical histograms of (a) alert and (b) drowsy

full analysis for comparison of traditional PERCLOS to proposed metrics is given in Appendix 2. Besides these newly defined metrics, the standard deviation of gaze x (to indicate scatter), the mean of gaze x (to show focusing on some particular point during the session), and the scatter and entropies of the head movement information in two dimensions are also included in the analysis.

3.2 Driver inputs (non-visual channels)

The force exerted on the steering wheel by the driver, and hence the bending of the SWC, is measured by strain gauges. The gauges are placed most conveniently on the SWC of the vehicle simulator. The signal is filtered to remove high-frequency noise prior to further processing. The reason for collecting this information is to identify its potential for detecting microsleep lapses and to fuse these data with the SWA to identify the mode of driving: relaxed, alert, etc. In the scope of this study, this signal will only be considered for its potential in helping to monitor the driver alertness level by using its complexity via entropy and its scatter via the standard deviation (Std). The meaningful information is expressed in the vector given by

$$\text{Force feature} = [\text{Std} \quad \text{entropy}] \quad (5)$$

The inclusion of speed measurements in the analysis helps to connect the attention distribution of the driver for two specifically defined tasks in the standard highway scenario. These tasks are longitudinal control of speed via the throttle and lateral position control (lane keeping) via the steering SWA. A sample output of non-visual information channels can be seen in Fig. 4.

For the speed measurements, the deviation from the reference speed that drivers are told to maintain is measured by the integral of the steady state error, which is known as the performance index in control theory used for proportional–integral–derivative controllers.

As can be seen from the simulator output in Fig. 4, the speed-keeping behaviour of a driver resembles a second-order system step response and steady state error showing that the longitudinal control performance of the driver, and hence the integral of standard error (ISE) and the integral of average error (IAE), are of interest. These integrals are given by

$$\text{ISE} = \int_{t_s}^T e^2(t) dt \quad (6a)$$

$$\text{IAE} = \int_{t_s}^T |e(t)| dt \quad (6b)$$

where t_s is the settlement time and T is the time window size.

The final metric definition combines these control variables with the Std and entropies of the lane deviation and SWA completing the non-visual channel sensor output according to

$$\begin{aligned} &\text{Driver input metrics} \\ &= [\text{ISE} \quad \text{IAE} \quad \text{SWA Std} \quad \text{SWA entropy}] \end{aligned} \quad (7)$$

4 METHODOLOGIES

4.1 Data Reduction

First, all the signals are normalized and baselines are found; then the basic statistical features (i.e. mean and Std) are calculated and any particular difference between ‘normal’ and ‘sleep-deprived’ data batches is sought. Second, entropies are calculated where applicable. The entropy measure is defined as the ‘information content’ in the signal in telecommunication engineering and is a good indicator of system response changes. The same concept is used here to represent the complexity of a signal.

Finally, three different forms of feature vectors $\{F_1, F_2, F_3\}$ are designed, as defined in Table 2. These are intended to embody the same phenomena either by visual cues only (F_1), control performance index and entropies (F_2), and visual cues together with the Stds (F_3). This grouping is based on intuitive perspective on metrics; no *post-hoc* analysis is performed at this stage. However, after comparison of the classification performance of each feature space, members not positively contributing to the classification will be eliminated.

The best of the feature vector space is formed by combining the outperforming members from different feature spaces. A typical change in feature vector members for F_3 can be seen in Fig. 5 for alert and drowsy states of the same driver.

The reduction of the raw data into a meaningful summary of the three different feature spaces can be seen in the flow chart in Fig. 6. This process of database preparation can be fully automated and the system can thus be made adaptive, in that it updates its own online database continuously.

After obtaining the feature vector spaces, available metrics are used to train and test three different systems for decision making on a drowsiness level.

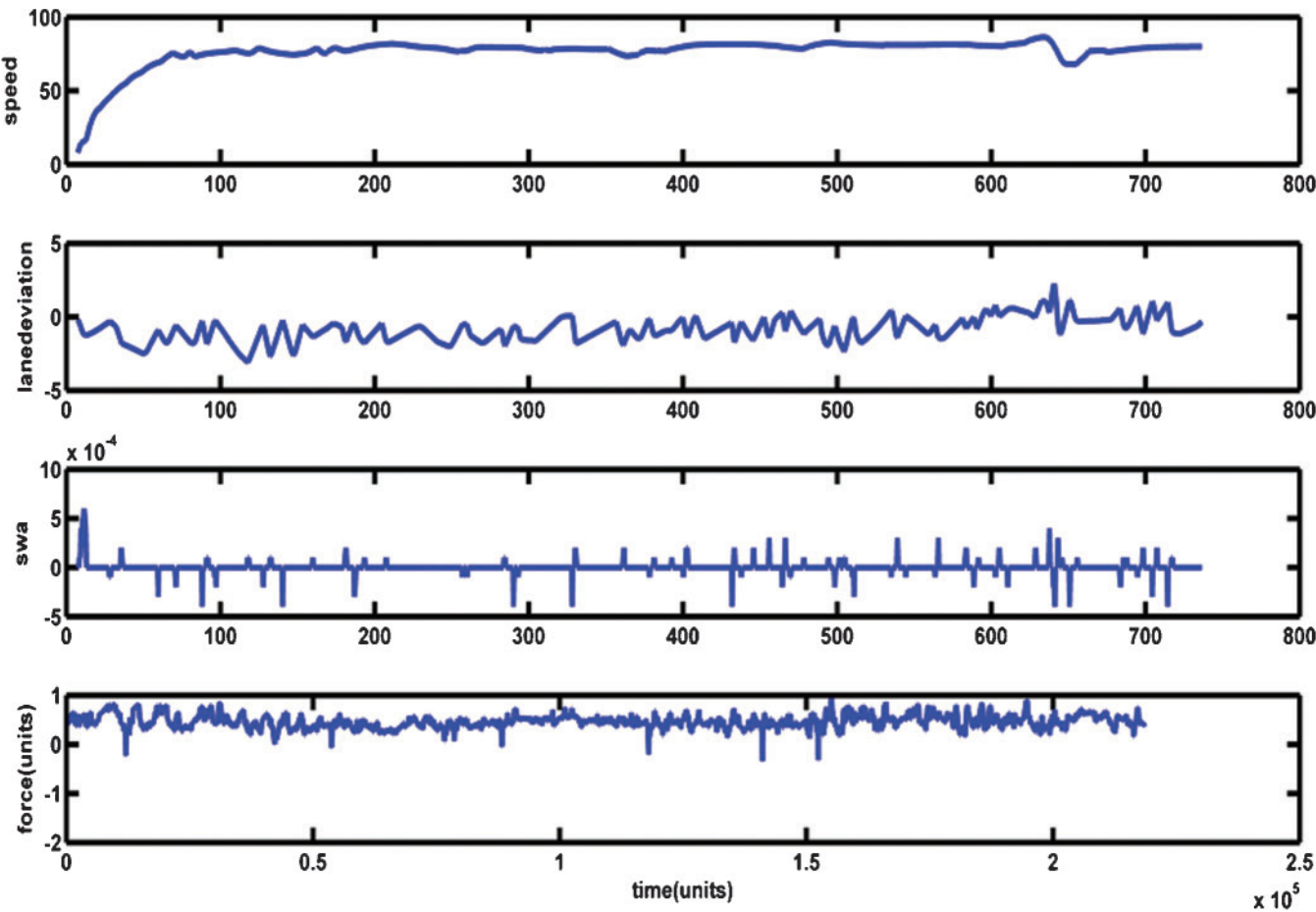


Fig. 4 Sample data from the non-visual information channel: speed (km/h), lane deviation (m), SWA (deg) and force (normalized to the $[-1, +1]$ kgf interval); time (simulator time (s))

Table 2 Different feature vectors designed to investigate the best representation of the phenomena

Visual cues (F_1)	Entropy and control (F_2)	Visual cues and Stds (F_3)
ECM1	IAE	SWA, Sd
ECM2	ISE	Force on SWC, Sd
ADR	SWA, entropy	ECM1
Gaze x , mean	Gaze x , entropy	ECM2
Gaze x , Std	Force on SWC, entropy	ADR
Head Motion in x , Std	Head motion x , entropy	Gaze x , mean
Head Motion in y , Std	Head motion y , entropy	Gaze x , Std
		Head Motion x , Std
		Head Motion y , Std

The methods are a fuzzy inference system (FIS) with a linguistic rule base, an FIS with an extracted rule base using subtractive clustering, and finally an artificial neural network (ANN). FISs are explained in section 4.2 and ANNs in section 4.3. The results of training and testing these systems are given in section 5. Finally, a time-window analysis is performed using only an FIS with the extracted rule base, to observe how narrow the time window can be selected

safely without losing accuracy in the prediction performance.

4.2 Fuzzy inference systems

The approach is to begin with the available relationships from previous studies [3, 6, 7, 10, 11] as linguistic rules, then to justify these rules and finally to add new rules via signal processing and statistical

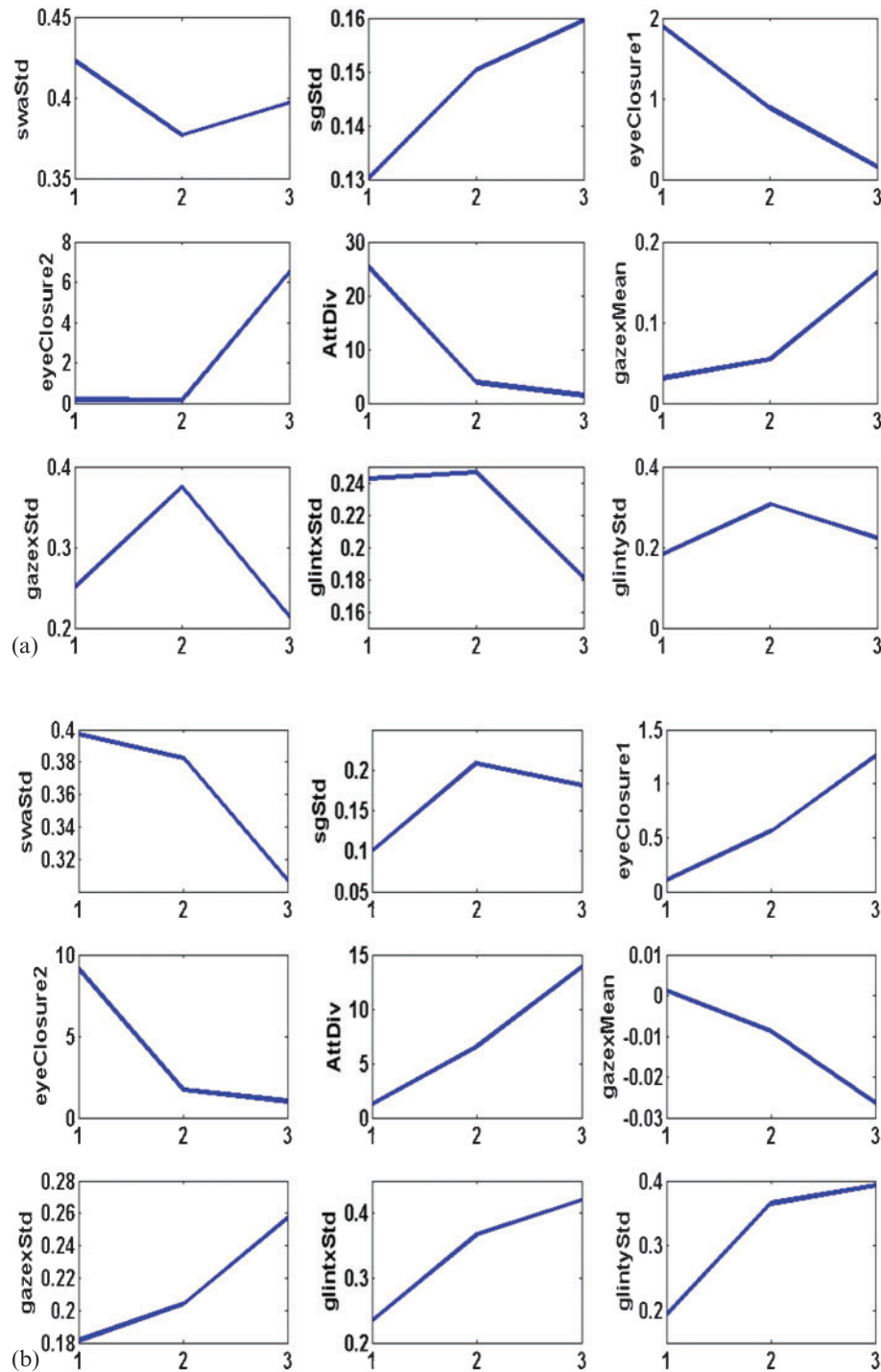


Fig. 5 Change in the variables of F_3 through three sessions for the same subject: (a) alert; (b) drowsy. The x -axis labels 1, 2, and 3 stand for the sessions start, midterm, and final sessions of the data collection

analysis. FISs are employed [9] using only visual cues from a CV system as well; however, feature vector members from both visual and non-visual channels will here be incorporated. The ultimate aim is to reach an algorithm that would predict the alertness state of drivers well before they reach a dangerous

condition. Therefore, the rule base is designed by two different approaches. The first approach uses only linguistic rules defined by previous studies, whereas the second approach aims to extract rules as a set of mathematical relationships from the data itself using fuzzy subtractive clustering. These two

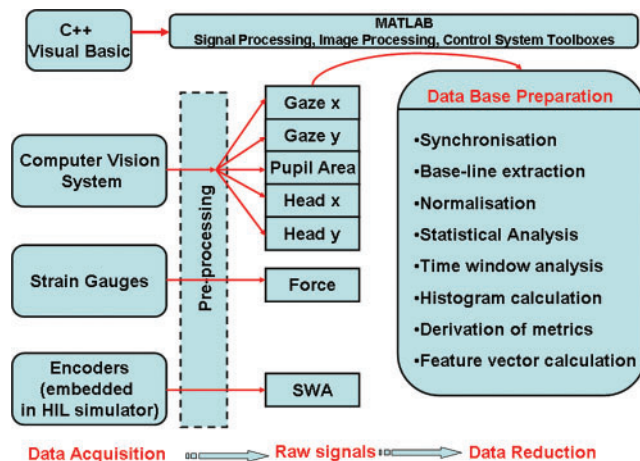


Fig. 6 Signal flow and data reduction (HIL, hardware in the loop)

approaches and three feature vector spaces are compared. As a result of this comparison, the best feature space with the best FIS is selected.

Initially, the rule base of the FIS is constructed as a combination of the rules stated in the literature using raw outputs from the sensors. However, it is soon realized that the system responds too rapidly, leading to false alarms and to failure to take into account properly the measurement history. Such a decision system should not behave as in an on-off mode. For this reason, feature vector calculation over a time window is employed, strengthening the 'reasoning capacity' of the fuzzy system from the observed trends and changes in the data pool, capturing previous and current measurements together. FISs designed using linguistic rules are of the Mamdani (linguistic) type [17] and all the rules are transparent to the designer. An example of such a rule used in the first approach, in the form of a linguistic statement, could be given as follows: 'If ECM1 is *high* and ECM2 is *low*, then the drowsiness or risk level is *high*.'

In fact, this system is like a co-pilot observing the changes in driver behaviour and yielding a prediction on drowsiness level. The second rule base is extracted by a fuzzy subtractive clustering algorithm [18] utilizing the whole database and concluding in a Sugeno-Takagi (S-T) [19] (constructing mathematical linear relationships between inputs and output) FIS with a variable number of rules depending on the distribution of the data. A subtractive clustering method is used in deriving the S-T-based FIS.

The clustering algorithm is an iterative optimization algorithm minimizing the cost function in

$$J = \sum_{k=1}^n \sum_{i=1}^c \mu_{ik}^m \|x_k - v_i\|^2 \quad (8)$$

where the degree of membership μ_{ik} is given by

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c (\|x_k - v_i\| / \|x_k - v_j\|)^{2/(m-1)}} \quad (9)$$

n = number of data points

c = number of clusters

x_k = k th data point

v_i = i th cluster centre

μ_{ik} = degree of membership of the k th datum in the i th cluster

m = constant

When the input precisely matches the centre of the cluster, this definition guarantees that the input will have zero membership coefficients for other clusters. It guarantees that the separate clusters are formed, allowing the rules based on them to be defined. The number of the clusters in this method is automatically concluded by the time that all the data points are attached by a membership function to very dense clusters. Next, the mapping from the input space to output space is then performed with a linear equation using the clusters extracted by this method.

4.3 Artificial neural networks

A robust ANN can be a powerful tool in applications involving a complex classification or decision problem where the answer is required in a short time and a convenient mathematical model is not available. The details of the ANN algorithms will not be given here; however, a comprehensive coverage of the topic can be found in reference [20]. The ANN structure selected for this work is the feedforward multiple-layer perception and trained by back propagation. The training algorithm is chosen as the Levenberg-Marquardt algorithm because of its convenient properties. There are several strategies to decide on the best topology of the network. Cross-validation and statistical approaches are widely known [21, 22]; however, as observed in reference [21], the strategies give an insignificant alleviation in finding the optimum topology. Therefore, starting with the number of inputs as input layer and one layer for output, ANNs with three-layer topology are employed here since they are universal classifiers.

The network structure is a family of networks with $9-x-1$ architecture with three layers; x will be changed to observe the changes in the performance. The best classification performance amongst the network structures starting with an input layer having nine nodes representing the best feature vectors taken from the optimum feature space, which is derived in section 5.2 and ending with a single node for the output of drowsiness level.

5 RESULTS AND ANALYSIS

In order to train prediction and classification systems with supervised learning, a ground truth output is needed. This output represents the true drowsiness level of the driving sessions. This ground truth risk level is obtained by averaging the following scores:

- the driver's own assessments of his or her condition;
- the independent grading from 20 assessors watching the videos from the simulation;
- the experiment designer's view from the NIR frontal face videos of drivers.

The assessment is based on a scale from 1 to 5, with 1 representing a fully alert and 5 a significantly drowsy state.

5.1 Comparison of prediction performances

As two approaches are followed to design different FIS systems, it is necessary to compare their performances over three previously defined feature vector spaces. The Mamdani system is built based on *a priori* linguistic relationships besides being transparent. This means that any rule can be added or removed linguistically, making the system flexible. However, the prediction of the system when tested with unknown vectors gave the average drowsiness level of 3. Hence, it is not able to predict extreme cases with the available rule base, mostly outputting scale 3 from the [1–5] interval indicating a moderate level of drowsiness (Fig. 7).

On the other hand, the S–T FIS system depends completely on the database and learns all the mapping from the data via cluster analysis. Its success is much higher than the Mamdani system in correctly labelling the drowsy sessions on a [1–5] scale; however, the system is highly data driven, and the rules are not intuitive. It may output false classification results in some extreme cases. It is able to follow the drowsiness level defined by average

of experimenter's, independent assessors', and subject's own drowsiness gradings on the [1–5] scale. An S–T FIS system based on the F_1 (visual clues) feature space gives reasonably good results when 18 completely unknown sessions were used as input (Fig. 8).

The performances of the S–T FIS system over F_2 (entropy and control) and F_3 (visual clues and Std) are given in Figs 9 and 10 respectively. As can be seen in Fig. 9, F_2 is not a good feature vector space and F_3 appears to be the best (Fig. 10). The results are also filtered using heuristics to damp the overshoots of the system and to rescale it into a [1–5] scale.

The total testing and validation session results are summarized in Table 3, illustrating that F_1 and F_3 are the best feature spaces amongst the heuristically designed spaces; however, F_3 did not perform as well as expected. This can be explained by the fact that some of the feature members are not following a consistent trend in the database. These members can be identified and a smaller weight factor can be attached to them, so that the system performance is not affected greatly.

5.2 Selecting outperforming members from feature spaces

In statistics, correlation analysis helps experimenters to find statistically correlated variables and their dependences. The correlation value r shows how the trend in two separate variables match, and the p value or significance value indicates their linear dependence. If the correlation analysis of two variables yields high correlation values and low significance, it is concluded that the two variables are linearly independent, however highly correlated. The previously defined risk function is correlated with each feature member over 150 sessions and correlation coefficients with significance or p values are calculated (Table 4) to identify the feature members highly correlated to the drowsiness level.

If the members with $p < 0.05$ are included as shown in Table 5, a significant improvement is observed giving only two false classifications over 150 mixed sessions with test and training data. In addition to this, there has been no post-processing of results using heuristic filtering and rescaling. The system is still able to detect the states correctly, giving only one false alarm and one miss in test data (Fig. 11).

Using nine members of the vector space F_4 , as explained in section 4.2, an ANN from a family of 9–

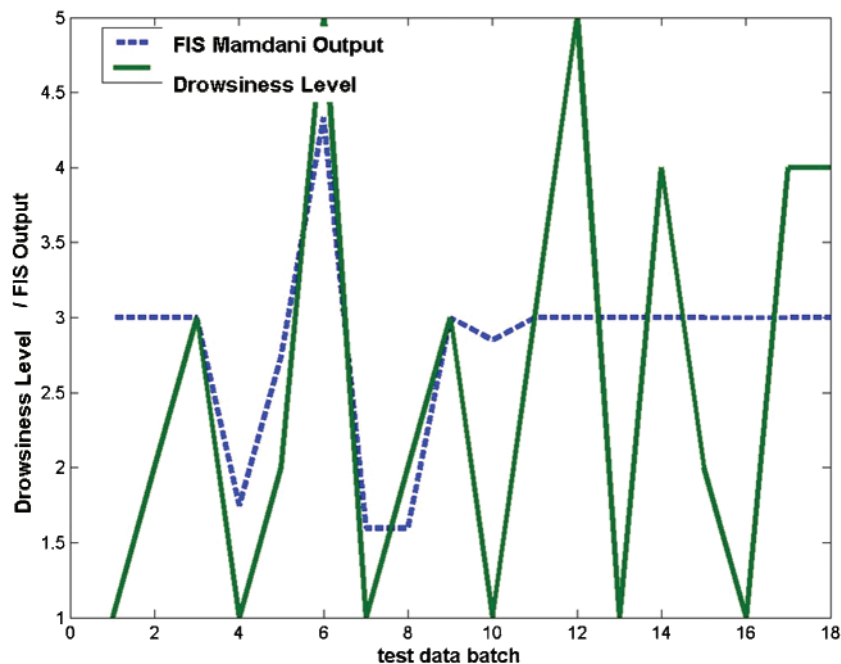


Fig. 7 Output of the FIS-Mamdani system with linguistic rules versus subjective risk function (F_1 feature vector space)

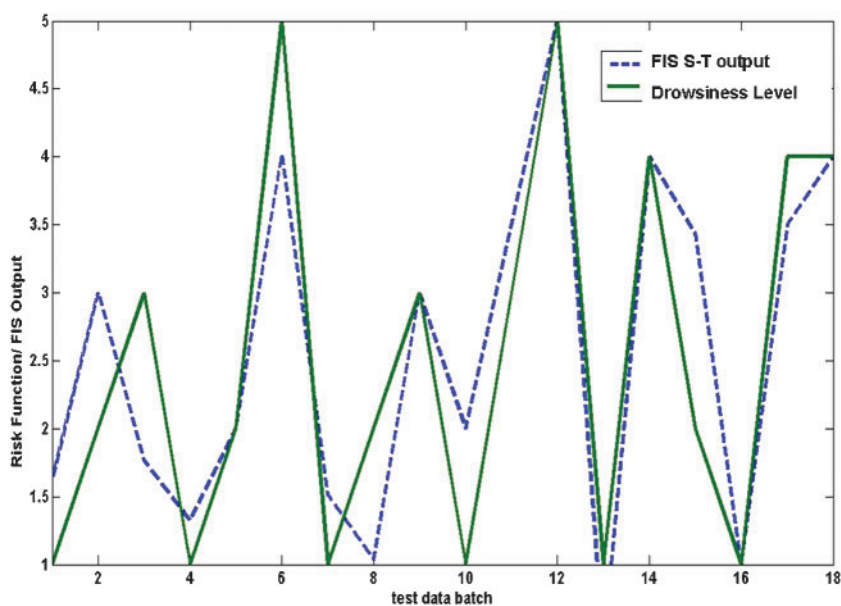


Fig. 8 Output of the S-T FIS versus risk function over F_1 space

$x-1$ architecture is used to see whether there is any difference in performance when a different algorithm is used for classification. It was observed that, while x has values in the $[2, 8]$ interval the performance of the ANN classification was found with the 9-6-1 architecture. However, as can be seen in Figs 12 and 13, there is no significant difference between the ANN and the FIS in terms of performance.

6 TIME WINDOW ANALYSIS

In this part, the quest is for the best time window over which the feature vectors were calculated. The best time window represents the following characteristics:

- the ability to capture the trend and necessary information content in the data, as it has to

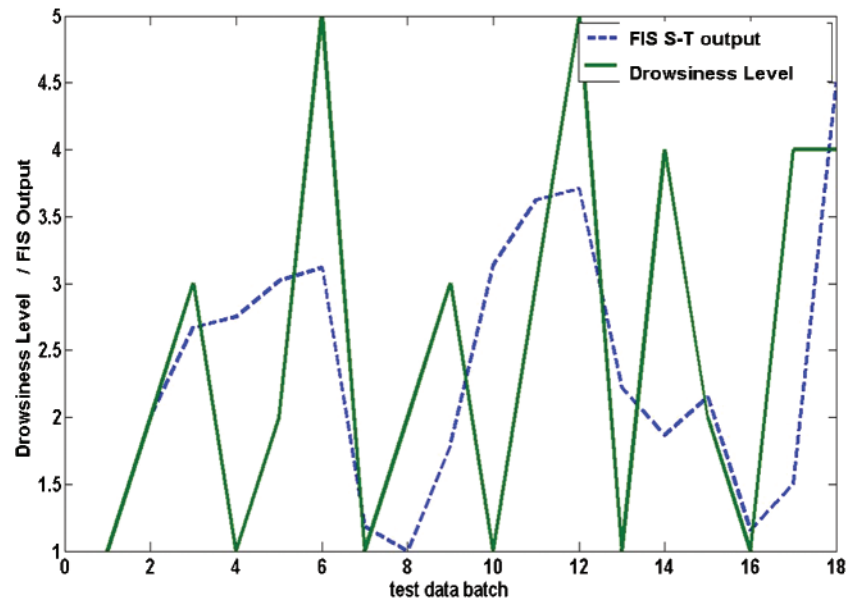


Fig. 9 Output of the S-T FIS system versus risk function over F_2 space

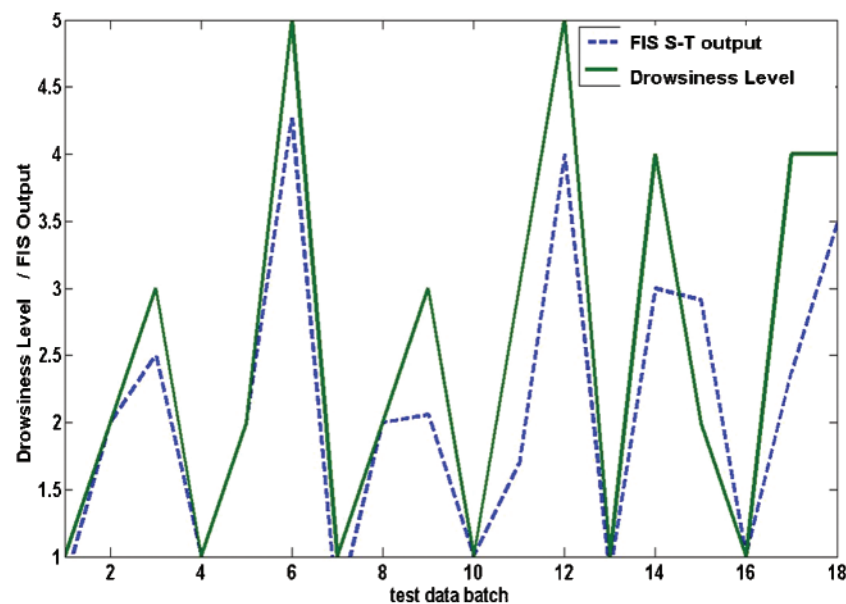


Fig. 10 Output of the S-T FIS system versus risk function over F_3 space

Table 3 Comparison of performances for two FIS systems utilizing the three chosen feature vectors

Performance comparison	F_1		F_2		F_3	
	Success (%)	False alarm (%)	Success (%)	False alarm (%)	Success (%)	False alarm (%)
Mamdani	90	10	80	20	85	15
S-T	98	0	90	10	95	5

represent sufficient history of the signals so that trends become more visible and distinguishable;

- (b) the ability to allow a suitable update on driver status, so that a preventative system can act promptly.

In order to investigate the time window effect on the system performance in decision making, four different data batches were prepared on the basis of the original reduced raw data. The time windows were set at 12 min, 6 min, 3 min, and 1.5 min.

Table 4 Feature members and corresponding correlation coefficients with significance values

Feature vector member	Correlation coefficient R	Significance p
ECM1	0.5886	0.000
ECM2	-0.3951	0.000
ADR	0.2642	0.002
Gaze x , mean	-0.0384	0.658
Gaze x , Std	-0.0383	0.659
Head motion x , Std	0.3256	0.001
SWA, Std	-0.1310	0.130
Head motion y , Std	0.3991	0.000
IAE	0.201	0.0193
ISE	0.174	0.0429
SWA, entropy	-0.227	0.0079
Gaze x , entropy	-0.083	0.3300
Force, entropy	-0.179	0.0360
Head motion x , entropy	-0.18	0.0342
Force, Std	0.196	0.0224
Head motion y , entropy	-0.198	0.0213

Table 5 Feature members with $p < 0.05$ defining best feature space (F_4)

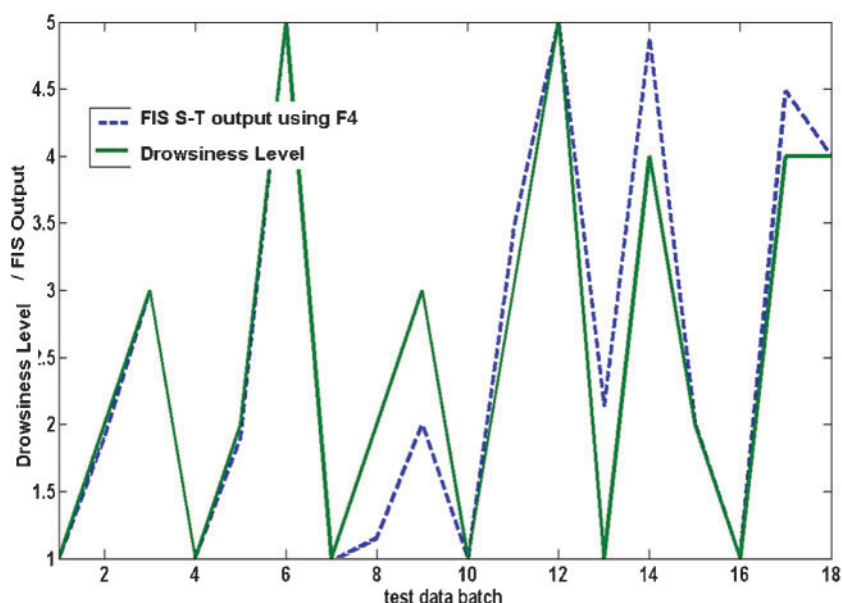
Feature vector member	Significance p
ECM1	0.000
ECM2	0.000
ADR	0.002
Head motion x , Std	0.001
Head motion y , Std	0.000
IAE	0.019
ISE	0.042
SWA, entropy	0.007
Force, entropy	0.036
Force, Std	0.022

In Fig. 14 the result for training a data batch using a 6 min time window can be seen. As can be deduced, the prediction is highly accurate, missing only two data points slightly amongst 270 data points. However, when presented with completely unknown feature vectors, the trained FIS gives a reduced performance, as seen in Fig. 15.

The next two further steps of reduction in time window size are 3 min and 1.5 min. The prediction performances of these time windows on training data and test data are given in Figs 16 and 17 respectively for 3 min, and in Figs 18 and 19 respectively for 1.5 min. It is concluded that 3 min or 1.5 min can be used to predict driver drowsiness level with some acceptable false alarm rate, not exceeding 10 per cent. For further reduction in the time window the sensor reliability has to be increased.

Overall, the following observations are recorded from time window analysis.

1. If the time window is narrowed too much as in the case of 1.5 min, the response becomes noisy.
2. The fluctuating response of the narrower time windowed fuzzy systems could be because of output vector interpolation, resulting in quantization error. Because the actual output values between two known output values is assumed to change linearly, this may not be representing the real case.

**Fig. 11** Test results of the best feature vector space

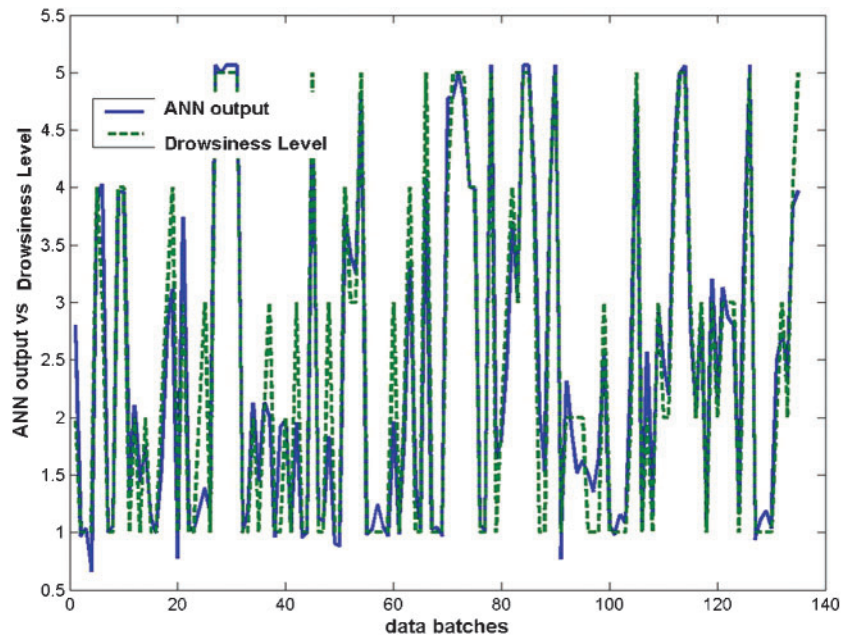


Fig. 12 ANN output versus ground truth drowsiness level showing a good match training goal = 0.17

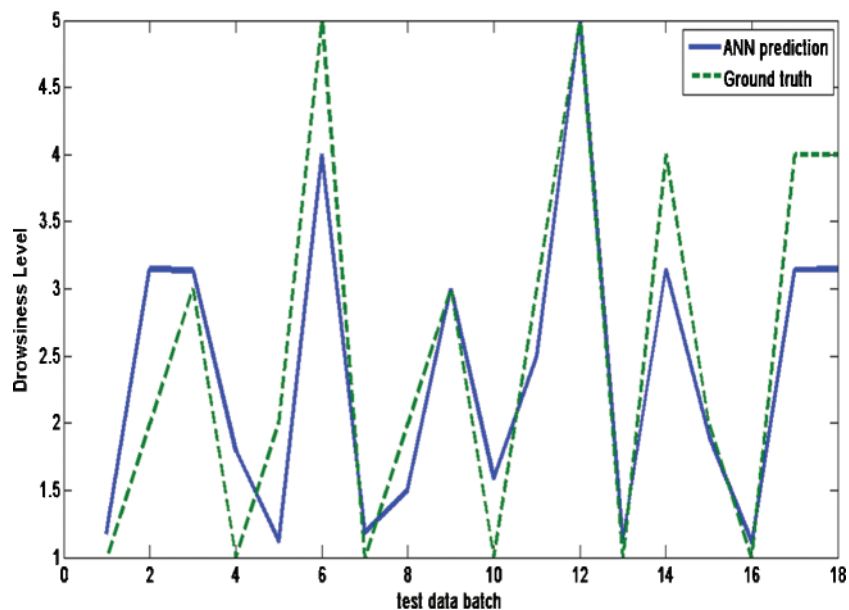


Fig. 13 ANN prediction results for a test data batch

7 CONCLUSIONS AND FURTHER RESEARCH DIRECTIONS

It was shown that the proposed driver monitoring system can predict drowsiness levels related to drowsiness in 98 per cent of the cases previously examined manually and 89 per cent of the unknown cases. The approach has a sound structure beginning with the application of linguistic rules and ending with the development of a mathematically

supported rule base with the best feature space identified through correlation coefficient calculation. No significant improvements are observed when an ANN was employed as a prediction algorithm.

The high correlation coefficients of selected (e.g. eye closure) and designed (e.g. ADR) metrics, with a subjectively defined risk function on a [1–5] scale, provide a quantitative way of representing drowsiness. The correct classification rate of the artificial intelligence system described suggests a starting

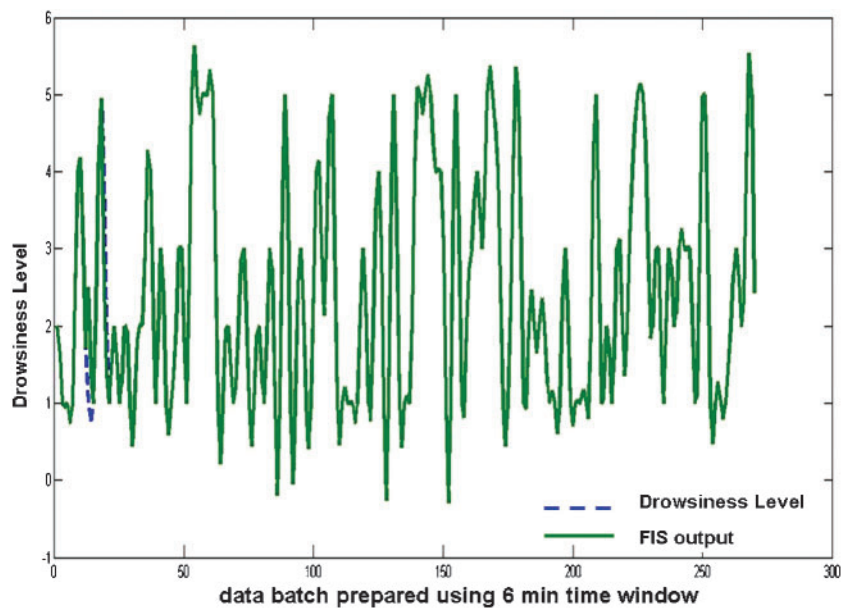


Fig. 14 Prediction results versus ground truth drowsiness level for training data with 6 min time windows

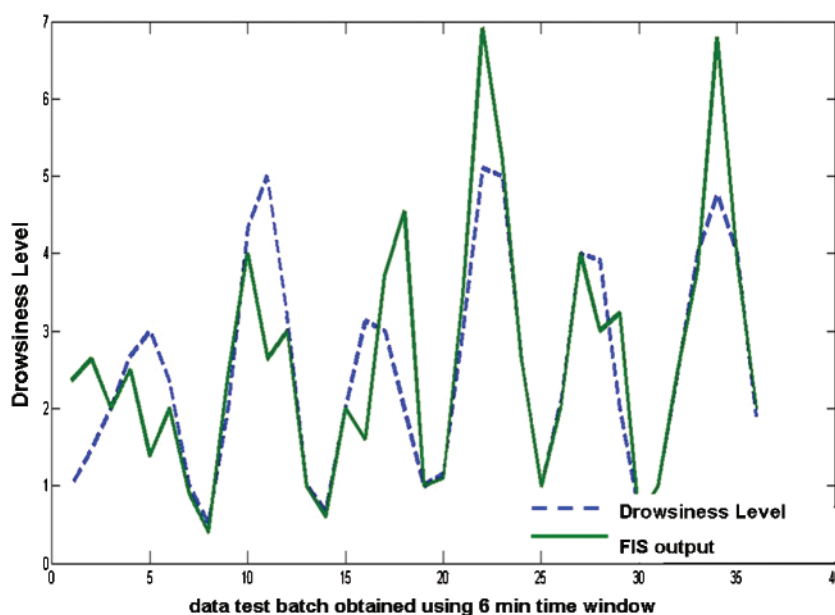


Fig. 15 Prediction results versus ground truth drowsiness level for test data with 6 min time windows

point for a robust driver monitoring system for future automobiles. In its present form, the system is capable of outputting the driver alertness level and associated risk level every 12 min. In fact, with the time window analysis, it is shown that this duration can be as narrow as 1.5 min. If the time windows are selected as overlapping time windows, a continuous prediction of drowsiness level becomes possible.

The reliability of the system is found to be significantly high, giving only one false alarm and one

miss over 150 sessions. This offers a good opportunity for applications in real conditions with high reliability and low false alarm rate. The false alarm rate increases as the time windows become narrower; however, this can be solved by improving the reliability in raw signals.

This study can be considered as a probe investigation to reconsider indicative metrics and to enhance the prediction results of the monitoring algorithm. However, several shortcomings of the system and

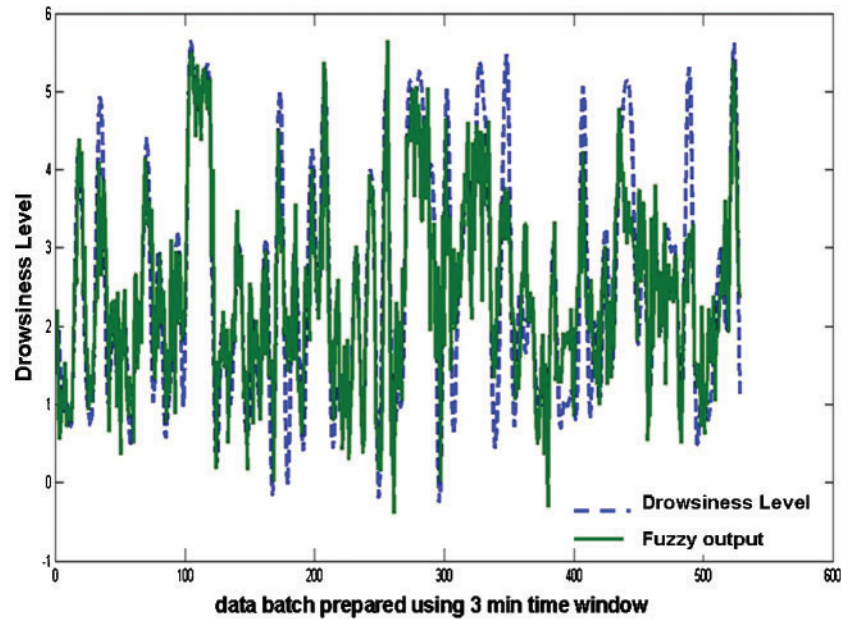


Fig. 16 Prediction results versus ground truth drowsiness level for training data with 3 min time windows

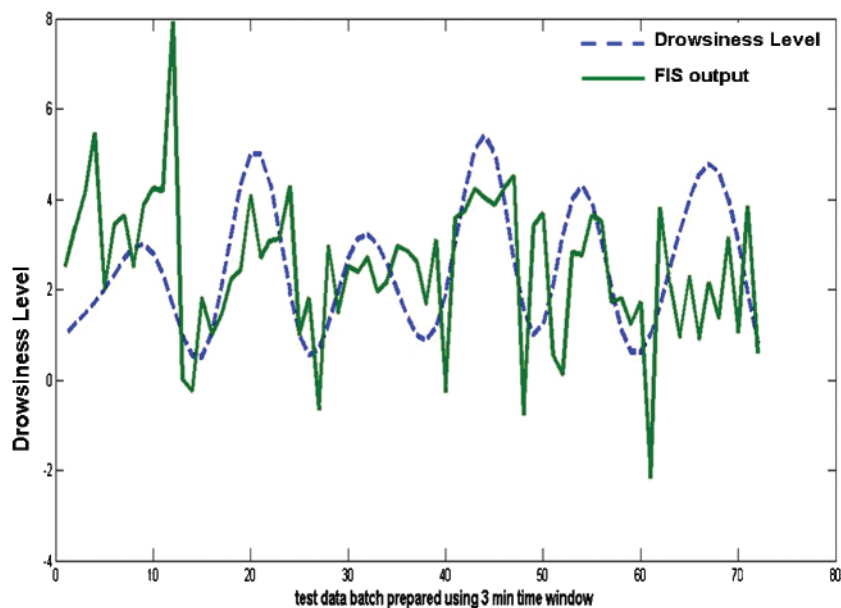


Fig. 17 Prediction results versus ground truth drowsiness level for test data with 3 min time windows

proposed algorithms must be noted and improved in further studies. Three directions are identified to expand or enhance this work.

1. If the overlapping time windows are used to obtain the continuous output from the system, to prevent the undesired fluctuation in the response an estimation filter needs to be applied. Since the hidden Markov models are the most general case of

the estimators enclosing the derivation of the Kalman filters as well, they can be employed to smooth out the system output. If the Kalman filter or its versions (unscented Kalman Filter and extended/Kalman Filter) is to be used it is beneficial to employ an analogy of the Kalman filter for sensor fusion; the states can be selected as the inputs to the decision system and the estimation result from an FIS or an ANN can be considered as the sensor

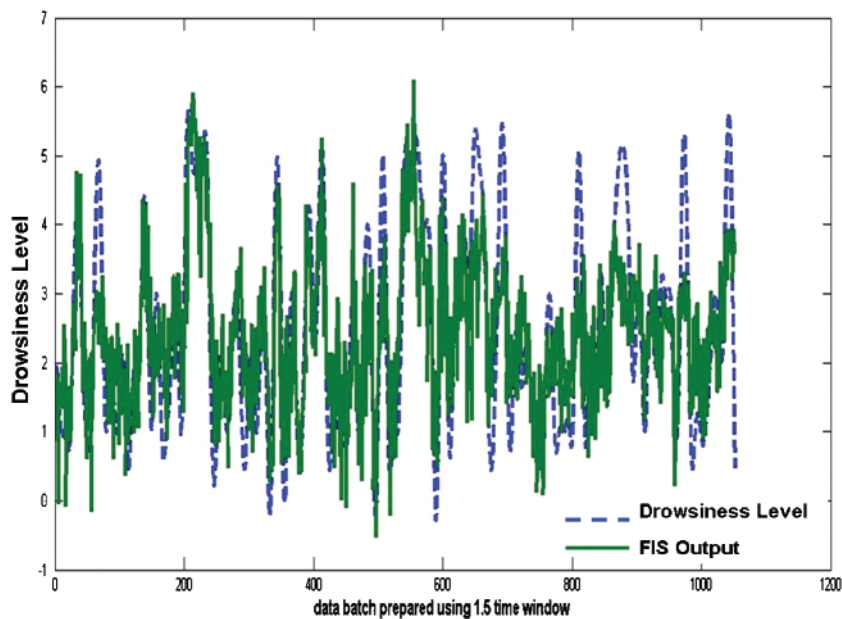


Fig. 18 Prediction results versus ground truth drowsiness level for training data with 1.5 min time windows

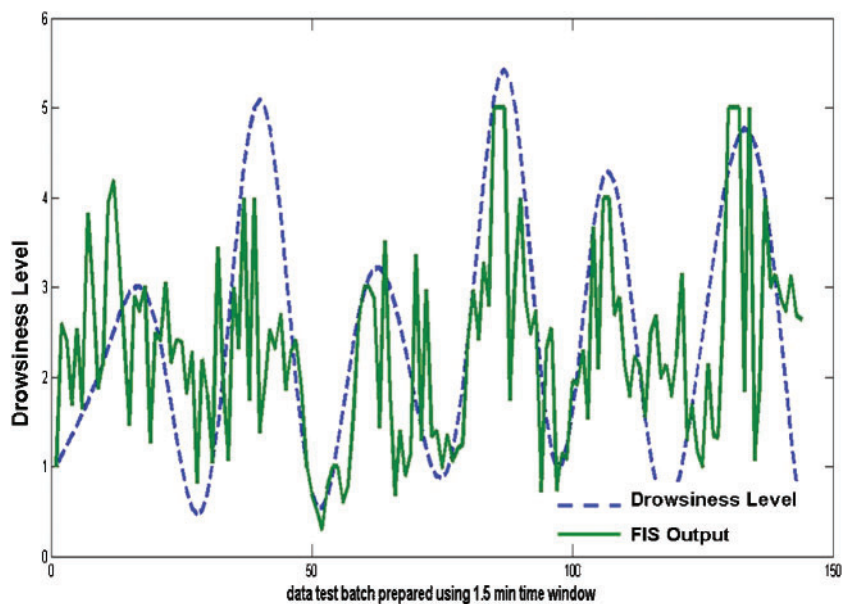


Fig. 19 Prediction results versus ground truth drowsiness level for test data with 1.5 min time windows

output whereas the ground truth represents the actual measurement excluding the noise in estimation introduced by the FIS or the actual sensors in the monitoring system.

2. Another topic to explore more is enhancing the CV system to be used in daylight conditions and in real vehicles. As is widely known, eye trackers using the NIR brightpupil technique are prone to fail because the sunlight causes the pupil response to become indistinguishable and often the

corneal reflection on the eyeball is hard to track. It should be emphasized that, for the probe study, the NIR technique is employed with partial adaptation to lighting conditions using an evolutionary approach. However, template matching and recognition schemes are suggested for use in daylight conditions to alleviate the sensor failure. During night driving sessions the NIR system is expected to give a better performance since there is no interference from sunlight. Therefore, a dual

mode in the CV is needed to monitor the driver's day and night with best available resolution and false alarm rate. Alternatively, surrogate metrics can be investigated to replace the visual metrics when the sensors fail or the source become unavailable.

3. The final efforts should be put into implementing this monitoring system in real vehicles under real road conditions. Since the nature of the experiment involves a high risk to participants, a test-track experiment closed to traffic will be the first step to test the system. A sufficiently large, demographically balanced group of participants should be included in these experiments to observe whether the observations on metrics agree with real conditions. Since the system will be applied in a real car, on-board digital signal-processing systems will be appropriate to use in the final version; however, a synchronized data acquisition unit is enough for preliminary test-track experiments. In terms of sensors, a dual-mode CV system for eye tracking, Controller area network bus signals from the on-board diagnostics port, pressure measurements on the steering wheel, brake, and accelerator pedal and a separate CV system for lane tracking should be included. For synchronization purposes the GPS signal needs to be included.

REFERENCES

- 1 Bergasa, L. M., Nuevo, J., Sotelo, M. A., and Vazquez, M. Real-time system for monitoring driver vigilance. In Proceedings of the IEEE Intelligent Vehicle Symposium, 2004, pp. 78–83 (IEEE, New York).
- 2 Bittner, R., Hana, K., Pousek, L., Smrcka, P., Schrieb, P., and Vysoky, P. Detecting of fatigue states of a car driver. In Proceedings of the First International Symposium on *Medical data analysis (ISMDA 2000)*, Lecture Notes in Computer Science, vol. 1933, Frankfurt, Germany, 29–30 September 2000, pp. 260–273 (Springer-Verlag, Berlin).
- 3 Brookhuis, K. A. and De Waard, D. The use of psychophysiology to assess driver status. *Ergonomics*, 1993, **36**, 1099–1110.
- 4 Chapman, P., Ismail, R., and Underwood, G. Waking up at the wheel: accidents, attention and the time-gap experience. In *Vision in vehicles VII*, Proceedings of the Seventh International Conference on *Vision in vehicles* (Ed. A. Gale), Marseilles, France, September 1997, 1999, pp. 131–138 (Elsevier, Amsterdam).
- 5 Dinges, D. F. An overview of sleepiness and accidents. *J. Sleep Res.*, 1995, **4**(Suppl. 2), 4–14.
- 6 Galley, N., Andr s, G., and Reitter, E. Driver fatigue as identified by saccadic and blink indicators. In *Vision in vehicles VII*, Proceedings of the Seventh International Conference on *Vision in vehicles* (Ed. A. Gale), Marseilles, France, September 1997, 1999, pp. 49–59 (Elsevier, Amsterdam).
- 7 Rau, R. and Knipling, P. A valid psychophysiological measure of alertness as assessed by psychomotor vigilance, 2005, available from <http://www.fmcsa.dot.gov/documents/tb98-006.pdf>.
- 8 Zhu, Z. and Ji, Q. Robust real-time eye detection and tracking under variable lighting conditions and various face orientations. *Computer Vision Image Understanding*, 2005, **98**, 124–154.
- 9 Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R., and Lopez, M. E. Real-time system for monitoring driver vigilance. *IEEE Trans. Intell. Transportation Systems*, 2006, **7**(1), 63–72.
- 10 Pilutti, T. and Ulsoy, A. G. Identification of driver state for lane-keeping tasks. *IEEE Trans. Systems, Man, Cybernetics Part A: Systems Humans*, 1999, **29**(5), 486–502.
- 11 Swingler, K. and Smith, L. S. Producing a neural network for monitoring driver alertness from steering actions. *Neural Computing Applic.*, 1996, **4**(2), 96–104.
- 12 Sayed, R. and Eskandarian, A. Unobtrusive drowsiness detection by neural network learning of driver steering. *Proc. Instn Mech. Engrs, Part D: J. Automobile Engineering*, 2001, **215**, 969–975.
- 13 Gress-Hernandez, N., Esteve, D., and Bekiaris, A. IMU: integrated monitoring unit of save diagnostic system. In Proceedings of the IFAC Workshop on *Intelligent components for vehicles* (Ed. A. Ollero), Seville, Spain, 23–24 March 1998, pp. 147–151 (Elsevier, Amsterdam).
- 14 Apostoloff, N. and Zelinsky, A. Vision in and out of vehicles: integrated driver road scene monitoring. *Int. J. Robotics Res.*, 2004, **23**(4–5), 513–518.
- 15 Victor, T. A technical platform for driver attention research, accessed 21 October 2006, available from <http://www.tech.volvo.se/human.html>, <http://www.nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/7.PDF>.
- 16 MacAdams, C. C. Understanding and modelling the human driver. *Veh. System Dynamics*, 2003, **40**(1–3), 101–134.
- 17 Mamdani, E. H. and Assilian, S. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-Mach. Stud.*, 1975, **7**(1), 1–13.
- 18 Bezdek, J. C. *Pattern recognition with fuzzy objective function algorithms*, 1981 (Plenum, New York).
- 19 Sugeno, M. *Industrial applications of fuzzy control*, 1985 (Elsevier, Amsterdam).
- 20 Haykin, S. *Neural networks: a comprehensive foundation*, 1999 (Prentice-Hall, Englewood Cliffs, New Jersey).
- 21 Andersen, T. and Martinez, T. Cross-validation and MLP architecture selection. In Proceedings of the IEEE International Joint Conference on *Neural*

networks (IJCNN 99), Washington, DC, USA, 1999, vol. 3, pp. 1614–1619 (IEEE, New York).

- 22 **Ripley, B. D.** Statistical ideas for selecting network architectures. In *Neural networks: artificial intelligence and industrial applications* (Eds B. Kappen and S. Gielen), 1995, pp. 183–190 (Springer-Verlag, Berlin).

APPENDIX 1

Computer vision algorithm details

Here, the technical details of the CV system are given as well as the algorithms designed for false-proof operation of the system. Appendix 1 is divided into hardware, image properties and algorithm development parts.

Hardware

A CMOS camera giving a video stream of 30 frames/s is selected. Since the bright-pupil technique is used, camera selection also included the specific wavelength used in NIR range. The human eye retina is most responsive and the humorous liquid of the eye is least absorbent at 880 nm wavelength; therefore the camera's sensor has to be the most efficient at this particular wavelength as well. Furthermore, to avoid the environmental interference at 880 nm, an optical absorption filter was placed in front of the camera lens. Finally, a circular illumination source was arranged coaxially with the camera optical axis to gather the light reflected from retina with the most efficiency. The full hardware arrangement can be seen in Fig. 20.

Image properties

Unprocessed grey-level images [0–255] from a video stream had a resolution of 640×480 with a signal-to-noise ratio (SNR) of 54.04 dB. Since the SNR caused segmentation problems the images are pre-processed using a Gaussian filter.

Algorithm development

An evolutionary spatial filter is proposed for autonomous eye tracking. ANNs are used to verify the eye region after obtaining eye candidates by a simple threshold exploiting corneal reflection of light source on the eye ball. On the other hand, the pupil area is

accurately segmented from detected the eye region using spatial filters optimized by genetic algorithms (GAs). When closely examined, the system separates the high-level supervisory algorithms from the low-level segmentation and detection part. In this way, high-level algorithms can supervise a low-level process operation (in an ANN, verifying image threshold candidates) and tune the parameters of the candidates (in a GA, optimizing spatial filter parameters) without disturbing their operation and affecting their speed.

In addition to this hierarchical structure, evolutionary spatial filters are used as the kernel of the mean shift algorithm to improve their tracking capability. Segmenting the pupil area by using the spatial connectivity of pixels and gradients, the spatial filter is the best way of representing empirical estimation of the pixel distribution constituting the pupil area. Therefore, the mean shift drift is not observed and the algorithm tracked the eyes when there are available pupils. The signal flow of the algorithm is given in Fig. 21.

APPENDIX 2

Perclos versus the proposed metrics and cumulative results

In this analysis, a justification of the proposed metrics is given as they were used instead of the traditional PERCLOS. Correlation coefficients are calculated in two separate sets: first, PERCLOS versus all proposed metrics (ECM1, ECM2, and ADR); second, ground truth versus all visual metrics. The results are given in Tables 6 and 7 respectively. The full database including the drowsy and normal condition data of 30 drivers is used except the cases where data were corrupted or there was a sensor failure. Each subject is given a number and 'sub2' represents subject two in normal conditions whereas 'sub2d' contains the same participant's driving data from drowsy sessions. Each driver normally has 12 min data selected from beginning, middle-term, and final parts of the 1.5 h of total data. Visual metrics are calculated over 1 min time windows; therefore giving 12 (mins) \times 3 (samples) \times 2 (drowsy/normal) min data. This analysis includes 1620 min of simulator data, since the ground truth is known only for the initial 12 min; the drowsiness level is assumed to remain the same in between. This assumption results in correlation coefficients lower than 0.9; therefore Table 7 must be examined under this assumption. With a ground-truth-obtaining method

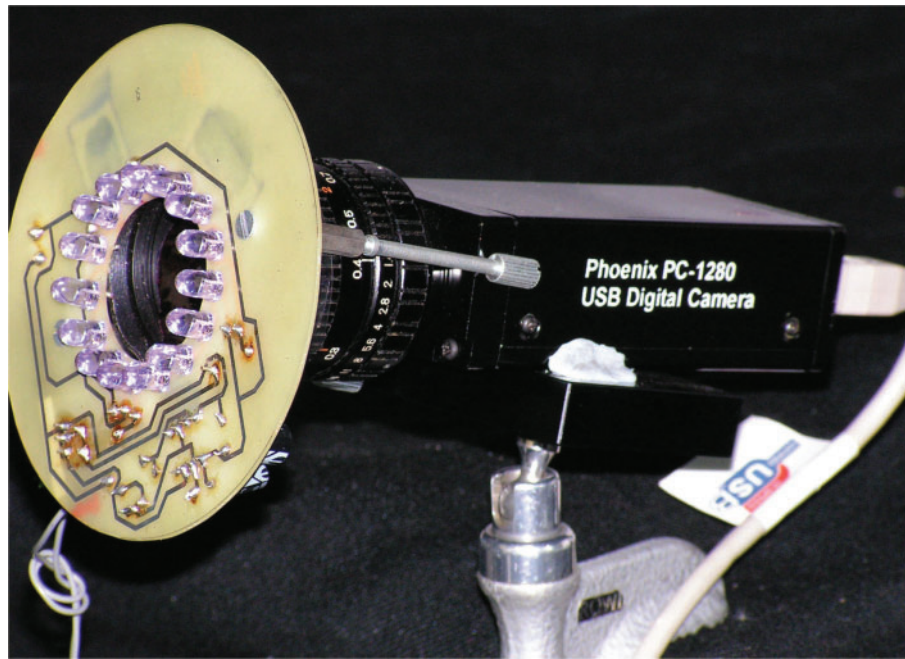


Fig. 20 CV system hardware

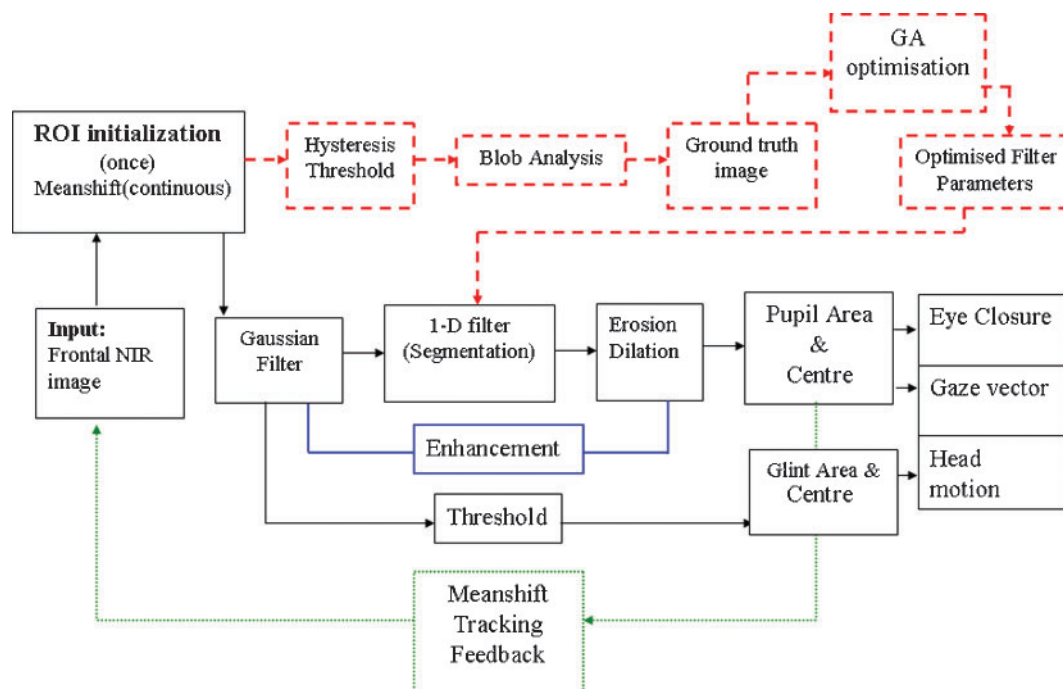


Fig. 21 Signal flow of the CV system

allowing assessments of drowsiness level with finer and denser time intervals, the correlation coefficients may greatly improve.

To summarize the main observations, in eight sessions, ECM2 and ADR are better correlated with PERCLOS whereas, in 37 sessions, ECM1 is better correlated. In 21 sessions, PERCLOS is better correlated with ground truth; in eight sessions ECM1, in 17

sessions ECM2, and in 15 sessions ADR are better correlated with ground truth. Taken into account as a group, for 24 of 45 sessions the proposed metrics had higher correlation coefficients with the ground truth signal. If it is taken into account that ECM1 is highly correlated to PERCLOS from Table 6, the proposed metrics can cover the cases where PERCLOS performed better as well. From Table 7 ECM1

Table 6 Perclos versus all metrics correlation

Subject	ECM1	ECM2	ADR
sub2	0.980 524	-0.382 33	0.987 838 77
sub2d	0.993 855	-0.504 81	0.889 215 64
sub3	0.902 02	-0.912 41	0.884 195 42
sub3d	0.943 14	-0.399 49	0.841 776 83
sub4d	0.985 319	-0.153 27	-0.183 138 53
sub5	0.994 491	-0.486 54	-0.152 021 03
sub5d	0.943 55	-0.691 02	0.003 359 46
sub6	0.989 718	-0.508 79	0.849 529 28
sub6d	0.991 583	-0.168 04	0.653 796 18
sub7	0.934 496	-0.627 16	0.205 337 51
sub7d	0.585 919	-0.981 39	0.668 589 21
sub8	0.996 959	-0.587 19	0.229 152 81
sub8d	0.994 227	-0.438 51	0.154 246 65
sub9	0.996 641	-0.526 95	0.526 021 95
sub9d	0.988 072	-0.417 16	0.825 033 86
sub10	0.994 063	-0.533 87	0.894 363 79
sub10d	0.138 654	-0.282 96	0.683 060 6
sub11	0.939 854	-0.365 411	0.078 704 57
sub11d	0.397 402	-0.7888	0.767 512 29
sub12	0.978 28	-0.380 74	0.183 483 96
sub12d	0.838 077	-0.591 45	0.038 884 06
sub13	0.024 411	-0.685 16	0.806 261 65
sub13d	0.971 27	-0.497 11	0.812 001 85
sub14	0.994 478	-0.875 55	0.613 558 28
sub14d	0.946 819	-0.980 77	0.153 763 24
sub15d	0.910 81	-0.893 51	-0.210 069 43
sub16	0.943 23	-0.4443	0.663 609 6
sub16d	0.994 345	-0.486 19	-0.404 355 47
sub17d	0.135 606	-0.291 62	0.242 180 72
sub18	0.876 196	-0.783 16	0.376 149 33
sub18d	0.953 211	-0.487 55	0.888 537 01
sub19d	0.995 498	-0.799 63	0.733 014 73
sub20	0.992 847	-0.232 97	0.618 826 77
sub20d	0.996 482	-0.638 23	-0.072 231 38
sub21	0.996 75	-0.651 53	0.065 888 49
sub21d	0.944 578	-0.512 85	0.885 256 63
sub22	0.943 305	-0.333 02	-0.035 561 24
sub22d	0.974 358	-0.496 69	0.358 078 89
sub23	0.515 884	-0.903 32	0.582 222 11
sub24	0.987 734	-0.824 85	0.163 920 85
sub24d	0.979 887	-0.676 89	0.446 773 94
sub25	0.981 433	-0.669 02	0.777 263 17
sub25d	0.922 291	-0.547 61	0.903 852 25
sub26	0.995 652	-0.5398	0.755 209 05
sub27	0.976 949	-0.621 28	0.709 741 21
sub27d	0.976 306	0.168 388	0.157 887 52

Table 7 Ground truth versus all metrics correlation

Subject	ECM1	ECM2	ADR	PERCLOS
sub2	0.410 627	-0.684 86	0.479 094	0.454 542
sub2d	0.909 71	-0.531 52	0.795 466	0.945 582
sub3	0.530 127	-0.837 84	0.633 444	0.834 281
sub3d	0.525 77	-0.429 56	0.488 69	0.693 86
sub4d	0.575 896	0.280 905	-0.330 95	0.557 009
sub5	0.459 407	-0.457 59	-0.169 16	0.488 035
sub5d	0.331 587	-0.545 99	0.516 726	0.424 888
sub6	0.806 276	-0.5112	0.579 716	0.843 57
sub6d	0.584 258	-0.131 06	0.758 317	0.593 455
sub7	0.850 615	-0.5712	0.340 076	0.964 156
sub7d	0.392 129	-0.7328	0.614 343	0.754 581
sub8	0.578 167	-0.457 83	0.636 02	0.589 824
sub8d	0.153 543	-0.527 25	0.407 009	0.151 405
sub9	0.310 136	-0.481 57	0.611 159	0.296 355
sub9d	0.437 94	-0.719 03	0.238 004	0.498 495
sub10	0.781 63	-0.414 28	0.768 09	0.767 773
sub10d	0.549 275	-0.439 37	0.736 499	0.440 256
sub11	0.738 341	0.625 328	-0.132 63	0.916 521
sub11d	0.328 863	-0.449 42	0.291 788	-0.000 78
sub12	0.612 578	-0.4189	0.226 489	0.622 48
sub12d	0.031 191	-0.413 03	0.650 728	0.198 691
sub13	0.351 321	-0.579 75	0.495 724	0.590 363
sub13d	0.682 807	-0.363 26	0.654 161	0.746 249
sub14	0.702 15	-0.687 23	0.520 146	0.711 119
sub14d	0.188 708	-0.3015	0.713 059	0.253 033
sub15d	0.268 09	-0.3309	0.336 51	0.313 434
sub16	0.876 696	-0.403 49	0.586 261	0.985 952
sub16d	0.110 277	0.532 723	-0.573 91	0.107 976
sub17d	0.352 376	0.451 272	0.374 619	0.215 784
sub18	0.652 881	-0.817 73	0.473 788	0.869 043
sub18d	0.713 266	-0.438 89	0.756 445	0.819 25
sub19d	0.591 13	-0.596 09	0.613 804	0.632 818
sub20	0.332 996	0.375 031	0.299 438	0.342 125
sub20d	0.680 254	-0.810 45	-0.124 74	0.685 347
sub21	0.199 827	-0.263 48	0.543 029	0.217 966
sub21d	0.701 362	-0.547 96	0.687 716	0.847 767
sub22	0.291 679	-0.598 07	-0.145 56	0.399 721
sub22d	0.103 12	-0.367 32	-0.095 04	0.183 055
sub23	-0.147 96	-0.295 75	-0.306 18	0.258 917
sub24	0.353 832	-0.362 14	0.395 705	0.365 81
sub24d	0.188 194	-0.383 12	-0.290 25	0.266 707
sub25	-0.103 77	-0.141 45	-0.080 17	-0.0384
sub25d	0.760 769	-0.349 22	0.820 498	0.827 183
sub26	0.073 119	0.171 084	0.170 674	0.076 058
sub27	0.189 763	-0.314 86	-0.027 47	0.280 531
sub27d	0.659 194	0.652 39	0.123 129	0.732 024

and PERCLOS are often very close in values as well. In cumulative analysis, it was observed that ECM1 is more sensitive to the changes in eye closure data, which does not necessarily correlate with the ground truth used in this study since they are obtained by

subjective assessments with coarsely distributed time intervals. If the ground truth is obtained in finer time intervals, ECM1 may give better correlation results with the ground truth.