MULTI-SENSOR DRIVER MONITORING AND ASSISTANCE SYSTEM USING STATE-OF-THE-ART SIGNAL MODELLING

Pinar Boyraz
Memis Acar
David Kerr
Wolfson Mechanical & Manufacturing Engineering Department, Loughborough University, Ashby Road, LE11 3TULoughborough, Leicester, UK
Paper Number: 07-0165

ABSTRACT

Driver assistance and performance monitoring systems are currently being applied in modern cars in order to enhance safety. However, these systems have to answer certain concerns raised by manufacturers, legislators and users. These include, degree of intrusiveness (warning messages, tactile feedback, taking control of the car), ability to respond to different driving contexts and system reliability under varying road and environmental conditions and driver reliability. By combining inexpensive and non-intrusive sensors with state-of-the-art signal processing, probabilistic theory and artificial intelligence for signal analysis and modelling, it is possible to present a solution to all the above concerns to a certain extent. To investigate this extent, highway scenario simulator experiments have been conducted including 30 drivers in normal physical condition and impaired conditions due to lack of sleep. A simulator equipped with a near-infrared eye-gaze tracker, strain gauges to measure force on the steering wheel column (SWC), and potentiometers to measure steering wheel and throttle angle has been used. In addition to these core sensors, two webcams have been implemented to view the driver and to track lane-keeping. Raw data have been obtained comprising eye movement, force on SWC, vehicle speed, lane deviation, and human activity from the webcam. The data are first processed up to a level where all signals are one dimensional and continuous. Secondly, metrics have been derived using derivatives, histograms and entropies of the signals. These metrics are then tested against a ground truth risk level obtained from a driver survey and from independent observers. After selecting the best metrics for driver performance indication, different time windows for metric derivation are compared and the driver sessions are classified by a Fuzzy Inference System The system works well on the simulator data, with a 98% correct classification rate and is now being implemented in real conditions on real roads.

INTRODUCTION

Active safety depends on how well the vehicle is equipped for accident avoidance and prevention. However, a well-equipped car can still be involved in a severe accident if the driver of the car is not monitored. Detection of low performance of the driver due to fatigue, sleepiness and inattentiveness is crucial for active safety systems to operate on time considering the condition of the driver. Any solution to this problem could significantly in reduce the number of the accidents because thousands of car accidents are caused by low driver performance and condition [1] Therefore, experimental studies in search of indicator signals and studies to define the best way of using these signals to obtain a high correct classification rate and low number of false alarms are conducted. Eye tracker systems become centre of attention in computer vision domain. Different eye tracking systems together with head tracking algorithm are suggested based on near infra-red or visible light using different hardware architectures. Eye closure metric PERCLOS is identified as a good psychomotor indicator and validated against EEG [2]. In [3] the steering wheel angle is considered as an indicator signal and Artificial Neural Networks (ANN) are used as decision mechanisms. There are studies to use statistics, regression analysis [4] and fuzzy systems [5] for decision making using the indicators. In addition to mainstream approach alternative signal modelling approaches are also suggested such as system identification (SI). [6] Despite the vast amount of research on the issue, the questions including degree of intrusiveness, the cost and feasibility of the system, and the final output form have not been satisfactorily addressed. In order to answer these questions, the best indicator signals which can be measured non-intrusively using a low-cost sensor system are investigated. The first section defines the proposed multi-sensor system from this point of view. Next, derivation of best metrics representing signal characteristics and extraction of high-level information from raw signals is discussed. These metrics are grouped under different combinations in search for an optimal feature space. In some feature spaces, some of the metrics are not included on purpose to observe the effect of missing sensor data on performance of decision making system. The effect of time window size during which the feature vectors are calculated, on the prediction performance is investigated and optimum window size is determined. Finally, the decision systems are investigated and the results of training and testing of decision systems are reported.

Multi-Sensor System Structure

Detection of driver vigilance and quantitative measurement of states are difficult problems due to three system design requirements:

* *Non-intrusiveness*: The measurement systems must be non-intrusive to driver.
* *Robustness and Reliability*: System should be as reliable as possible to represent the real vigilance state and robust to compensate sensor failures.
**Low-cost and Feasibility:* The sensors selected for measurement system should be low-cost and should be connected in a feasible way.

Low cost, non-intrusive sensors and measurement methods that can be connected to CAN system of the cars are selected. Robustness and reliability are addressed under decision selection fusing the information from different information channels using a multi-sensor system.

The metrics derived from eye movements are the most reliable indicators .Therefore, in the core of the multi-sensor system, a near infrared computer vision system is placed for eye tracking. Because in certain cases (e.g. bright sun light, tinned glasses, driver out of field of view) eye tracking system cannot give reliable results or any results at all, it should be supported by peripheral sensor systems. This consists of strain gauges to measure the force applied by driver on the steering wheel and two encoders for measuring the steering wheel angle and the throttle angle. In addition to this system, there are two webcams viewing driver and the road for human movement analysis and for lane tracking performance measurement respectively. In brief, the peripheral system supports the computer vision system for eye tracking and also adds extra information about the attentiveness level of the driver in terms of vehicle dynamics (e.g. speed via throttle angle, steering wheel angle, lane deviation via webcam) and human-car interface related measures (e.g. force on steering wheel). Multi-sensor monitoring system arrangement and experiment geometry can be seen in Figure 1.

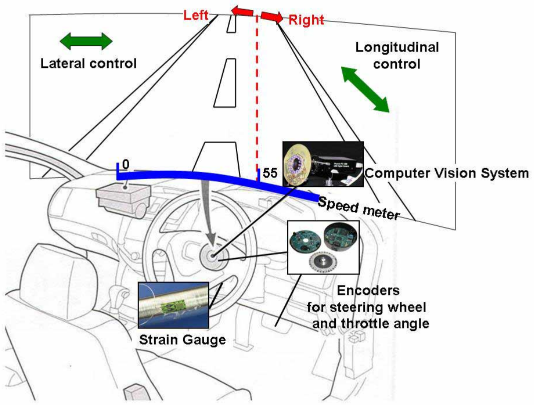


Figure 1. Multi-sensor system structure and experiment geometry

Experiment Design and Conditions

Thirty drivers with different level of driving skills and driving behaviour took place in the experiment and they drive the STISIM car simulator for about 1.5 hours. Each subject drove the same route twice under normal conditions and sleep-induced conditions. In the normal driving sessions drivers had their normal daily sleep need before taking part in the experiment, whereas in the 'sleep deprived' session they were requested to sleep at least 3 hours less than their usual sleep need. In order to induce sleepiness, this session took place between 2-4 pm in which the circadian rhythm of the body is known to decrease. Driving scenario is a monotonous highway scenario with no curvatures on the road, helping to induce sleepiness as well.
In order to separate the driving task into longitudinal and lateral control actions and to observe the distribution of the attention during the driving, drivers were given special instructions. Firstly, they were requested to keep their speed at 55 kmph during the session; therefore they needed to adjust their longitudinal control commands by changing throttle angle. Second instruction was to choose a lane and keep the lateral position of the car as stable as possible minimising lane deviation. In fact, these two requirements represent two rules that drivers should obey in a highway not to have any risk. The speedometer is arranged in front of the screen as a slide bar just underneath the road view. The drivers needed to look at different heights to check for the speed (speedometer) and for the lateral position (road view) changing their eye gaze. By this way the distribution of their attention during the experiment is expected to be measured from the gaze vector output of the computer vision unit.

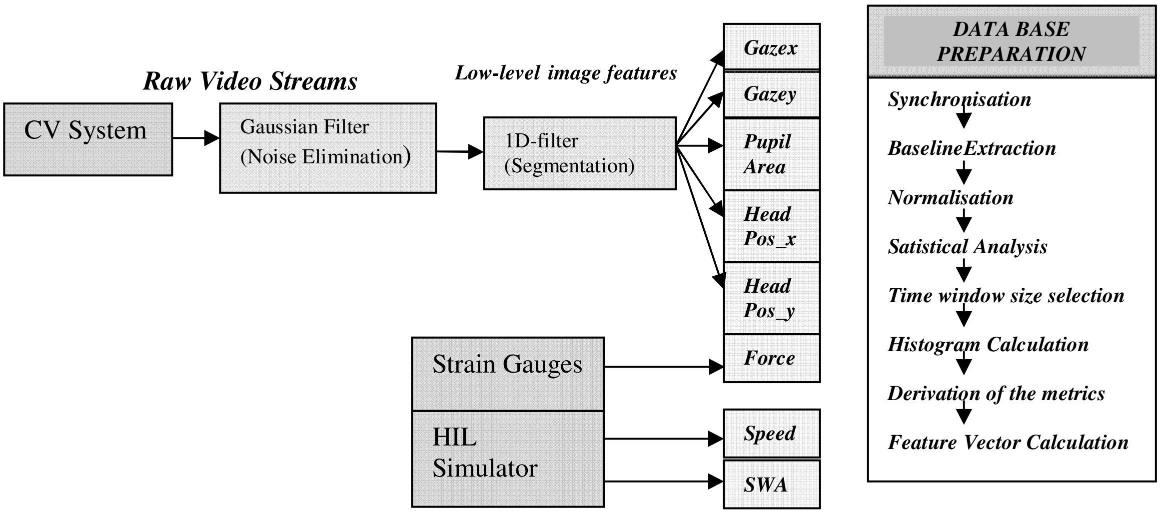


Figure 2. Data reduction and database preparation procedure

Data structure and analysis

The raw data obtained from the controlled experiment comprises the following list:

- Video stream containing near infrared frontal face images of the driver.
- Strain data from strain gauges on the steering wheel.
- Speed and steering wheel angle from the potentiomers in the simulator on steering wheel and on throttle respectively.

All the data should be reduced into one dimension first and then to some metrics/features characterising the signal. This data reduction procedure can be seen in Figure 2.

Metric Development- In order to derive the metrics all the information are reduced to one dimensional signals changing versus time and are synchronised. Development of good metrics depends on how well the nature of the signal is understood. Different metrics developed for this study are explained here briefly without giving details of low-level processing algorithms for the sake of brevity and focus.
****Visual Metrics:*** The first metric group is derived using the raw signals from computer vision system. The near infrared video stream of driver faces is processed to segment the eye image containing pupil and glint features. An example of segmented pupil area and glint can be seen in Figure 3. The vector between the centres of the pupil and glint can be mapped to the real coordinate of where the eye gaze is directed. The gaze is the direction of the eyes when the fovea is centred on the scene being seen, thus at the time when the frame is captured the attention is focused on that point. The one dimensional visual signals are constructed measuring pupil area, x and y component of the gaze vectors and x and y components of the pupil centre to measure the head coordinates.

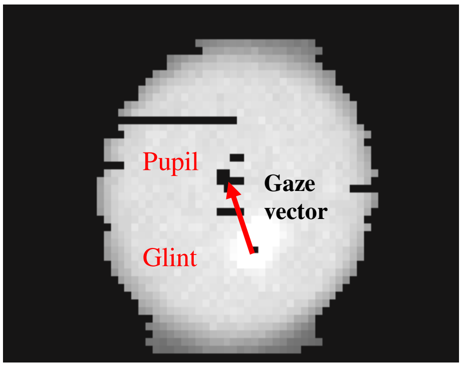


Figure 3. Segmented pupil area showing the pupil area, glint (corneal reflection) and defined gaze vector

Samples of gaze vector y component, pupil area and head x and y component measurements are given in Figure 4. In order to derive the metrics from one dimensional signal, an exploratory analysis is conducted. In this analysis, standard deviations, mean values and entropies of the signals are taken. In addition to these three variables, the histograms of the signals are taken over a predefined time window in order to follow signal value distribution over time. Each driving session is divided into 12-minute long sub-sections taken from start, middle and finish parts of the session. By constructing histograms of gaze x, gaze y and pupil area some characteristics that are not visible from signal-time diagrams are obtained. Three visual metrics are defined based on histograms: Eye closure metric 1(ECM1), Eye closure metric 2 (ECM2) and Attention division ratio (ADR).

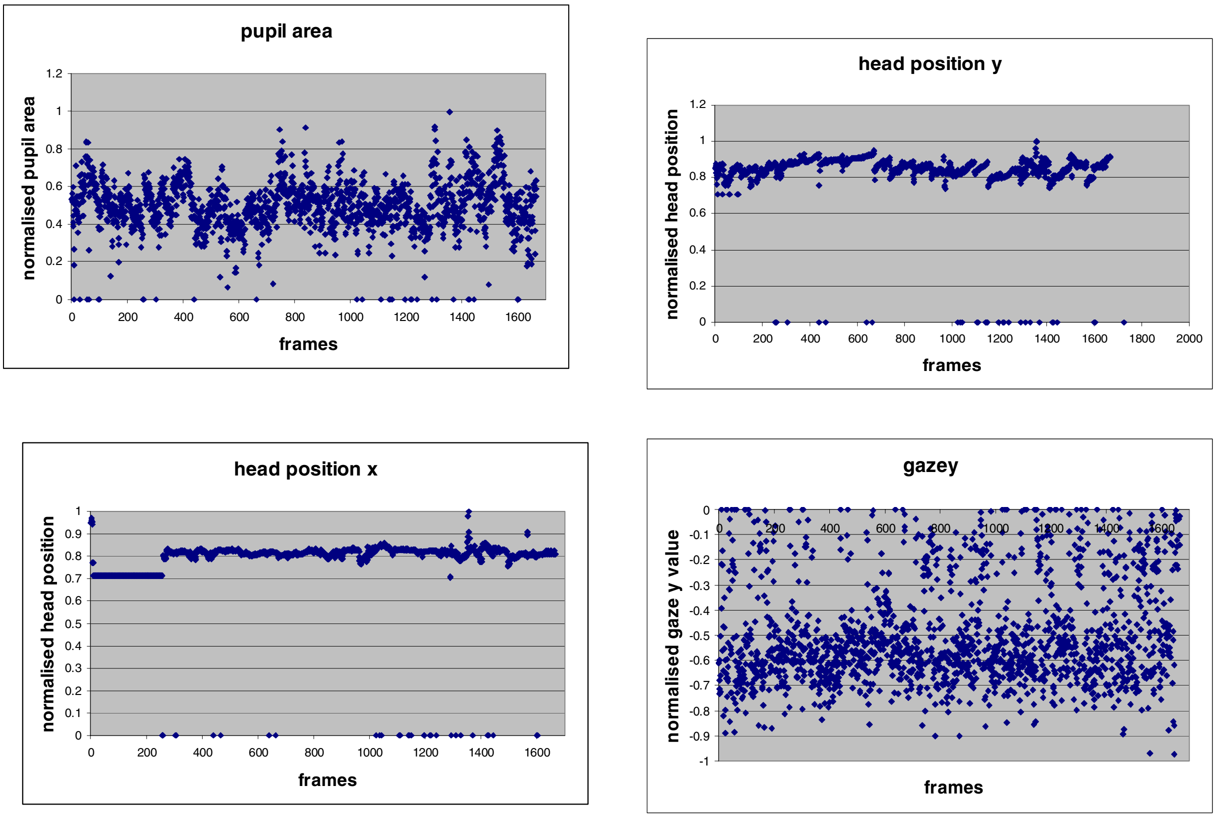


Figure 4. Samples of pupil area head position x and y and gazey

The histograms of gaze x, gaze y and pupil area are shown in Figure 5 for a normal/alert and impaired/drowsy driver. There are two clear observations that can be drawn from histograms:

1. Gaze y values between [-1 and -0.5] represents road scene changing eye gaze, thus the attention is on the road scene. For the values between [-0.5 and 0] the attention is focused on speedometer. Gaze y histogram has an equal distribution of attention to these two defined intervals at the start of the simulation for both alert and drowsy drivers. However, as the time proceeds the distribution of gaze y concentrates in speed checking region. Both alert and drowsy driver follows the same trend, however, the distribution change towards the speed checking interval is more rapid and dramatic in drowsy driver.
2. The number of closed eyes [0] and open eyes [1] and intermediate states if any can be seen in pupil area column of the histogram. As the time proceeds the proportion of number of closed eyes to the number of open eyes increases. Both alert and drowsy drivers follow the same trend however the proportion increases dramatically in drowsy state.

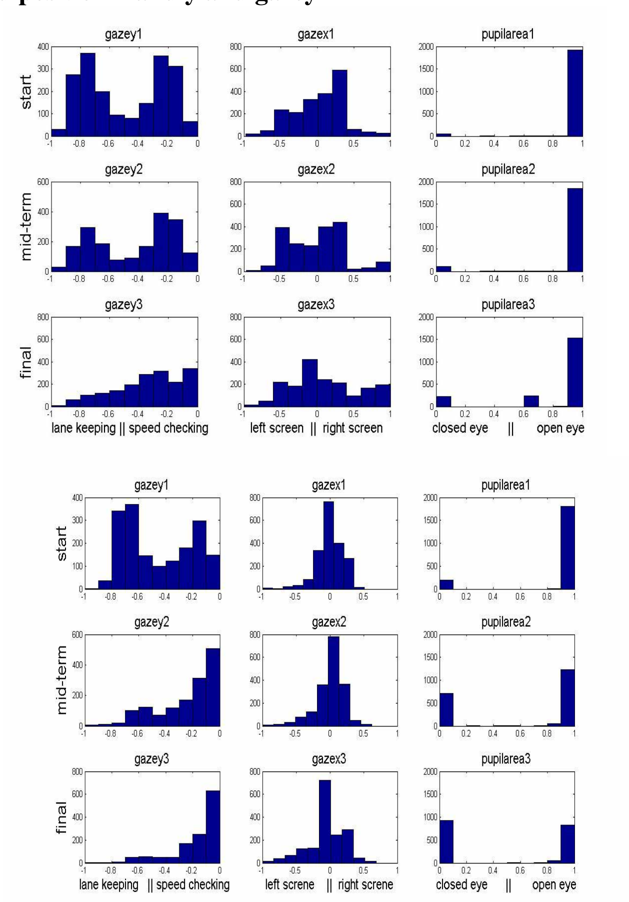


Figure 5. Histograms of alert (top) and drowsy (bottom) drivers

In the light of these observations ECM1 and ECM2 are derived from pupil area histograms of three sub-sessions of the whole driving session. These eye closure metrics are similar to PERCLOS measured by [7], however defined over a longer time period. The attention division ratio (ADR) is derived from the gaze y column of the histograms by dividing the number of gaze measurements in speed checking interval to the number of gaze measurements in road checking interval.(1)

Eye Closure Metric (ECM1) emphasizes the number of blinks by taking the proportion of fully closed to fully open eye cases, whilst ignoring the cases in between. The second metric of eye closure (ECM2) represents partially closed cases as well as fully closed eye cases in the denominator of the ratio. (2 and 3)

$$ADR = \frac{N_{gazey\ >-0.5}}{N_{gazey\ <-0.5}} \tag{1}$$

$$ECM1 = \frac{\sum_{1}^{n} N_{pupilclused}(i)}{\sum_{1}^{N} N_{pupilwideopen}(i)} \tag{2}$$

$$ECM2 = \frac{> PupilArea*0.9}{< PupilArea*0.9} \tag{3}$$

In addition to these newly defined metrics, the entropies and standard deviations of the signals are also used. The entropy of the signals is calculated as in information theory given in (4).

$$entropy = \sum_{i-1}^{n} x_i \log x_i \tag{4}$$

To sum up, the visual metrics are the standard deviation, entropy and mean value of gaze x, standard deviation and entropies of head motion x and y components, ECM1, ECM2 and ADR.

****Non-visual Metrics:*** These are the metrics obtained from vehicle dynamics (speed and steering wheel angle) and human-car interface (force on steering wheel) signals. A sample of these signals can be seen in Figure 6.

As in the visual metric development the standard deviation and entropy is used to measure the scatter and complexity in the signal. In addition to this general approach, two control metrics are derived from the speed indicating longitudinal control performance. As can be noticed from Figure 6, the speed graph resembles the response of a PID controller. In fact, drivers are told to keep their speed at 55 kmph; therefore, driver acts like a PID controller to keep this reference value with some small deviations. These small deviations can be added up to give total steady state error of the driver after settlement. Integral of the standard error and integral of the average error is taken as performance indicators of longitudinal control as given by (5 and 6).

$$ISE = \int_{0}^{T} e^{2}(t)dt \tag{5}$$

$$IAE = \int_{0}^{T} |\, e(t) \,|\, dt \tag{6}$$

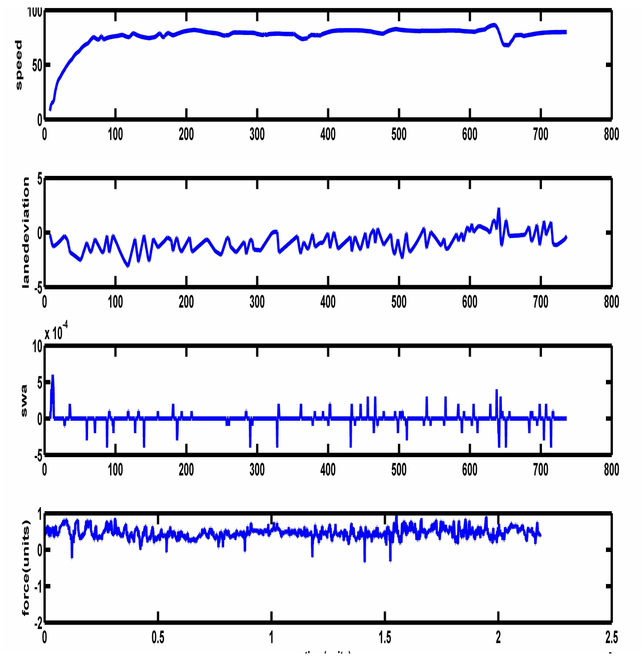


Figure 6. Sample data from non-visual information channel: speed (km/h), lane deviation (m), steering wheel angle (degree) and force (normalized to [-1, +1] N interval), time (simulator time unit is seconds)

Feature Spaces and Feature Selection

In order to find the best feature space to represent the signals, three different feature space has been constructed. The first feature space (F1) contains only the visual metrics, the second (F2) contains entropy and control values and finally the third feature space (F3) has visual metrics and standard deviations. The first feature space is constructed to observe how well the visual metrics can predict the drowsiness level of the driver. The second feature space leaves out the visual metrics so that it becomes observable how well the control and entropy related metrics can predict without using visual cues. Finally, the third space represents a visual metric space backed up by standard values of other metrics. The member of these three spaces can be seen in Table 1.

Apart from these three feature spaces a 'best feature space' is constructed after considering the results from correlation analysis. The ground truth risk level obtained from independent assessment and

Table 1. Different feature vectors are designed to investigate the best representation of the phenomena

Visual Cues (F1)	**Entropy and Control (F2)**	**Visual Cues and Standard Deviations(F3)**
Eye Closure Metric 1 Eye Closure Metric 2 Attention Division Ratio Gaze x Mean Gaze x Standard Deviation Head Motion in x Standard Deviation Head Motion in y Standard Deviation	IAE ISE SWA Entropy Gaze x Entropy Force on SWC Entropy Head Motion x Entropy Head Motion y Entropy	SWA STD Force on SWC STD Eye Closure Metric 1 Eye Closure Metric 2 Attention Division Ratio Gaze x Mean Gazex Standard Deviation Head Motion Standard Deviation Head Motion y Standard Deviation

Table 2. Best feature space members selected by p<0.05 criterion

FV Members	***P (signf.)***
ECM1	0.000
ECM2	0.000
AttDivRatio	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

surveys and metrics are taken into correlation analysis as in (*metric, ground truth drowsiness level)*pair for each metric separately . The metrics having high correlation coefficient ,r, and small values of signifance ,p, with ground truth drowsiness level as a result are selected as best metrics to construct best feature space. Best feature space members are given in Table 2 together with their p values. The members are selected by taking their p and r values into account. The metrics having $p<0.05$ and $r>0.2$ is considered as having high enough correlation to actual drowsiness level expressed by ground truth.

Decision Systems and Results

After constructing the feature spaces from calculated metrics, a decision system should be trained to detect impaired and normal states of the drivers. In order to train the decision systems, database containing the feature spaces are divided into train and test data groups. Supervised learning method is used and ground truth acted as teacher to shape the artificial intelligence system to produce rules using available metrics. Because the drowsiness level can be judged by people and the available information is fuzzy nature, the Fuzzy Inference system is chosen as a good candidate to mimic this decision making process. Thus the system is expected to behave like a co-pilot detecting the impaired driver and the level of the risk involved. Finally, the effect of the time window selection in calculating the metrics are analysed using different time windows and training separate FIS for each time window selection.

Fuzzy Inference Systems- FIS can be of mainly two types: Mamdani [8] and Sugeno-Takagi [9]. The first System constructs the rule base of the system from expert knowledge and is transparent to the designer of the system. Any rule can be added to or removed from a Mamdani FIS. On the other hand, Sugeno-Takagi system uses data to extract the rules in terms of linear relationships between the inputs to yield the output, thus it is data driven. If how the measured metrics were connected to the drowsiness level was known, the choice should be a Mamdani system. However, in our problem the rules expressing the relationship between the metrics and level of impairment and involved risk is not clear. Therefore, this investigation asks a two way question to find the best feature spaces and best decision system. For this reason, Sugeno-Takagi system is used to reveal the relationships mathematically to construct a rule base. A sub-clustering method is used in deriving the Sugeno-Takagi (S-T) based FIS.

The fuzzy C-means algorithm is an iterative optimization algorithm minimizing the cost function in (5).

$$J = \sum_{k=1}^{n} \sum_{i=1}^{c} \mu^{m}{}_{ik} \parallel x_k - v_i \parallel^2 \qquad (5)$$

where degree of membership μ_{ik} is given by (6).

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\parallel x_k - v_i \parallel}{\parallel x_k - v_j \parallel} \right)^{2/(m-1)}} \qquad (6)$$

where

n : number of data points, c : number of clusters

x_k : kth data point

v_i : ith cluster center

μ_{ik}: degree of membership of kth data in ith cluster,

m : constant

When the input precisely matches with 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 mapping from the input space to output space is then performed using the rule base extracted by this method.

A Sugeno-Tkagi (S-T) FIS system is trained using 150 sessions of database collected from 30 subjects and tested using 18 sessions that are completely new to the trained system. The result of the S-T FIS trained for F1 feature space is given in Figure 7showing that the decision system was able to predict most of the cases.

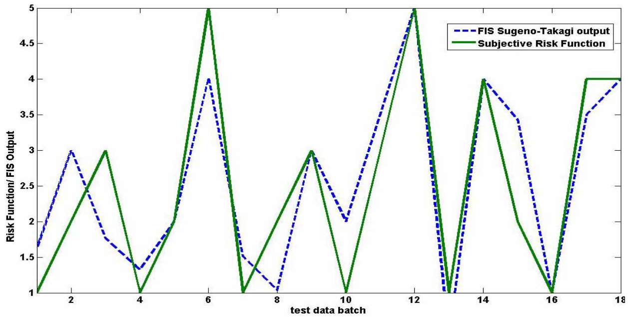


Figure 7. Output of S-T FIS vs Risk function over F1 space

The performances of F2 and F3 feature spaces are given in Figure 8 and 9 respectively. F2 feature space containing the control and entropy metrics and lacking the visual metrics is not successful in predicting the test data precisely.

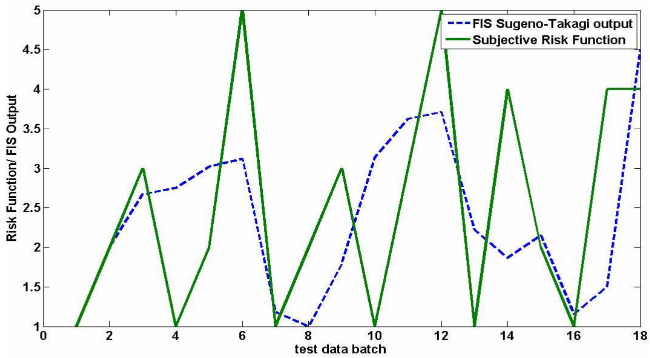


Figure 8. Output of FIS-S-T system vs Risk function over F2 space

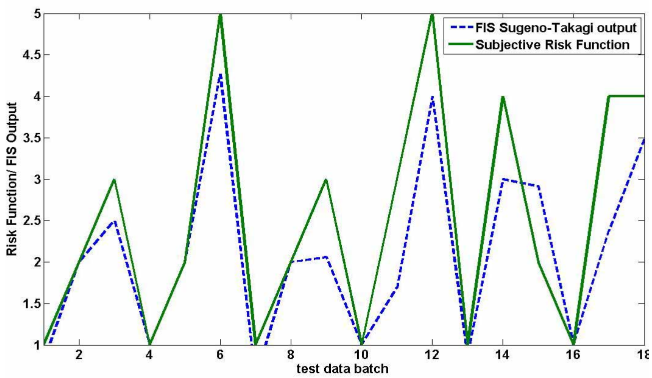


Figure 9. Output of FIS-S-T system vs Risk function over F3 space

Adding standard deviation values of non-visual metrics to visual metrics caused a drop in performance when compared to F1. However, F3 is still advantageous to F1 because it does not solely depend on the visual metric. When the visual metrics become completely unavailable F3 can still judge the drowsiness level based on non-visual metrics it contains. On the other hand F2 feature space shows that it is not enough to include only control and entropy related metrics, the visual metrics are of crucial role in the system.

The performances of these three feature spaces are summarised in Table 3.

PERFORMANCE COMPARISON	*F1*		*F2*		*F3*	
	Success (%)	**False Alarm (%)**	**Success (%)**	**False Alarm (%)**	**Success (%)**	**False Alarm (%)**
Mamdani	90	10	80	20	85	15
Sugeno-Takagi	98	none	90	10	95	5

Table 3. Performances of three feature spaces using S-T FIS for decision making

After observing the feature space performances from F1, F2 and F3, it is decided that the best feature space,F4, should be put under test to obtain an optimum performance form the available metrics. This feature spaces includes only the metrics having high correlations with ground truth as explained before. The test results from F4 space is shown in Figure 10.

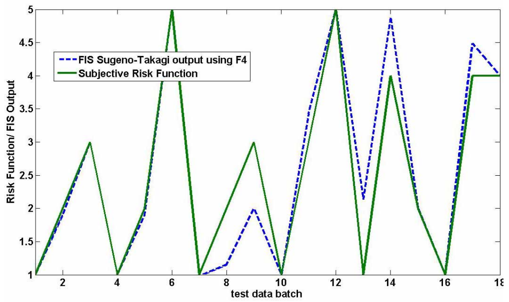


Figure 10. Test results of S-T FIS using F4 space

Feature space F4 was able to correctly identify 98% of the sessions; therefore, proving to be the best feature space. For the rest of the analysis including time windows, therefore, F4 will be used and tested.

The next step in our investigation is to analyse the effect of the time window on the performance of the decision system. For this reason, the time window is successively halved to obtain 6 min, 3 min and 1.5 min intervals corresponding to 1000, 500 and 250 frames of the video segments of the NIR computer vision system. The results of 6-min time window is given in Figure 11.

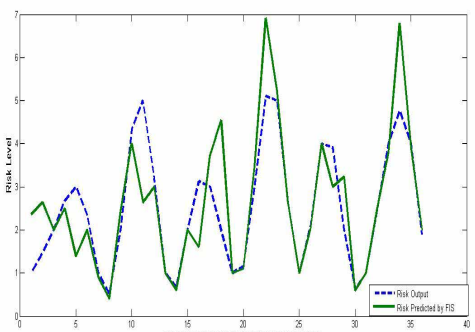
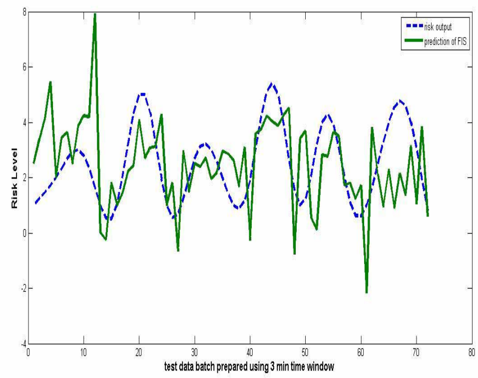


Figure 11. Output from S-T FIS using 6-min time window in calculating feature space



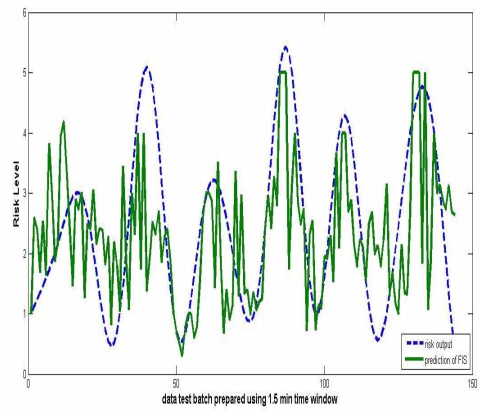


Figure 12. Output from S-T FIS using 3 and 1.5 min time window in calculating feature space

As the time window for the calculation of the feature vectors narrows down the system is able to track the general trend efficiently however after 3 min it begins to fluctuate. The fluctuation occurs because of the quantization and re-sampling between the data points. However, the system is able to give reasonable response for each time window. Preferably the 3 min time window is a good compromise between fast response and generalization capability or tracking capability of the general trend. These observations can be tracked from Figure 12.

CONCLUSIONS

In this study, a multi-sensor driver vigilance monitoring system is designed using the-state-of-the-art signal modeling techniques and Fuzzy Inference Systems. This study investigates three important aspects of monitoring system design problem: reliability, availability and robustness. In order to find an answer these three requirements, best feature space, the time window used in calculations and an optimum decision system are sought after. As can be seen from the results, the visual channels of information are proven to be the most powerful signals to detect the drowsiness and associated risk. In addition to conventional eye closure metrics, a new metric developed to measure the distribution of the attention of the driver. It has been found highly correlated to the involved risk level due to drowsiness. The best feature space is defined according to correlations with the perceived risk level. Finally, it is concluded that a Fuzzy Inference System using a feature space containing visual metrics of eye closure, attention distribution, head movement and non visual metrics of entropies of the steering wheel angle and the force on the steering wheel supported by control performances (IAE and ISE) of the longitudinal speed is the best. The best time window for calculating the metrics is identified as 3 minutes.

REFERENCES

[1] Drowsy Driving and Automobile Car Crashes, NCSDR/NHTSA Expert Panel on Driver Fatigue and Sleepiness, http://www.nhtsa.dot.gov/people/injury/drowsy_driving1/drowsy.html#EXECUTIVE%20SUMMARY Last visited: 4.03.2007

[2] R. Rau and Knipling P., "A Valid Psychophysiological Measure of Alertness as Assessed by Psychomotor Vigilance", http://www.fmcsa.dot.gov/documents/tb98-006.pdf, 2007

[3] Sayed, R. and Eskandarian, A., "Unobtrusive drowsiness detection by neural network learning of driver steering", *Proceedings of Institution of Mechanical Engineers*, **Vol. 215, Part D**, pp. 969-975, 2001

[4] Tijerina, L., Gleckler, M., Stoltzfus, D., Johnston, S., Goodman, M.J., Wierwille, W., W., "A Preliminary Assessment of Algorithms for Drowsy and Inattentive Driver Detection on the Road", NHTSA, DOT HS, 808, March 1999

[5] Onken, R. and Feraric, J.P., 'Adaptation to the driver as a part of a driver monitoring and warning system', Accident Analysis and Prevention, **vol. 29, no: 4,** pp.507-513, 1997

[6] T. Pilutti and A. G. Ulsoy, "Identification of driver state for lane-keeping tasks," *IEEE Transactions on Syst Man Cybern. Pt. A Syst. Humans*.**Vol.29** ,pp. 786 , 1999

[7] Dinges, D., F., Mallis, M.,M., Maislin, G., Powell, J., W., 'Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and the Basis for Alertness Management', NHTSA DOT HS 808 762, April 1998

[8] Mamdani, E.H. and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, Vol. 7, No. 1, pp. 1-13, 1975.

[9] Sugeno, M., *Industrial applications of fuzzy control*, Elsevier Science Pub. Co., 1985.