Abstract

Driver distraction is recognized as a major factor in the cause of automobile accidents. Therefore, it is extremely important for an intelligent driver support system to be able to monitor the driver’s attentive state. This paper proposes a system to monitor driver attention based on a variety of information sources. The LISA-Q test vehicle is used to synchronously capture video, audio, vehicle information, LASER RADAR information, and GPS information as input to the driver state evaluation. Information about the driver’s facial affects, lane keeping, steering movements, and time headway are all extracted from the multimodal data streams and evaluated.

1 Introduction

Automatic systems for drivers state monitoring have garnered a large amount of attention due to the large amount of fatalities every year resulting from driver fatigue and driver inattention. According to a study by the NHTSA, driver inattention is the leading contributing factor to crashes based on both the 1997-2000 Crashworthiness Data System and the 2000 General Estimates System [1]. A complete system for driver attention analysis is therefore a key component for intelligent driver support systems. Such a system must be able to acquire many different modalities of data and analyze them to generate a robust estimate of the driver’s state.

Many different data sources have been looked at for researching driver state. Steering corrections have been shown to provide cues for driver fatigue and inattention [2]. Fatigued and inattentive drivers perform fewer corrections with larger magnitudes than vigilant drivers. In addition to this, other metrics such as lane keeping, time headway maintenance, and velocity maintenance are also important metrics for driver state [3].

Vision systems for driver fatigue monitoring have been developed that examine eye gaze, blink rate, head pose and movements [4], [5], [6]. These techniques have shown to be effective, but neglect important information which can be obtained though other sources.

Hidden Markov Models [7] and Bayesian Networks [4] have provided frameworks for estimating the driver’s state and behavior based on one or more of these obtain cues. The Bayesian framework allows for the inclusion of a priori knowledge of the driver’s state such as the amount of sleep or physical activity the driver has had.

This paper outlines the development of a system for combining multi-sensory cues to create a robust estimate of driver attention. Section two describes the architecture of the system. Section three describes the data processing involved in obtaining cues for driver attention from raw data. In sections four and five, results and conclusions from our findings are presented.

2 Driver Attention Monitoring System

The driver attention monitoring system described in this paper is based on the Laboratory for Intelligent and Safe Automobiles Infiniti Q45 Test Bed (LISA-Q) [8]. Data is acquired using this test bed which is outline in greater detail in section 4. This data is then processed to generate four main cues for driver state analysis. These four cues are driver affects, vehicle lane position, steering wheel and pedal movement, and time headway analysis. Based on these cues we can classify the driver state into attentive and inattentive states. A flow chart for this system is shown in Figure 1.

3 Driver State Cue Extraction

A critical step in the driver state monitoring system is to accurately and reliably extract cues that are informative of the driver’s attentive state. To this end, it is important to rely not only on one single cue, but to
merge the information gathered from multiple cues. Our system currently uses the cues from driver affect analysis, lane position, steering wheel and pedal movements, and time headway.

3.1 Driver Affect Analysis and Cue Extraction

Driver affect analysis incorporates the driver's facial expressions, head pose, eye movements, and eye blinks to generate reliable cues for the driver's state. The current system is implemented with a real-time facial expression analysis system based on thin-plate splines [9] for expression feature vector extraction. Future work will include adding head pose, eye movements, and eye blinks into the driver monitoring system.

3.1.1 Facial Expression Analysis

Using thin-plate splines for facial expression analysis allows for the easy separation of non-rigid motion from facial affects from the rigid motion from head movements. Furthermore, this decomposition requires only two matrix multiplications and one 2x2 matrix inversion making the method well suited for real-time applications.

The analysis begins with the initialization of specific facial landmarks that include the eyes, eyebrows, nostrils, and points on the mouth. These landmarks are then tracked between frames. By minimizing the energy function (1), the warping between the neutral model face and the current face (2)-(3) can be calculated using equations (4)-(7) [10]. Finally, the nonlinear warping parameters can be transformed using equation (8) to remove the dependence on the affine parameters which can be attributed to the rigid body motion.

\[
I_f = \arg\min_f \int \left( \left( \frac{\partial^2 f}{\partial x^2} \right)^2 + \left( \frac{\partial^2 f}{\partial x \partial y} \right)^2 \right) dx dy + \left( \frac{\partial^2 f}{\partial y^2} \right)^2 \]

\[
f = t_i + a_{x}x + a_{y}y + \sum_{j=1}^{n} w_i \left[ \begin{array}{c} x_{m,j} \\ y_{m,j} \end{array} \right] - \left[ \begin{array}{c} x \\ y \end{array} \right]
\]

\[
Z = -U = -r^2 \cdot \log(r^2)
\]

\[
r^2 = x^2 + y^2
\]

\[
W = T = A = L^{-1}Y
\]

\[
L = \begin{bmatrix} K & P \\ P^T & 0 \end{bmatrix}
\]

\[
P = \begin{bmatrix} 1 & x_{m,1} & y_{m,1} \\ 1 & x_{m,2} & y_{m,2} \\ M & M & M \\ 1 & x_{m,n} & y_{m,n} \end{bmatrix}
\]

Graphically this warping can be seen in Figure 2 which shows the warping for 6 facial expressions.

![Figure 2 This plate spline warping from facial expressions](image)

3.2 Lane Position Analysis

The driver's ability to maintain focus on the lane keeping task has been shown to be related to attention and fatigue [3]. As well as providing a metric for the driver's state, lane position detection can determine how critical the situation is if the driver is determined to be inattentive. An urgent alert may be signaled if the vehicle is crossing into an adjacent lane or leaving the road when the driver is not being attentive to the driving task or fatigued.

It is therefore important to have a robust lane tracking system as part of the driver state monitoring system. The more accurate the lane detection, the less likely false alarms will occur which may degrade the driver's responsiveness when critical situations arise.
The lane tracking system used in this system is based on a unified approach to detection and tracking \[11\]. A flow chart for this system is shown in Figure X. Steerable filters \[12\] are used to selectively look for circular reflector lane markings (Bots Dots) and lines oriented near specific directions. Feedback from the lane tracker is used to more robustly detect lane markings by determining the expected orientation of the lane lines.

\[G^{\theta}(x,y) = G_{xx} \cos(\theta)^2 + G_{yy} \sin(\theta)^2 - 2G_{xy} \cos(\theta) \sin(\theta)\]

Using this fact we can calculate the minimum filter response using equations (13)-(15).

\[\theta_{\text{min}} = \tan^{-1}\left(\frac{G_{xx} - G_{yy} - A}{2G_{xy}}\right)\]

\[\theta_{\text{max}} = \tan^{-1}\left(\frac{G_{xx} - G_{yy} + A}{2G_{xy}}\right)\]

where

\[A = \sqrt{G_{xx}^2 - 2G_{xx} G_{yy} + G_{yy}^2 + 4G_{xy}^2}\]

Solid line lane markings can be detected by computing the filter response at a given angle from equation (12).

Examples of the detect lane markings can be seen in the figures above.

The lane position is then tracked using a discrete time Kalman filter. The state of the Kalman filter includes the lane position, the first and second derivatives of the lane position, and lane width. Output from the lane tracking system can be seen in Figure 12.
3.3 Steering Wheel and Pedal Movement Analysis

Another important cue for driver state monitoring is the frequency and magnitude of corrections to steering and speed. It has been shown that for inattentive drivers the frequency of corrections decreases while the magnitude of the corrective actions increases [2]. By capturing the steering wheel and pedal information via the can bus, we can use this information to analyze the amount and magnitude of corrections. Figure 13 shows a plot generated during the data analysis of the steering angle. It shows small high frequency corrections typical of an alert driver.

3.4 Time Headway Analysis

Similar to lane position, time headway is a good indicator for driver attentiveness [3]. Also, falling below a particular time headway might indicate that a critical situation is about to occur. This information is captured via the LASER RADAR system built into the LISA-Q capture system. The following distance divided by the current speed of the vehicle obtained from the CAN bus. This provides a reliable estimate of time headway.
Experimental Evaluation from Highway Driving

At the backbone of this system is the ability to acquire multiple full frame video streams, audio, CAN bus and LASER RADAR data. This is all accomplished using the LISA-Q intelligent vehicle test bed shown in Figure 16.

The system captures and synchronizes data from video cameras, a global positioning system, and the vehicle's internal computer. Video cameras monitor the driver’s face, the driver’s feet, and the vehicle's surround. Data from the internal computer includes steering position, pedal positions, LASER RADAR lead vehicle distance measurements, vehicle speed, and other vehicle dynamics variables. By allowing for the synchronized capture of multiple modalities of data streams, the system is capable of extracting a rich set of cues for estimating driver state.

Concluding Remarks

In this paper we have shown a framework for a system for driver attention monitoring. Its main emphasis is not centering on just one or two cues to determine driver attention, but rather to explore a wide variety of sensors to obtain information about the driver and his surroundings. This system involves capturing and synchronizing data from many different types of sensors and generating relevant metrics from this information. By combine a wide variety of different sensors and metrics, a more robust estimate of the driver state is created and more aspects of driver behavior and attention can be studied.

After all of the metrics have been obtained, it is important to be able to combine them together to form a reliable estimate of driver attention. There are many different methods for doing this. Bayesian networks...
have been shown to be effective and bring in the ability to combine prior knowledge of the driver to help make more accurate estimates. Data driven classification such as support vector machines or boosting can also provide good classification results.

This system can also be expanded to include other sensors and information. An example of this might be to use the driver head pose to determine the driver’s view [13]. From the driver’s view it could be possible to obtain the focus of his attention.

Acknowledgements

The authors would like to thank UC Discovery Grants Digital Media Innovation program, Nissan Motor Co. LTD., Japan, their colleagues at the CVRR. A special thanks to Ofer Achler for his effort in creating the LISA-Q test bed.

References