This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Portable and Scalable Vision-Based Vehicular Instrumentation for the Analysis of Driver Intentionality

Steven S. Beauchemin, Michael A. Bauer, *Member, IEEE*, Taha Kowsari, and Ji Cho

***Abstract*—Probably the most promising breakthroughs in vehicular safety will emerge from intelligent, Advanced Driving Assistance Systems (i-ADAS). Influential research institutions and large vehicle manufacturers work in lockstep to create advanced, on-board safety systems by means of integrating the functionality of existing systems and developing innovative sensing technologies. In this contribution, we describe a portable and scalable vehicular instrumentation designed for on-road experimentation and hypothesis verification in the context of designing i-ADAS prototypes.**

***Index Terms*—Behavioral science, cognition, instrumentation and measurement, optical sensors, vehicle driving.**

I. Introduction

WORLDWIDE deaths from injuries are projected to rise from 5.1 million in 1990 to 8.4 million in 2020, with traffic-related injuries representing the major cause for this increase [1], [2]. Our research aims at reducing these fatalities by first developing a deeper understanding of the cognitive (cephalo-ocular) task of driving, identifying related risk factors and integrating these findings into predictive models of driver intentionality. The long-term goals of this program include the identification of the cognitive factors involved in driving that impact traffic safety, the definition of sound principles for the design of automated vehicular safety technologies, and the development of intelligent, Advanced Driving Assistance Systems (i-ADAS), with driver behavior prediction and correction as the central tenet of safety improvement.

While research on ADAS may integrate a number of different functions such as forward collision detection and lane departure tracking [3], little attention is devoted to the monitoring of events and factors that directly concern the driver of the vehicle. It is only recently that cognitive aspects have been considered as a legitimate part of i-ADAS [4]. Since 95% of all accidents are caused by human error, it is crucial that these aspects of driving be a central part of i-ADAS [5]. Keeping the driver as an active participant in the feedback mechanisms allows for providing contextually motivated informational support and offers immediate applications for enhancing safety [6].

The extended possibilities of integrated, i-ADAS are very relevant research areas as they do not intend to replace the driver as much as to assist in the process of driving safely. As it has been pointed out by Petersson *et al.* [6], what remains to be automated to reach the state by which vehicles become completely autonomous in a practical manner turns out to be difficult and elusive in everyday driving situations. In light of this, it is our belief that driver support through i-ADAS can be deployed more readily, with consequent socioeconomic benefits.

This contribution rests on earlier work in which preliminary instrumentation and tests were recently conducted [7]. It however differs significantly in that it motivates the instrumentation in the form of clearly stated hypotheses derived from a central conjecture and provides a performance evaluation of the platform, along with identified physical, sensory, and computational limitations.

Our approach, while sharing common elements with those of others, is unique in several ways. First, we designed a portable instrumentation requiring no modification to the vehicular platform, using low-cost off-the-shelf components that are widely available. Second, our on-board computational approach rests on scalability. That is to say, additional computing power can easily be added to the current instrumentation, without any modifications to the existing system. This of course is a core requirement, as algorithms must be run in real time. Third, our approach integrates the driver in the system as an inherent behavioral agent, in the aim of understanding and predicting driving actions.

II. Related Literature

While many research groups provide brief descriptions of their vehicular instrumentation in the context of driving assistance, such as [8], [9] for vision systems, and [10], [11] for multisensor instrumentation, few contributions directly address instrumentation strategies, concepts, and implementation in the context of ADAS. A notable exception is by Thrun [12] in the context of autonomous driving in which the sensory interface, perception modalities, planning and control, user interfaces, and software services are described in extensive detail. The motivation for our contribution partly stems from the observation that the related literature is currently sparse.

Manuscript received March 14, 2011; revised July 6, 2011; accepted July 10, 2011. The Associate Editor coordinating the review process for this paper was Dr. Jesús Ureña.

The authors are with the Department of Computer Science, The University of Western Ontario, London, ON N6A-5B7, Canada.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2011.2164854

III. Hypothesis Verification

Our primary goal is to determine whether driver intentionality and driving-related actions can be predicted from quantitative and qualitative analyses of driver behavior. This broad question, while only partially answered [13], conveys its importance in more than one way. For instance, predictive formulations of the cognitive aspects of driving open the way to the design of reliable models of driver intentionality prediction and may lead to advances in safety-related vehicular technologies and accident prevention strategies from the perspective of on-board safety systems, up to assisting in guiding policies regarding the regulation of i-ADAS.

A. Primary Conjecture

Studies of driver behavior have approached the problem of intentionality from various perspectives. Driver behavior models have been suggested along with their empirical validations, with varying degrees of success [14]. However, *"the most effective technology may be that which monitors driver state and driving behavior to help attend to the roadway and recognize unsafe behavior"*[15]. In addition, it has been demonstrated time and time again that eyes, in general, *"look directly at the objects they engage with"*[16] and that the *"fixation that provides the information for a particular action immediately precedes that action"*[17]. These observations support our primary conjecture, which states that *if one considers a vehicle as an extension to the inherent human capability for motion, then one must also admit the possibility that eye movements are as predictive of driving actions as they are of physical movement*. The underlying rationale from which our conjecture stems rests on the demonstration that eye movements reflect processes aimed at locating the information required to generate actions in relation to the environment [13], [18], [19].

B. Hypotheses

Since 95% of all accidents are caused by human error, it is imperative that drivers be the central element of systems that provide driving support [5]. Consequently, our short-term goals consist of the empirical testing of hypotheses derived from the primary conjecture, in the hope of demonstrating that on-board vehicle safety systems which focus on the predictability of driver behavior are capable of significantly increasing driving safety. Toward this end, our primary conjecture is functionally fragmented into a number of hypotheses which can be investigated effectively and objectively:

1) *Cephalo-ocular behavior correlates with driver intentionality and precedes driving actions:* This hypothesis has been demonstrated in certain driving circumstances, as it is known that drivers negotiating a bend fixate on its tangent point to gather information on its curvature. This fixation precedes steering adjustments by an average of 0.8 s [19]. Are there other driving circumstances (negotiating intersections, merging, highway driving, etc.) for which particular ocular behavior precedes driving actions? While it is clear that ocular behavior cannot be constantly predictive of actions due to secondary drivers tasks (such as attending to vehicle functions), it is important to determine which behaviors possess a predictive value. However, the possibility exists that ocular behavior may not be sufficiently correlated with intentionality for use in prediction models. In this case, other investigative avenues may be possible, particularly through the observation of maneuvers being applied to the vehicle by the driver. Current actions may be predictive of future actions and determining to which extent this may be the case would be central to this scenario. A third investigative avenue may be that both current driver maneuvers and ocular behavior are sufficient for useful prediction purposes.
2) *Driver levels of attention are indicative of the meaningfulness of cephalo-ocular behavior:* Driver visual attention is a central part of safe driving. It has been shown that driver glances away from the road for 2 s or more resulted in 3.6 times more involuntary lane departures than glances of 1 s [20]. Conversely, long visual fixations are not necessarily synonymous with attention. While eyes may be fixating, attention may not be elicited by events in the visual field. However, certain ocular patterns such as fixations accompanied by regular saccades are descriptive of the visual search behavior for information acquisition processes and correlate with drivers attending to the roadway [21]. The identification of factors providing indications of meaningful cephalo-ocular movements is necessary to assess whether the ocular behavior represents intent.
3) *Information delivered to drivers does not increase their cognitive loads:* Drivers are exposed to increasing flows of information provided by modern on-board vehicle functions. Recent studies have revealed that drivers are not always capable of eliciting a correct response to such solicitations due to, among other factors, the complexity of the driving context, or an increased cognitive load generated by actions not directly related to driving [22]. While it is suspected that the aforementioned hypothesis does not hold in general, it is crucial to experimentally determine the modalities, timings, and types of delivered information that can be tolerated and understood sufficiently rapidly by drivers, such that there is available time to perform corrective maneuvers [15]. Still, it can be conjectured that in most circumstances, the cognitive loads of drivers may already be high when safety-related information must be issued, probably increasing driving risk rather than reducing it. In the case this conjecture proves correct, it may become fruitful to investigate automated driving interventions (without delivery of information) in particularly demanding traffic contexts, or when information would not come in time, or otherwise distract drivers even more.
4) *Visual stimuli drivers attend to can be identified:* Salient elements in the visual field of drivers elicit cephalo-ocular responses aimed at attending to such stimuli. Correspondences between cephalo-ocular behavior and visual stimuli must be established to identify the elements

within the visual field to which driver attention is turned. This knowledge will allow predictive models to assess whether drivers are attending to the appropriate stimuli, given current traffic contexts. This requirement implies that elements in the environment be correctly identified, located, and intersected with the 3-D gaze direction of the driver. Consequently, systems in charge of processing the output of stereo sensors must reliably detect the presence of other vehicles, pedestrians, and obstacles in general. This objective has only been partially attained with the use of passive sensing (CCD cameras) mainly because the reliability of most (if not all) techniques greatly depends on visual scene conditions [23]. While it is expected that passive vision systems will fail from time to time in difficult driving conditions, there may be effective methods of providing enhanced reliability by way of combining other sources of vehicular information. Vehicle-to-Vehicle (V2V) intercommunication may be used in situations where vision systems fail or underperform, such as times when fog, snow, or rain are present. Such communication modalities have the potential to both enrich and extend the range of visual sensors when surrounding vehicles signal their presence and position. These ideas may enhance the robustness of on-board vision systems and are further investigated in [24].

The creation of effective predictive driving behavior models rests on the confirmation of these hypotheses. While it is not expected that every aspect of these ideas can be empirically demonstrated, it is believed that their investigation will extend the current knowledge of the cognitive task of driving and allow for the establishment of strong principles for the design and operation of future on-board safety systems.

IV. Layered Approach to Vehicular Instrumentation

The next generation of i-ADAS will require extensive data fusion and analysis processes owing to an ever increasing amount of available vehicular information. In this context, a layered approach is best suited for real-time processing. In particular, such an approach enables bringing real-time data from sensors to a common level of compatibility and abstraction which significantly facilitates fusion and analysis processes. Our proposed computational model consists of four layers, with increasing levels of data abstraction (see Fig. 1). The innermost layer consists of the hardware and software required to capture vehicle odometry, sequences from visual sensors, and driver behavioral data. The second layer pertains to hardware synchronization, calibration, real-time data gathering, and vision detection processes. The third layer is where the data is transformed and fused into a single 4-D space (x, y, z, t). The last layer makes use of the fused data to compare driver behavioral data with models of behavior that are appropriate given current odometry and traffic conditions. While we proceed to describe the four layers, it is to be noted that this contribution specifically addresses the instrumentation (layers one and two) and its performance evaluation.

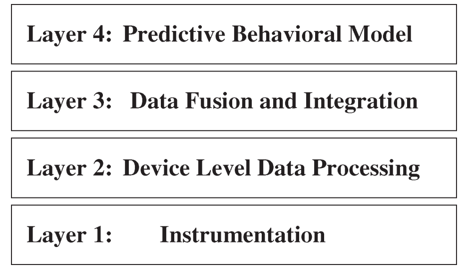

Fig. 1. Four layers comprising the data processing strategy on-board the instrumented vehicle.

A. Instrumentation

Contemporary vehicles equipped with on-board diagnostic systems (OBD-II) allow vehicle sensors to report on current status and constitute the interface through which odometry is made available in real time. Since 2008, the CANbus protocol[1] has become mandatory for OBD-II. This standardization simplifies the real-time capture of vehicle data. OBD-II to USB hardware interfaces with appropriate drivers are now common devices used to feed vehicle-related information to on-board computers or similar devices. The available information relevant to i-ADAS applications include current speed and acceleration (longitudinal and lateral), steering wheel rotation, state of accelerator and brake pedals, and independent wheel speed, which are real-time data captured at frequencies generally comprised between 20 and 200 Hz. These elements provide the information that is required to understand the maneuvers effected by the driver.

In addition, several vision systems must instrument the vehicle to appropriately monitor the immediate environment (lanes, other vehicles, pedestrians, obstacles, etc) and the behavior of the driver (gaze direction, level of attention, etc). These hardware systems must be capable of high sampling rates (30 Hz or more) such that sufficient accuracy in image processing and automated vision processes is achieved. It is useful to keep in mind that the position of a vehicle moving at 120 km/h changes by 33 meters every second.

Similar observations apply concerning the changes in visual gaze direction (known as saccades) as they occur very rapidly. For this reason, vision hardware monitoring the gaze direction of the driver must have sufficiently dense sampling rates as to allow for deriving driver intentionality prior to the execution of the anticipated behavior [19]. This part of the vehicle instrumentation is realized with commercial hardware and software[2] from which data such as eye gaze direction, vergence distance, and saccade events are obtained at a frequency of 60 Hz.

Also, part of the instrumentation layer is a GPS device which is used by V2V communications systems to provide other nearby instrumented vehicles with knowledge of traffic conditions beyond the range of their visual sensors.

Last but not least, on-board computing capabilities must also be sufficient to process the sum of incoming data in real time. To this end, we have designed and assembled a computer for

[1] The CANbus (Controller Area Network bus) provides microcontrollers with the means to communicate with each other within a vehicle.

[2] FaceLAB 5 implements our instrumentation for eye tracking.

Fig. 2. Color-coded calibrated stereo depth maps are obtained at 30 Hz. The distance between the instrumented vehicle and the roadside curbs, and other vehicles, is estimated in real-time.

real-time data processing and fusion consisting of 16 cores, each running at 3.0 GHz, with 16 GB of internal memory and a 128 GB solid state drive (SSD), with Linux Debian 5.01 as the operating system. The nodes are networked with a high-end gigabit network switch and configured as a diskless cluster, with the master node providing the operating system image to other nodes.

B. Justification

In the context of our hypothesis, it is vital that the instrumentation be able to provide information on current (and expected) driver behavior and vehicle operation. Two subsystems contribute to this goal. First, an OBD-II to USB interface[3] sends vehicular data (odometry and vehicle operation) to the on-board computer for recording or real-time analysis, or both. Second, an eye and head pose tracking system provides the necessary data for cephalo-ocular behavior recording and analysis. The sum of these subsystems provide the information required to determine the interactions between the driver and the vehicle, in addition to ocular behavior parameters. The resulting instrumentation allows to identify the visual stimuli drivers respond to in relation with the driving surroundings and the most probable behavior to be observed next.

Our choice of passive sensors is motivated by the fact that data acquisition is noninvasive and provides information conveyed by visual elements such as road markings and signs, which are critical to the task of driving and yet unavailable to active sensors such as radars or range finders [25]. In addition, multiple lens and sensor configurations are possible. For instance, IR or near-IR (Infra-Red) filters or sensors may readily be installed to provide night vision. Conversely, lenses of various types may be mounted on the sensors without any design modifications.

C. Device-Level Data Processing

For visual sensors, it is critical to obtain precise calibration parameters such as lens distortion, the optical center, and the external orientation of sensors with respect to each other. This calibration is required to perform stereo and to estimate distances of objects (other vehicles, pedestrians, etc.), which in turn greatly simplifies other vision-related tasks such as estimating motion, tracking, and obstacle detection. The RoadLAB stereo calibration interface was designed for this process. The interface is implemented using a calibration algorithm from the OpenCV 2.1 open source library based on Zhang's technique [26]. The calibration process consists of two steps. Intrinsic parameters are first estimated for each sensor and then, based on these, the extrinsic parameters for all possible sensor pairs are obtained. It is also possible to estimate the extrinsic parameters dynamically [27]. All the image frames from visual sensors are synchronized to within 125 μs. Once the synchronized frames are obtained, stereo depth maps are computed at frame rate, based on the calibration parameters (see Fig. 2). The GPS data is obtained through "gpsd," a GPS service daemon from "http://gpsd.berlios.de/" which provides an event-driven architecture. The data from the OBD-II/CANbus is obtained in a similar manner by creating a software layer for this purpose. Additionally, the incoming data from the instrumentation provides timestamps, allowing the system to fuse and select data elements in a synchronized fashion.

D. Data Fusion and Integration

Streams of data and video frames coming from monitoring the driver, the environment, and vehicle odometry must be placed in a suitable context for use by the behavioral prediction engine. We define a driver-centered frame of reference, in which elements of the cognitive state of driver (CSD) descriptor (head pose, gaze direction, blink events, lip movement), the contextual feature set (CFS) descriptor (road lanes, other vehicles, pedestrians, etc), and the vehicle state of odometry are transformed into, from their local sensor frames of reference (see Fig. 3 for a depiction of the CSD and CFS descriptors in the context of our layered model). This is performed by using the extrinsic parameters obtained with the calibration of the visual sensors with respect to each other. With these elements fused into a single frame of reference, the current CSD, CFS, and VSO descriptors are updated at 30 Hz and made available to the behavioral prediction engine.

Two modes of operation exist at this level. A recording mode captures the data and video streams from the instrumentation for in-laboratory, off-line analysis. A processing mode which performs as an i-ADAS operating in real time is also possible. Each sequence generated for off-line analysis obeys a strict format standard, in which the calibration data, the timestamped frames from the stereo systems, and the vehicle odometry are recorded at 30 Hz.

[3] A Kvaser Leaf Light OBD-II to USB device implements this part of the instrumentation.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

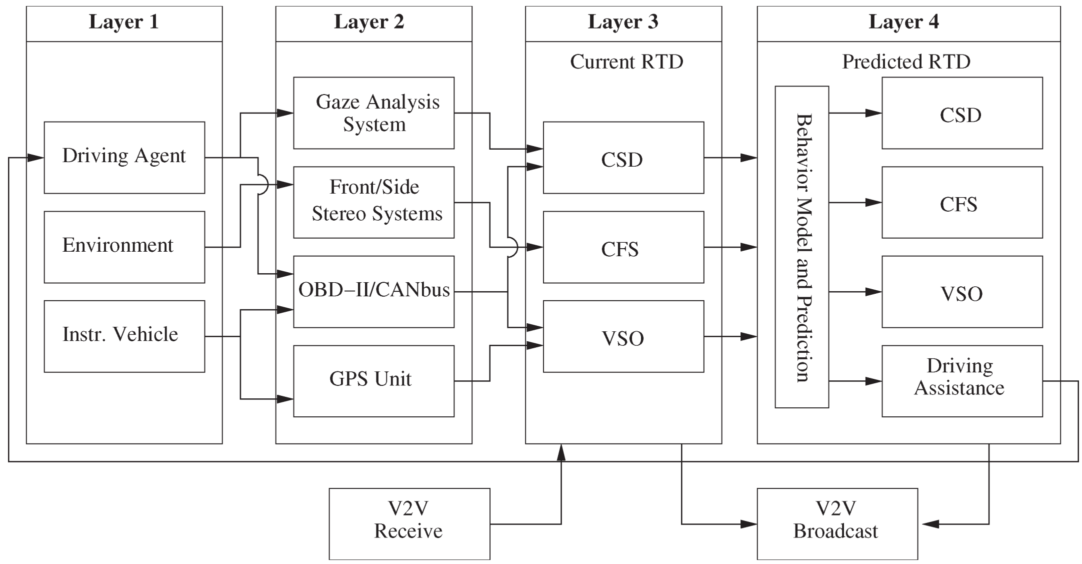

Fig. 3. Description of the retroactive mechanism operating between the current and predicted RTDs with respect to the outlined layered approach, in which driving assistance impacts both the current and predicted behavioral state of the driving agent. The reception of V2V information enriches the current CFS, which in turn impacts the predicted RTD. Informational elements from both the current and predicted RTDs are broadcast to other instrumented vehicles.

E. *Predictive Behavioral Model*

Our general hypothesis stems from research demonstrating that eye movements reflect moment-to-moment cognitive processes used to locate the information needed by the motor system for producing actions in relation to the environment [19], [18], [28]. This hypothesis is the foundation for our conjecture stating that the analysis of driver gaze direction (and other facial features) fused with the knowledge of the environment surrounding the vehicle (and its odometry) lead to the possibility of predicting driving behavior for short time frames (a few seconds). To accomplish these goals, it is necessary to infer a behavioral driving agent model that puts in relation the cognitive state of the driver, the vehicle odometry, and its surrounding environment as captured by sensors. For this purpose, we devise a real-time descriptor (RTD) for a moving vehicle essentially consisting of a CFS, a CSD, and a VSO descriptor.

These elements represent the knowledge required in composing an extensive RTD suited for our purposes. While we are interested in deriving practical and predictive driving agent models, it is worth noting that both the CFS and the VSO possess predictive models which are less difficult to formulate. We further propose to structure the elements of the RTD within a retroactive mechanism (see Fig. 3) in which both the current and predicted descriptors (CSD, CFS, and VSO) assist in determining not only the safety level of the context derived from the current RTD, but also that posed by the predicted RTD.

At the heart of the behavioral prediction engine is a Bayesian model which takes the current CSD, CFS, and VSO as inputs and predicts actuation behavior of the driver in the next few seconds. It also gathers statistical information about driving decisions and errors in a driver statistical record (DSR) which can be used over time to improve the prediction accuracy. The current CSD and CFS are in turn used to establish a driver memory of surroundings (DMS) based on the attention level and gaze direction analysis of the driver. A general forgetting factor is applied to the DMS as time elapses to reflect common characteristics of short-term visual memory. In addition, a driver cognitive load factor is inferred, based on the activities engaged by the driver, which in turn impacts the DMS, among other things.

V. Intervehicular Communication

Vehicular networks have been an area of research for the past two decades [29]. Interest has been shown by researchers, government agencies, and automobile manufacturers in developing the technologies and protocols for vehicular networks. There is a number of major areas of interest where unique problems must be solved, including protocols for the physical and link layer, higher layer protocols to deliver traffic, safety, and security information. Beyond these purely technical challenges, vehicles created by different manufacturers must be able to communicate, thus rendering standardization essential.

Some standardization has already occurred with the IEEE 802.11p draft standard and allocation of 75 MHz in the 5.9 GHz spectrum for dedicated short-range communications for the physical and link layer protocols. Further standardization with the IEEE 1609 draft standards for higher level protocols and services is ongoing [30]. Nonetheless, there are numerous open areas of research where solutions must be found before vehicular networks are adopted in consumer vehicles. How these two sets of technologies intersect is a topic that currently has not been looked at in depth. Vehicular networking technologies can provide detailed information about other vehicles in a large area, while sensor-based technologies can provide more detailed information about the environment immediately surrounding a vehicle in real time. How these two sets of technologies intersect is a topic that currently has not been looked at in depth.

Being able to combine both sources of information provides greater detail and breadth than any one technology can provide

Fig. 4. RoadLAB in-vehicle laboratory. (a) (Left): on-board computer and LCD screen. (b) (Center): dual stereo front visual sensors. (c) (Right): side stereo visual sensors.

Fig. 5. Various mounting configurations. (a) (Left): dual stereo sensors mounted on top of vehicle. (b) (Center): dual stereo sensors mounted on hood of vehicle. (c) (Right): experiment with an external sensor configuration.

on its own. How to do this exactly is an open research problem. There is also no guarantee that the information provided by these separate system will agree. There is a wide variety of circumstances in which data from both systems may not match and the vehicle will need to deduce which one is most likely correct.

Our approach consists of the systems, strategies, and implementation of the concept of using V2V to extend on-board visual sensor range. Coupling V2V and sensory input may increase detection reliability and range for visual sensors. Conversely, sensors may inform i-ADAS of the presence of noncommunicating elements such as pedestrians and nonvehicular obstacles. The potential that is held by integrating V2V communication with on-board sensory perception is considerable.

An instrumented vehicle navigating in an environment where other vehicles are similarly equipped would have access to critical traffic information well beyond the range of its sensors. Additionally, cascading information between communicating vehicles would allow a single vehicle to decide upon the range within which traffic information is deemed desirable.

While it seems natural to integrate sensory information with V2V in the context of i-ADAS, few research efforts have been conducted toward this goal. We believe that the complementarity of information obtainable from on-board sensors and V2V communication can form the basis for new approaches in driver-assisted systems and technologies [24].

VI. In-Vehicle Laboratory

The design of the instrumented vehicle follows principles of sensor portability and computing scalability. Sensor portability is achieved by using vacuum devices to attach the instrumentation equipment to the the interior or exterior surfaces of the vehicle (see Fig. 4), such as stereo camera rigs and LCD displays. Similarly, computing scalability is addressed with a diskless, master-slave cluster of computing nodes, configured with open source software from Sandia National Laboratories (OneSIS). Additional computing nodes and graphical processing units (GPUs) may be added at will, with the obvious cargo limitation imposed by the instrumented vehicle. Portability enables the use of a wide variety of vehicles without compromising their physical integrity, while computing scalability ensures an adequate supply of processing cores, matching the many possible sensor configurations (see Fig. 5).

A. *Physical Equipment*

Each minute, the sensory equipment sends 2 to 6 GB of data to the on-board computer, depending on the chosen sensory configuration. With such large amounts of data to process, the computing equipment was designed with scalability as a guiding principle. For this purpose, A diskless cluster arrangement was chosen essentially to provide the option of adding computing nodes as necessary. Currently, the on-board computer is composed of 16 computing nodes distributed over four boards networked with a gigabit switch. The nodes and the switch are contained inside a portable server case which in turn can be installed on the back seat or in the trunk of the vehicle. The computer and instrumentation are powered with a 1500 W inverter connected directly to the battery of the vehicle. The instrumentation can be run continually without battery drainage. See Fig. 6 for the schematics of the physical instrumentation.

B. *Mounting Configurations*

The visual sensors instrumenting the vehicle can be mounted in three distinct configurations. Fig. 5(a) and (c) depict an external, top mounting of the dual stereo head, while Fig. 5(b) shows an external hood-mounting configuration. Both of these

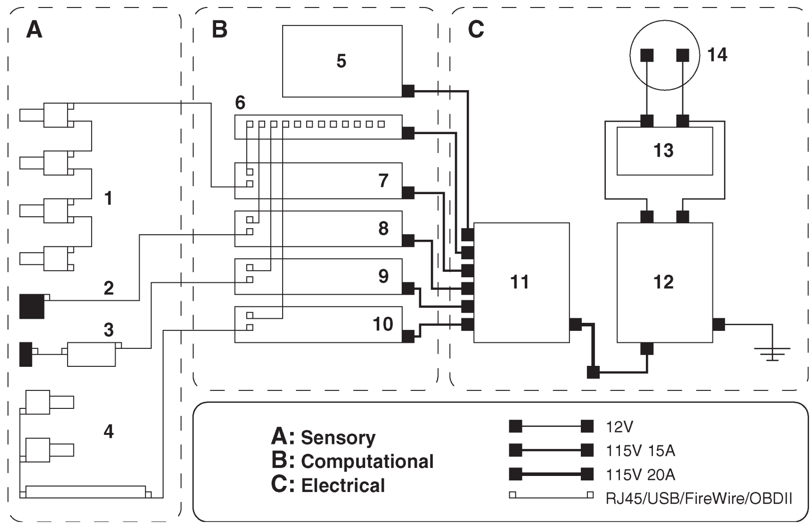

Fig. 6. Schematic vehicle instrumentation. (1). Dual stereo sensors. (2). GPS unit with USB interface. (3). OBD-II to USB interface. (4). FaceLAB 5 5): 19-in LCD display. (6). Gigabit network switch. (7). 3.0 GHz quad-core master node with 128 G Solid State Drive (SSD). (8), (9), (10). 3.0 GHz quad-core slave nodes. (11). 20 A power conditioner. (12). 1500 W power inverter. (13). Vehicle battery. (14). 140 A vehicle alternator.

set-ups do not hinder visibility for drivers. However, such external configurations limit the use of the instrumented vehicle to periods of clement weather (without rain, fog, or snow). To counter this limitation, the dual stereo head system was also designed to be mounted inside the front windshield of the vehicle [see Fig. 4(b)]. While this configuration allows the operation of the vehicle in variable weather conditions, it nonetheless hinders driver visibility substantially (a 2-h training session in closed-circuit is required before the vehicle can be safely driven on public roads). Another unintended effect created by this configuration is the visual distortion introduced by the presence of the windshield, which is not currently modeled by our calibration process, as it differs from radial and tangential distortion. Consequently, the quality and density of the raw stereo data is subjected to noticeable degradation (which lacks quantification at this time).

C. Software Services

The instrumented vehicle operational software architecture is based on a threaded publisher/subscriber model (see Fig. 7). Each component executes on its own core, to ensure real-time performance. The RoadLAB recorder, depicted in Fig. 8, receives images from the stereo heads at 30 fps, performs rectification, computes raw depth maps at frame rate, and saves the stereo images in a cyclic queue, which are to be written onto a solid state driven by an independent process which synchronizes with the recorder by means of semaphores.

The publisher/subscriber system receives information published by other software components such as the driver monitoring system[4], the OBD-II CANbus interface, and the GPS device. The recorder, in turn, may subscribe to various published elements and create instrumented sequences specifically designed for use in subsequent experiments. Alternatively, general-purpose instrumented sequences containing the totality of the published information can be produced. In general, the RoadLAB recorder may be used to provide real-time information to the resident i-ADAS application, or to produce instrumented sequences for in-laboratory experiments regarding the testing of sensing, integration, and i-ADAS algorithms.

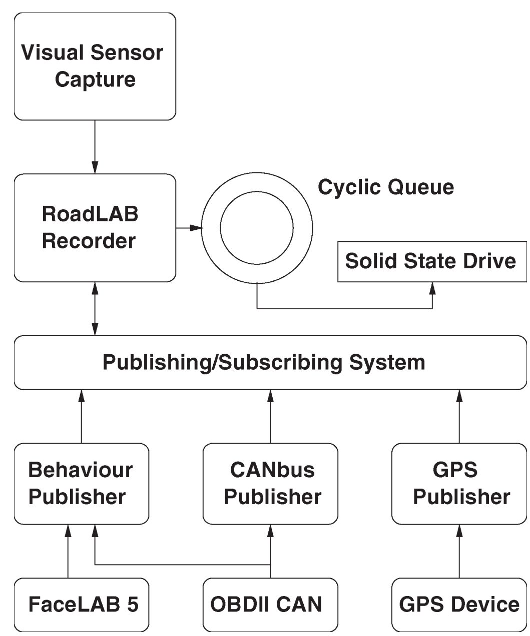

Fig. 7. Software services provided by the instrumented vehicle.

D. Vehicular Operation

Operating the instrumented vehicle consists of several steps which must be carefully followed. First, the instrumentation must be installed in the vehicle. The computing nodes, gigabit switch, and the power conditioner (located inside the portable server case) are mounted onto the back seating area, while the stereo sensors are installed in the chosen configuration by

[4]FaceLAB 5

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

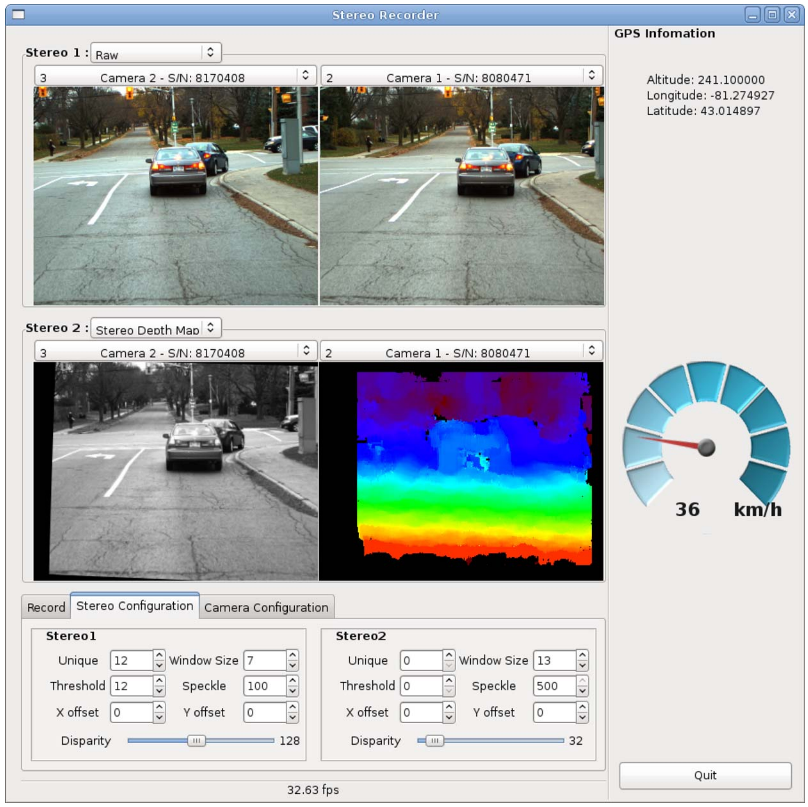

Fig. 8. RoadLAB Sequence Recorder in operation inside the instrumented vehicle.

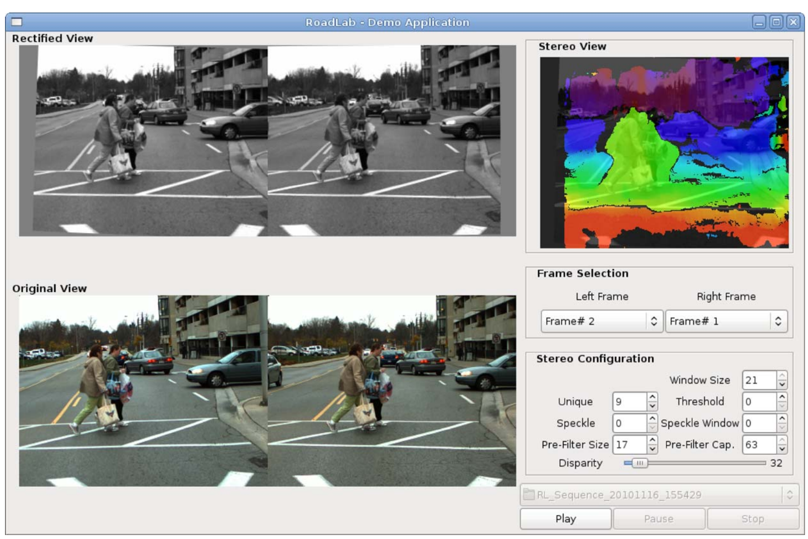

Fig. 9. Typical RoadLAB application using instrumented sequences produced with the vehicle operating in the recording mode.

way of vacuum devices. The GPS device magnetically attaches to the outside surface of the vehicle, and the CANbus to USB interface connects to the OBD-II outlet located under the vehicle instrumentation on the driver side. Once the sum of these elements are connected to the on-board computer, the calibration process takes place. A large calibration plane (125 by 155 cm) is used to capture sets of calibration images (25 to 30) each with a different orientation of the plane. A distance from 8 to 15 m must be respected between the vehicle and the calibration plane to obtain accurate calibration parameters, depending on the lenses being used with the stereo systems. A minimum of three trained research assistants and 60 min are required for completing the instrumentation and the calibration processes. At this stage, the vehicle can be operated in the recording mode, the i-ADAS mode, or both. Fig. 9 shows a typical off-line RoadLAB application using instrumented sequences produced with the vehicle in recording mode. Fig. 10 shows a real-time vehicle detection application, which is part of the resident i-ADAS software and constitutes an example of the vehicle being operated in the i-ADAS mode.

Fig. 10. Real-time vehicle tracking experiment using the RoadLAB instrumented vehicle in i-ADAS mode.

E. Limitations

Several limitations are experienced while instrumenting a vehicle for purposes such as ours. On the sensing side, the use of vacuum devices to attach the instrumentation to the vehicle limits the time of continuous vehicular operation to 30 min. After such time, the vacuum device pumps must be operated once more, to securely maintain the equipment in place. In addition, long-range lenses (with long focal lengths), when installed on the stereo systems, are sensitive to vibrations generated by both the condition of the pavement and the operation of the vehicle, resulting in a degradation of the raw 3-D depth data. This problem is made worse when the mounting configuration is located inside the windshield, as it introduces distortions that cannot be easily calibrated for. When conditions allow, an external mounting configuration coupled with short to medium range camera lenses leads to noticeably improved 3-D depth perception performance.

The availability of on-board computing power is inherently limited by the available space and electrical power in the vehicle. For instance, the use of high-resolution imagery would severely compromise our requirements for frame-rate processing. In this case, the problem may be addressed by replacing the computing nodes with GPUs, involving significant material costs. There is also the possibility of vehicle battery drainage with the use of high-end computing equipment, requiring the installation of a high-output, after-market vehicle alternator. In addition, our use of solid-state drives limits the amount of time the vehicle can be operated in recording mode. In our case, this limit is between 10 and 30 min, depending on how many visual sensors are in use while recording.

While these limitations are significant, the use of the instrumented vehicle for the validation of our previously stated hypotheses is justified, as we proceed to demonstrate.

VII. Methodological Considerations

Our current vehicular instrumentation is subservient to the validation of our hypotheses as described in this contribution and results from the following methodological considerations:

1) Instrumented sequences produced with test drivers are analyzed to determine what driving contexts correlate with cephalo-ocular behavior and to what extent this behavior can be considered predictive of driving actions. For this hypothesis to be tested correctly, drivers must be in an adequate state of alertness, which is measured by both eye saccade frequency and fixation mean duration [21]. Subsequently, correlations between cephalo-ocular movements and resulting driving actions are measured. We hope to find out which cephalo-ocular behavior predict driver intentionality. Insights gained from this approach assist in the creation of effective predictive models of driver behavior.
2) Driver level of attention may or may not provide significance to observed cephalo-ocular behavior when various driving environments are factored in. From instrumented sequences, it is possible to measure correlations between attention (defined as frequency and mean duration of glances away from the roadway) and driving environments (urban, rural, highway, congestion), to infer the meaningfulness of cephalo-ocular behavior (excluding fatigue-related considerations). These results assist in determining what factors are descriptive of meaningful cephalo-ocular behavior as it relates to driving.
3) Correlation between increases in cognitive load, defined as degradation of mean reaction time, and density of information delivery using a variation of modalities (audio, tactile, and visual), defined as events per time unit, is measured in an attempt to evaluate the effects of warning systems on the cognitive loads of drivers.
4) Our last hypothesis relates to computer vision processes and our advances are evaluated against those that operate in similar contexts. In this case, metrics are standard and relate to performance, measured as computational efficiency and quantitative accuracy. In addition, protocols for V2V in terms of improving on-board sensory range and robustness require other instrumented vehicles to communicate with, which are not available at this time, motivating our choice to explore this path with traffic simulators [24].

The in-vehicle laboratory as described in this contribution is capable of effecting the required measurements toward the validation of our hypotheses. Of particular importance is the extraction of driver behavior by using eye tracking and facial expression recognition techniques coupled with the maneuvers drivers apply to the vehicle, as obtained through the CANbus interface to form a basis for driver behavior prediction.

VIII. Performance Evaluation of Platform

The dual stereo systems constitute an essential component of the instrumented vehicle and for this reason, their performance (related to raw 3-D depth data) is crucially important. We first consider the problem of range resolution, which is inversely

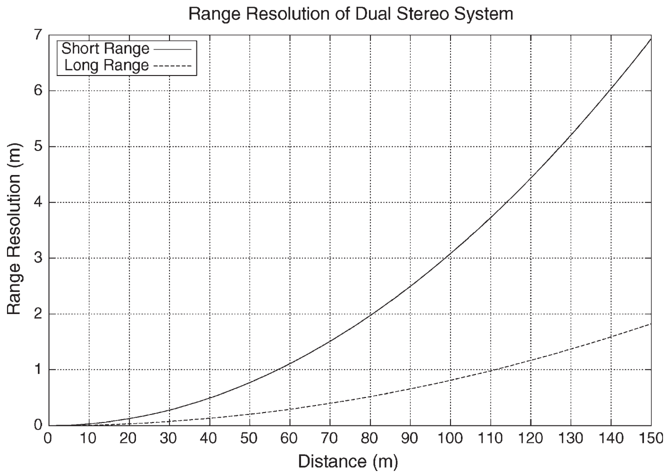

Fig. 11. Range resolution functions for dual stereo system, from 0 to 150 m.

related to object distance. The relationship governing range resolution is given by

$$\Delta r = \frac{r^2}{bf}\Delta d \tag{1}$$

where r is distance to object; f, focal length of imaging lens; b, stereo baseline length; and Δd, pixel size divided by the interpolation factor of the epipolar scan-line algorithm (for sub-pixel-precision 2-D matching). The range resolutions for our dual stereo systems constitute a reliable indication of the error levels contained in the depth data, provided that calibration is accurate and that the depth measurements do not stem from incorrect 2-D matches (due to occlusion, spatial aliasing, image noise, or related problems). Many dense stereo vision algorithms have been comparatively evaluated (including that of OpenCV, which we use) with image sequences for which true depth is available in terms of incorrect match density and resilience to noise [31]. The short-range stereo system has a baseline of length $b = 357$ mm, a smallest detectable 2-D disparity of $(1/16)$ of a pixel, a focal length of $f = 12.5$ mm, and a physical pixel square size of 4.40 μm. The long-range stereo system differs only in its baseline ($b = 678$ mm) and focal length ($f = 25.0$ mm). Fig. 11 displays the range resolution functions for both stereo systems. As expected, the range resolution of the long-range stereo pair surpasses that of the short range, due to an extended baseline and a longer focal length of the lens.

We have computed the average match density of both the long- and short-range stereo systems using instrumented sequences produced with the vehicle on public roads[5]. Results are reported in Table I, where different values of the minimum disparity[6] were used. As can be observed, the short-range stereo system performs better in terms of density, due to several factors, including the reported fact that operational vibrations introduce more noise in long-range systems.

TABLE I
STEREO MATCH DENSITY FOR SHORT AND LONG RANGE SYSTEMS, WHERE d IS MINIMUM DISPARITY AND D IS MATCH DENSITY WITH STANDARD DEVIATION σ

Stereo Average Density							
Short Range				**Long Range**			
$d = 32$		$d = 64$		$d = 64$		$d = 96$	
D	σ	D	σ	D	σ	D	σ
71.6%	9.0%	82.5%	10.1%	49.4%	7.7%	41.3%	7.5%

Each instrumentation layer as shown in Fig. 1 has access to four cores (one node) to perform its real-time tasks. A total of sixteen cores are available for the four instrumentation layers. Currently, only one in the four available cores for each layer is in use. While the software is in the later stages of development, its current performance at 30 Hz (for all layers) is consonant with the rate at which the visual sensors sample the environment. As the software modules are completed, the use of the remaining cores may become necessary to sustain the current performance. In the case where this would still be insufficient, an entire node can be added within the current configuration without any difficulty.

The performance of the quad-core computing nodes is largely sufficient to execute the stereo software at frame rate (30 fps). While one core suffices for the stereo computation, other cores may also be involved in processing other visual aspects of the captured frames and hence the speed at which frames can be transferred from one node to another is a critical constraint. By way of a high-end gigabit switch, the cores transfer frames (with resolution of 320 by 240 pixels) between nodes at 1.4 MHz (or 0.7 ms per frame), a speed which does not impede on the performance of the system. Additionally, the highest transmission rate on the OBD II CANbus was measured at 200 Hz, while our system reads and stores CANbus status at 2 MHz, ensuring that no incoming message could be missed out.[7]

IX. CONCLUSION

We have addressed the problem of vehicle instrumentation as an experimental platform for the design of i-ADAS, while maintaining our requirements for physical portability and computational scalability. We framed the data processing strategy of the instrumentation within a layered approach in which data abstraction increases with the number of layers. The predictive behavioral model was also integrated with our layered structure, yielding a comprehensive implementation for hardware, software, and data abstraction framework. The resulting in-vehicle laboratory, its various configurations, software services, and operation modes were described in depth. We demonstrated that this platform, in spite of its limitations, can be effectively used to address the hypotheses we formulated in relation to the design of i-ADAS.

[5]The instrumented sequences used to perform these computations are publicly available at www.csd.uwo.ca/faculty/beau/roadlab_download/index.html.

[6]The minimum disparity parameter controls the offset to the disparity search window. Increasing positive values has an effect identical to augmenting the convergence of the stereo cameras.

[7]Performance ratings of other aspects of our instrumentation such as the GPS device (GloablSat BU-353) and FaceLAB 5 are published by manufacturers and not reported herein.

REFERENCES

[1] E. Krug, "Road traffic injuries," WHO Overview Fact Sheet, 2000.

[2] N. H. T. S. Administration, "Traffic safety facts 2006: A compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system," US DOT, Washington, DC, Tech. Rep. HS 810 818, 2006.

[3] J. Liu, Y. Su, M. Ko, and P. Yu, "Development of a vision-based driver assistance system with lane departure warning and forward collision warning functions," in *Proc. Comput. Tech. Appl.*, 2008, pp. 480–485.

[4] M. Rockl, P. Robertson, K. Frank, T. Strang, and G. Center, "An architecture for situation-aware driver assistance systems," in *Proc. IEEE 65th Veh. Technol. Conf.*, 2007, vol. 4, pp. 2555–2559.

[5] Y. Umemura, "Driver behavior and active safety," *R&D Rev. Toyota CRDL—Special Issue*, vol. 39, no. 2, pp. 1–8, Apr. 2004.

[6] L. Petersson, L. Fletcher, and A. Zelinsky, "A framework for driver-in-the-loop driver assistance systems," in *Proc. IEEE Intell. Transp. Syst.*, 2005, pp. 771–776.

[7] S. Beauchemin, M. Bauer, D. Laurendeau, T. Kowsari, J. Cho, M. Hunter, and O. McCarthy, "Roadlab: An in-vehicle laboratory for developing cognitive cars," in *Proc. 23rd Int. Conf. CAINE*, Las Vegas, NV, Nov. 2010.

[8] I. Masaki, "Machine vision systems for intelligent transportation systems," *IEEE Intell. Syst. Appl.*, vol. 13, no. 6, pp. 24–31, Nov./Dec. 1998.

[9] D. Gavrila, U. Franke, C. Wohler, and S. Gorzig, "Real-time vision for intelligent vehicles," *IEEE Instrum. Meas. Mag.*, vol. 4, no. 2, pp. 22–27, Jun. 2001.

[10] B. Stoneman and D. Bowers, "Development of an instrumented vehicle for traffic and safety assessmennt," in *Proc. 7th Int. Conf. Road Traffic Monit.*, London, U.K., Apr. 1994, pp. 115–118.

[11] D. Gavrila, M. Kunert, and U. Lages, "A multi-sensor approach for the protection of vulnerable traffic participants—The protector project," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, Budabest, Hungary, May 2001, pp. 2044–2048.

[12] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. van Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, A. Nefian, and P. Mahoney, "Stanley: The robot that won the DARPA grand challenge," *J. Field Robot.*, vol. 23, no. 9, pp. 661–692, Sep. 2006.

[13] M. Land and J. Horwood, "Which parts of the road guide steering?" *Nature*, vol. 377, no. 6547, pp. 339–340, Sep. 1995.

[14] J. Lee, "50 years of driving safety research," *Hum. Factors*, vol. 50, no. 3, pp. 521–528, Jun. 2008.

[15] B. Donmez, L. Boyle, J. Lee, and G. Scott, "Safety implications of providing real-time feedback to distracted drivers," *Accid. Anal. Prev.*, vol. 39, no. 3, pp. 581–590, May 2007.

[16] D. Ballard, M. Hayhoe, F. Li, and S. Whitehead, "Hand-eye coordination during sequential tasks," *Philos. Trans. R. Soc. Lond. B*, vol. 337, no. 1281, pp. 331–338, Sep. 1992.

[17] D. Ballard, M. Hayhoe, and J. Pelz, "Memory representations in natural tasks," *J. Cognit. Neurosci.*, vol. 7, no. 1, pp. 66–80, 1995.

[18] K. Ker, I. Roberts, T. Collier, F. Beyer, F. Bunn, and C. Frost, "Post-license driver education for the prevention of road traffic crashes; a systematic review of randomized controlled trials," *Accid. Anal. Prev.*, vol. 37, no. 2, pp. 305–313, Mar. 2005.

[19] M. Land, "Eye movements and the control of actions in everyday life," *Progr. Retin Eye Res.*, vol. 25, no. 3, pp. 296–324, May 2006.

[20] Y. Lee, J. Lee, and L. Boyle, "Visual attention in driving: The effects of cognitive loads and visual disruption," *Hum. Factors*, vol. 49, no. 4, pp. 721–733, Aug. 2007.

[21] D. Shinar, "Looks are (almost) everything: Where drivers look to get information," *Human Factors*, vol. 50, no. 3, pp. 380–384, Jun. 2008.

[22] A. Amditis, K. Pagle, S. Joshi, and E. Bekiaris, "Driver-vehicle-environment monitoring for on-board driver support systems: Lessons learned from design and implementation," *Appl. Ergonom.*, vol. 41, no. 2, pp. 225–235, Mar. 2010.

[23] W. van der Mark and D. Gavrila, "Real-time dense stereo for intelligent vehicles," *IEEE Trans. Intell. Trans. Syst.*, vol. 7, no. 1, pp. 38–50, Mar. 2006.

[24] M. Bauer, K. Charbonneau, and S. Beauchemin, "V2eye: Enhancement of automated visual perception from v2v," in *Proc. CCNC*, Las Vegas, NV, Jan. 2011, pp. 737–738.

[25] M. Bertozzi, A. Broggi, M. Cellario, A. Fascioli, P. Lombardi, and M. Porta, "Artificial vision in road vehicles," *Proc. IEEE*, vol. 90, no. 7, pp. 1258–1271, Jul. 2002.

[26] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[27] J. Collado, C. Hilario, A. de la Escalera, and J. Armingol, "Self-calibration of an on-board stereo-vision system for driver assistance systems," in *Proc. IEEE Intell. Veh. Symp.*, Tokyo, Japan, 2006, pp. 156–162.

[28] K. Rayner, "Eye movements in reading and information processing: 20 years of research," *Psychol. Bull.*, vol. 124, no. 3, pp. 372–422, Nov. 1998.

[29] M. Sichitiu and M. Kihl, "Inter-vehicle communication systems: A survey," *Commun. Surveys Tuts.*, vol. 10, no. 2, pp. 88–105, Second Quarter, 2008.

[30] D. Jiang and L. Degrossi, "IEEE 802.11p: Towards an international standard for wireless access in vehicular environments," in *Proc. IEEE Veh. Technol. Conf.*, 2008, pp. 2036–2040.

[31] H. Sunyoto, W. van der Mark, and D. Gavrila, "A comparative study of fast dense stereo vision algorithms," in *Proc. IEEE Intell. Veh. Symp.*, Parma, Italy, Jun. 2004, pp. 319–324.

Steven S. Beauchemin received the M.Sc. and Ph.D. degrees from the University of Western Ontario, London, ON, Canada, in 1991 and 1997, respectively.

He is Associate Professor of Computer Science at the University of Western Ontario, London, ON, Canada. His research interests revolve around computer vision techniques for the automotive industry. As part of past and present research projects, Dr. Beauchemin has been involved in the training and retraining of elderly drivers in simulators and in instrumented vehicles. These projects are funded in part by the Auto21 center of excellence (www.auto21.ca).

Dr. Beauchemin also leads a team of graduate researchers and students in the use of vision technologies for intelligent, Advanced Driving Assistance Systems.

Michael A. Bauer (M'78) received the M.Sc. and Ph.D. degrees from the University of Toronto, Toronto, ON, Canada, in 1971 and 1978, respectively.

He is a Professor of Computer Science at the University of Western Ontario, London, ON, Canada. He was Chair of the Computer Science Department at Western Ontario from 1991 to 1996 and from 2002 to 2007. From 1996 to 2001, he was the Associate Vice-President Information Technology at the University of Western Ontario. He served on NSERC's (National Science and Engineering Research Council) Computer Science Grant review committee from 2005 to 2008 and was chair of the committee in 2008–2009. He was Principal Investigator for the CFI project that initially funded the creation of SHARCNET (www.sharcnet.ca)—a multi-university high performance computing grid. He is currently the Associate Director for SHARCNET. His research interests lie in the fields of distributed systems, mobile and vehicular applications and communication, high performance and parallel computing. He has published over 220 refereed articles, has served on the organizing and program committee of numerous conferences, and has refereed for a variety of international journals. He has supervised over 70 graduate students.

Prof. Bauer is a member of the IEEE and the Association for Computing Machinery and has served on various committees of both organizations.

Taha Kowsari received the M.S. degree in AI and robotics from the University of Tehran, Tehran, Iran. He is currently working toward the Ph.D. degree in computer vision at the University of Western Ontario, London, ON, Canada.

He is a Researcher at the RoadLAB Research Laboratory. His research interests include computer vision, intelligent transportation systems, intelligent, Advanced Driving Assistance Systems, robotics, and human–machine interface.

Ji Cho received the B.Sc. degree in computer science from the University of Waterloo, Waterloo, ON, Canada. He is currently working toward the M.Sc. degree in computer vision at the University of Western Ontario, London, ON.

He is a Researcher at the RoadLAB Research Laboratory. His research interests include computer vision, intelligent transportation systems, and intelligent, Advanced Driving Assistance Systems.