Looking-In and Looking-Out of a Vehicle: Computer-Vision-Based Enhanced Vehicle Safety

Mohan Manubhai Trivedi, Tarak Gandhi, *Member, IEEE*, and Joel McCall

***Abstract*—This paper presents investigations into the role of computer-vision technology in developing safer automobiles. We consider vision systems, which cannot only look out of the vehicle to detect and track roads and avoid hitting obstacles or pedestrians but simultaneously look inside the vehicle to monitor the attentiveness of the driver and even predict her intentions. In this paper, a systems-oriented framework for developing computer-vision technology for safer automobiles is presented. We will consider three main components of the system: environment, vehicle, and driver. We will discuss various issues and ideas for developing models for these main components as well as activities associated with the complex task of safe driving. This paper includes a discussion of novel sensory systems and algorithms for capturing not only the dynamic surround information of the vehicle but also the state, intent, and activity patterns of drivers.**

***Index Terms*—Active safety, driver-support systems, intelligent vehicles, real-time machine-vision systems.**

I. Introduction and Research Motivation

AUTOMOBILES were at the core of transforming lives of individuals and nations during the 20th century. The century started with production of a few hundred automobiles per year and ended with over 50 million units produced for global consumption annually. Unfortunately, along with the growth of the automobile usage, the numbers of accidents leading to fatalities and serious injuries have seen dramatic increases. Traffic-related accidents are recognized as a serious and growing problem with global dimensions. A recent study by World Health Organization mentions that annually, over 1.2 million fatalities and over 20 million serious injuries occur worldwide [1]. Enhancement of traffic safety is pursued as a high-priority item not only by various government agencies such as National Transportation Safety Administration [2] but also by most major automobile manufactures. University-based researchers are also contributing to this important mission.

Manuscript received March 6, 2006; revised August 19, 2006, August 30, 2006, and September 9, 2006. This work was supported in part by the UC Discovery Grant's Digital Media Innovations Program, in part by the Nissan Motor Corporation, and in part by the Volkswagen Corporation. The Associate Editor for this paper was N. Zheng.

M. M. Trivedi and T. Gandhi are with the Computer Vision and Robotics Research Laboratory, University of California, San Diego, La Jolla, CA 92093-0434 USA (e-mail: mtrivedi@ucsd.edu; tgandhi@ucsd.edu).

J. McCall was with the Computer Vision and Robotics Research Laboratory, University of California, San Diego, La Jolla, CA 92093-0434 USA. He is now with Microsoft Corporation, Redmond, WA USA 98052-6399 (e-mail: jmccall@ucsd.edu).

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/TITS.2006.889442

In order to develop effective counter measures for enhancing safe and smooth operation of an automobile in traffic, it is helpful to examine the full context in which driving occurs. There are three main components of the overall driving system:

1) **Environment**: including roadway infrastructure and dynamic climatic situations;
2) **Vehicle**: including ever-increasing telematic devices and infotainment gadgetry;
3) **Driver**: an essential part of the human–vehicle system, which needs to be maneuvered safely in the environment.

The complex dynamics of various events and interaction of various entities in the above tripartite "EVD" system components affect the overall safety of a vehicle as well as the condition of the traffic flow. For instance, properly designed roads, traffic signs, and traffic regulations and policies have all been recognized as important factors in making traffic safer on the U.S. Interstate Highways. Improved design of vehicles and safety systems, such as seat belts, brakes, and airbags are key factors in reducing injuries.

The vehicle-based safety systems are typically viewed as one of the two kinds. The first one is termed as "Passive." The purpose here is to minimize the severity of injuries sustained in case of accidents. Examples of these are seat belts, airbags, collapsible steering columns, and shatter-resistant windshields. The second kind is "Active," which is supposed to prevent vehicular accidents. Good examples of these are antilocking brakes. Obviously, it is more desirable to prevent an accident rather than reduce the severity of injuries. However, active-safety systems pose a lot more difficult and challenging problems. One of the key requirements in the design of an active-safety system is the ability to accurately, reliably, and very quickly identify the conditions which would lead to an accident and to force corrective actions so that the accident can be prevented.

An active-safety system has three parts as shown in Fig. 1. The front end of an active-safety system is a sensing subsystem, which needs to provide an accurate description of the dynamic state of the EVD system. The second important subsystem is an analysis subsystem which needs to analyze the EVD dynamic state using a model-based approach to compute some sort of a measure of safety underlying that particular EVD state. If this measure falls under a predefined threshold of margin of safety, then the analysis module needs to direct the active-safety control unit to initiate a corrective course of action so that the vehicle can always operate within the margins of an accident-free safety zone. There are some very challenging problems involved in each of the above three subsystems of an

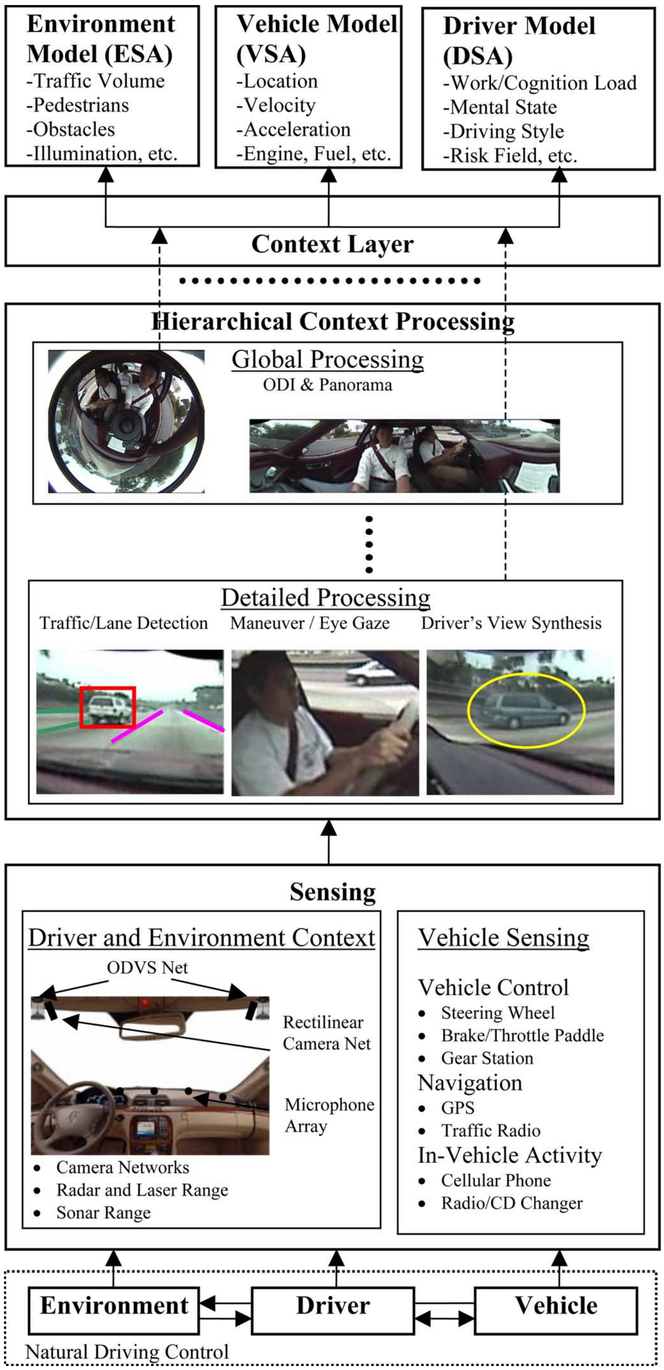


Fig. 1. Framework for multifunctional "active" computer-vision-based dynamic context-capture system.

active-safety system. Primarily these challenges can be stated in the following.

1) What are the critical states of the EVD system, which need continuous monitoring?
2) How do we sense these states without violating all of the constraints associated with "real-world" driving?
3) What are the models, measures, and metrics that define margins of safety?
4) What are the appropriate corrective actions, which should be initiated to bring the vehicle back in the safe operating EVD zone?

Satisfactory resolution of these challenging problems will require sustained efforts of multidisciplinary research teams over a long period of time. It is encouraging to note that a number of research teams, comprising of academic, industry, and government agencies have initiated serious efforts in this area. In this paper, we present an overview of a selected research studies conducted in our research laboratory, where novel concepts and systems based upon computer-vision technology are developed for enhancement of vehicle safety. It includes a brief overview of related literature in Section II and a description of the basic Looking-in Looking-out (LiLo) vision framework in Section III. A presentation of a number of systems for capturing useful visual contextual information from inside as well as outside the vehicle and for holistic analysis of such information are given in Section IV. Summary and concluding remarks are presented in Section V. A short description of the vehicle testbeds used in this research is described in the Appendix.

II. Related Investigations

A. Computer Vision for Active Safety

Recognition of computer vision as a critical technology for intelligent vehicles can be traced to the earlier efforts dealing with autonomous mobile robots and autonomous driving [3]–[5]. Such efforts helped to demonstrate the power of camera-based systems to support real-time control of vehicles. In [6] and [7], Bertozzi *et al.* give a comprehensive survey of the use of computer vision in intelligent vehicles. Approaches for lane, pedestrian, and obstacle detection are described and analyzed.

A new trend started emerging in the late 1990s, where research in computer vision focused on enhancement of the safety of automobiles [8], [9]. Already, camera-based modules with safety-oriented features such as "back-up" (or reverse) viewing, lane-departure warning, and blind-spot detection are offered in commercial vehicles. It was realized that in addition to monitoring the surroundings, the monitoring of the driver state is also important for improving safety. For example, a system for drowsy-driver detection, predominantly for nighttime, driving is presented in [10].

B. Driver Behavior and Active Safety

Active-safety systems need to accurately predict conditions which would lead to accidents unless proper corrective actions are undertaken. Human factors researchers have, over a long period of time, examined the influence of driver behavior on safety [11]. Driver distraction is considered to be a factor in over 25% of accidents [2]. Many of these studies have relied upon extensive laboratory and simulator-based trials. A noteworthy recent study called "100-Car Naturalistic Study" [12] recognized the need for observing driver behavior, vehicle state, and environment conditions. This study was conducted over a one-year period and involved over 240 drivers and 100 instrumented vehicles. Preliminary analysis of the extensive data captured during the study indicates driver behavior to be a contributing factor in crashes and near crashes. In addition to the human factor research, contributions by perceptual psychologists are essential in developing models for driver's attentive state. There is a large body of literature based upon both theoretical as well as experimental research in these areas. Interested readers can find useful pointers in some of the recent papers published in the literature [13]–[16].

III. LiLo Framework

Driving an automobile on an urban street or a crowded highway is a demanding task. Drivers are inherently limited in their ability to maintain accurate awareness of the driving context, especially when the situation is complex, the weather conditions adverse, or their task load too high. Hence, it is important to develop a system that improves the efficiency and effectiveness of drivers' attention allocation. It is important to provide drivers with a rich and highly informative driving environment. The goal of such system is to increase safety at the same level of workload or decrease workload for the same level of safety. Much effort is needed to assure that drivers do not lower the driving workload at the expense of safety. Such undesirable consequences of especially active-safety systems are generally the result of the assignment of an excessive level of capability and trust in the system by the driver [17], [18]. If the driver places too much confidence on the system, then the driver may not be ready to respond in an efficient and effective manner when the system encounters a situation that reaches beyond its design specifications. Instead, we propose to keep the driver in the loop and directly communicate the confidence that the system has in the support and guidance it offers the driver. This approach bypasses the difficult issue in automation, namely dynamic allocation of function or who is responsible and when.

Dynamic context capture for such a Human-Centered Intelligent Driving-Support System requires analysis and fusion of multimodal sensory information at multiple levels of abstraction. The sensor system should be capable of maintaining dynamic representations of the external world surrounding the vehicle, the state of the vehicle itself, and of the driver. To develop a robust dynamic context-capture system, computer vision, and machine-learning techniques play an important role. In this paper, we have pursed development and evaluation of an active multimodal sensory approach for "dynamic context capture and situational awareness" using cameras, radars, audio, and other sensors for establishing representations of the state of the environment, the vehicle, and the driver with accurate dynamic uncertainty characterization.

The overall objective of our studies is to seek answers to the following important questions.

1) What sets of sensors are robust enough under a wide variety of environmental conditions?
2) What contexts can support a sufficiently complete representation of the environment, the vehicle state, and the driver state?
3) What is the best computational framework to extract contexts from sensor networks that is compatible with human perception?
4) What are the best models of the environment, the vehicle state, the driver state, and the knowledge that drivers may have of the driving ecology?
5) How to classify and explain driver behavior according to the tasks that the driver is engaged in, the tasks that the driver is intending, and the safety margin of the driver to perform the task?

Resolution of these questions requires expertise from multiple disciplines. Teams from engineering and computer sciences need to examine multimodal-signal processing, pattern recognition, and decision and control theories. Cognitive scientists need to consider human–machine interfaces, whereas psychologists need to study models for perception, attention, and multitasking.

For dynamic context capture, vehicle-based state-of-the-art integrated sensor suites are pursued. We propose a hierarchical framework, as shown in Fig. 1, for context processing capable of dynamically allocating computational and sensor resources depending on the demands imposed by the complexity of environment and the vigilance level and cognitive load of the driver. These can be captured with relatively few resources and used to modulate the level of detail and integration in processing the vast amount of data from the multimodal-sensor network. Vision is one of the primary sensing modality, as driving is considered to be a visually guided activity. For dynamic capture of visual context, three basic types of viewing perspectives for the cameras can be considered:

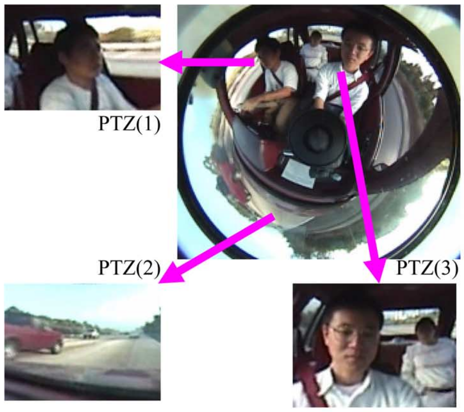
1) Looking in the vehicle: to capture the important visual context associated with the driver, occupant, and their activities and physical- and mental-state monitoring;
2) Looking out of the vehicle: to capture the visual context of the vehicle, including that of the surrounding road conditions and traffic;
3) Simultaneous Looking-in and Looking-out of the vehicle: to correlate the visual contextual information of vehicle interior and vehicle exterior. This approach would allow for systematic investigations of driver's behavior and intent. The objective is to derive useful feedback mechanisms for managing driver distraction.

In the remainder of this paper, we present an overview of a selected computer-vision-based systems, which provide the essential driver-assistance systems. The basic objective is the development of novel vision systems and appropriate algorithms for enhancement of safety. Such a development is pursued within the context of a wide range of application possibilities, including those for occupant safety, pedestrian safety, driver assistance, driver workload and "attention" monitoring, lane keeping, and dynamic panoramic-surround capture.

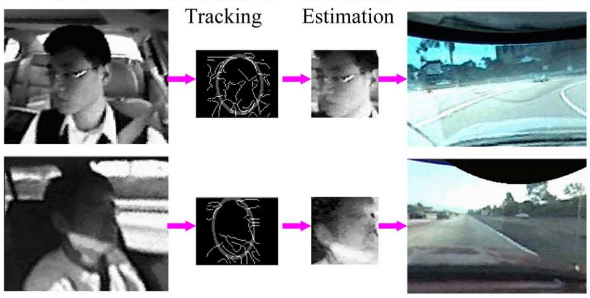
IV. Vision Systems for Enhanced Safety: Illustrations

A. Looking-In the Vehicle: Occupant Position and Posture

Analysis of the state of driver and passengers is important for capturing their physical and mental state, behavior, activities,

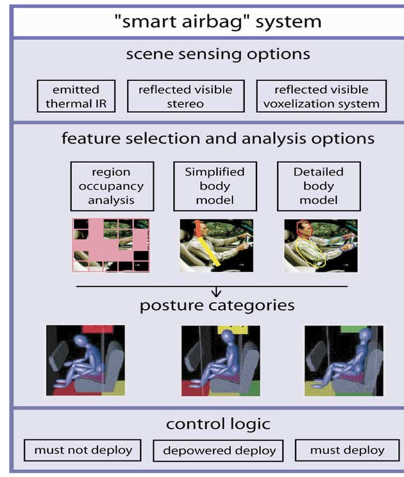


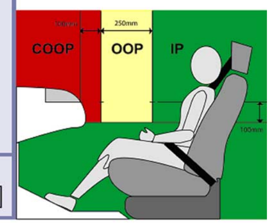
(a)



(b)

Fig. 2. (a) Generation of multiple virtual-perspective views of driver's face as well as surroundings from single omni camera mounted inside the vehicle. (b) Generation of driver's view based on estimation of face orientation.





(a) (b)

Fig. 3. Smart airbag system that makes airbag-deployment decision based on position of the occupant. (a) Block diagram. (b) Areas for In-position (IP), Out-of-position (OOP), and Critically out-of-position (COOP) regions.

and intentions. The interior context plays an important part in many intelligent vehicle systems, in particular, for enhancing safety. For example, the monitoring of driver's face and eyes can give information about the focus of attention, fatigue, drowsiness, and other attributes of the driver. The position and motion of hands and feet provide information about the steering and braking actions and intentions. In case of other occupants, the position of the body is important for making decisions about deployment of airbags. The following sections describe our research in analyzing the visual context of the interior of the vehicle.

1) Driver-View Generation Using Head-Pose Estimation: The direction in which the driver is looking is important for knowing the focus of attention of the driver, as well as the parts of the scene that are likely to be missed by the driver. This knowledge is important for generating appropriate warnings for dangerous events taking place outside the driver's attention. In [19], we describe an application to generate instantaneous driver's view by estimating the driver's head pose. A single omnidirectional camera mounted inside the vehicle is used to observe the driver's face as well as the vehicle surroundings, with capability to generate virtual views in any given direction, as shown in Fig. 2(a). The virtual view of the head is used to extract the driver's face and then estimate the viewing direction from the face image. Based on the viewing direction, another virtual view in the direction that the driver is looking is generated. The algorithm consists of the following stages as shown in Fig. 2(b).

1) Head detection by extracting edge map of the face image and fitting an ellipse using randomized Hough transform. For robust detection and accurate localization, a Kalman filter is used to track the head and predict its location in the subsequent frame.
2) Face-orientation estimation by comparing the extracted face with principal component analysis (PCA) templates generated from training samples with head in various orientations. The head orientation is selected based on maximum likelihood and tracked using a Kalman filter.
3) Driver's view generation using virtual-perspective view based on driver's face orientation that gives the approximate viewing direction of the driver.

Details of the approach are provided in [19]. With a 2.8-GHz Xeon processor (2.8 GHz) running the head-detection step, the processing speed is about 15 ft/s for 1024×768 images from a Firewire camera. Most of the computational time is spent on ellipse fitting. For the current implementation, the face-orientation estimation is implemented offline on MATLAB. A C++ implementation of hidden Markov model (HMM) should run quite fast and not reduce the processing speed significantly.

2) Occupant-Posture Analysis for Safe Airbag Deployment: In this section, we briefly discuss our research [20] on the development of a highly reliable and real-time vision system for sensing passenger occupancy and body posture in vehicles, ensuring safe airbag deployment, and helping to prevent injuries. The design of the "smart airbag" system can be divided into four parts, as shown in Fig. 3(a): 1) real-time scene sensing; 2) feature selection and analysis; 3) body size, posture, and movement analysis; and 4) decision logic for various levels of airbag deployment.

To determine whether a person is in the right position for airbag deployment, the area between the back of the seat and the dashboard can be divided into sections, as shown in Fig. 3(b). By analyzing these regions, we can categorically examine the human body under various positions that an occupant can take in the passenger seat, including sitting in a normal position, leaning forward, reaching down, seated with the seat advanced, reclined, slouched, knees on the dashboard, or the edge of the seat.

The following options were considered for scene sensing: 1) emitted long-wave infra-red (LWIR) imaging; 2) stereo-based depth imaging; and 3) multicamera-based voxel reconstruction. For feature selection and analysis, we consider simple region occupancy features as well as detailed human-body-model-pose estimation. Using stereo or multicamera systems with high-level human-body modeling would also provide information useful for other applications with minimal extra effort.

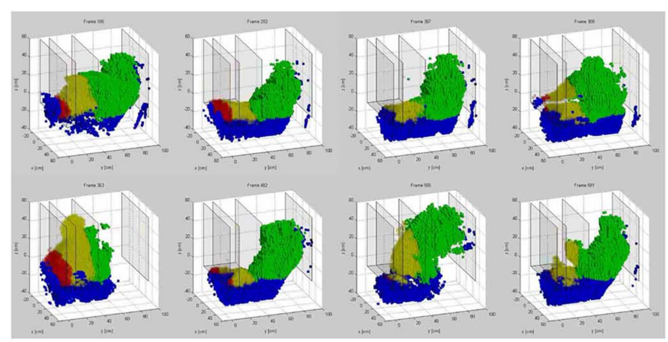
The algorithm for determining the occupant position is based on the head-detection algorithm developed by Eleftheriadis and Jacquin [21]. The original algorithm performs image preprocessing, computes edge map based on image gradients, performs ellipse detection on the edge image, and tracks the ellipse shape to give the orientation and size of the head in real time.

We have modified the algorithm in order to obtain improved results using stereo as well as LWIR sensors. In case of stereo, the available depth information allows for a larger degree of lighting changes than allowed by the reflectance model. Pixels with disparity values corresponding to the background and falling outside the car are eliminated. This improvement helps to remove extraneous background and foreground data. After detecting the ellipse corresponding to the head, its size is cross-checked with the depth information provided by stereo. In case of thermal images, the pixel intensity values are mapped to probability of membership to human class based on temperature using a Gaussian probability distribution. The average skin probability is also checked after ellipse detection in order to eliminate false detections. Fig. 4(a) and (b) shows examples of head detection using stereo as well as LWIR cameras. On a standard Pentium 4 PC computing platform, the IR and stereo-based approaches achieve frame rates of 3 and 6 frames per second, respectively. The majority of computational time in IR-based approach is spent in template matching, whereas in the case of stereo, the stereo disparity map computation as well as the head detection takes most of the computational time. Specialized hardware is likely to yield improvements in processing speed.

We have also investigated the use of voxel data obtained using multiple cameras by the shape-from-silhouette technique (SFS) to extract occupant-posture information. SFS is a technique that uses the silhouette images to reconstruct the visual hull of the object. The visual hull is the closest reproduction of an object's actual volume using an arbitrary number of silhouette images of the object [22]. The application of full or even partial body go beyond occupant-posture estimation for the purpose of "smart airbag" deployment and could be useful for driver-fatigue analysis, driver attentiveness, and human–machine interfaces inside the car. Fig. 4(c) shows the reconstruction of the upper body using multiple cameras. Detailed results are described in [20]. The voxel-based approach gives the frame rate of 0.23 frames per second corresponding

(a)

(b)



(c)

Fig. 4. (a) Head detection using stereo cameras. (b) Head detection using thermal cameras. (c) Three-dimensional reconstruction of upper part of body using multiple cameras.

to 4.3 seconds for each frame. The capture, segmentation, and voxelization are implemented in C++ and take only 180 ms. However, for the current implementation, the head detection is performed in MATLAB and takes the majority of the computational time. A C++ implementation should give significant improvement in processing time.

B. Looking-Out of the Vehicle: Dynamic Panoramic Surround (DPS)

A complete driver-support system should be aware of all of the surroundings of the vehicle. The main components of the vehicle surround are lanes, vehicles, other obstacles, and pedestrians. Detection of lanes and road boundaries provides the lateral position of the vehicle on the road, which is important for predicting lane departures. An extensive survey of lane-detection techniques and their characteristics is described in [23]. Stationary objects such as poles and guard-rails, as well as independently moving objects such as other vehicles, pose danger to the host vehicle. Motion [24] and binocular

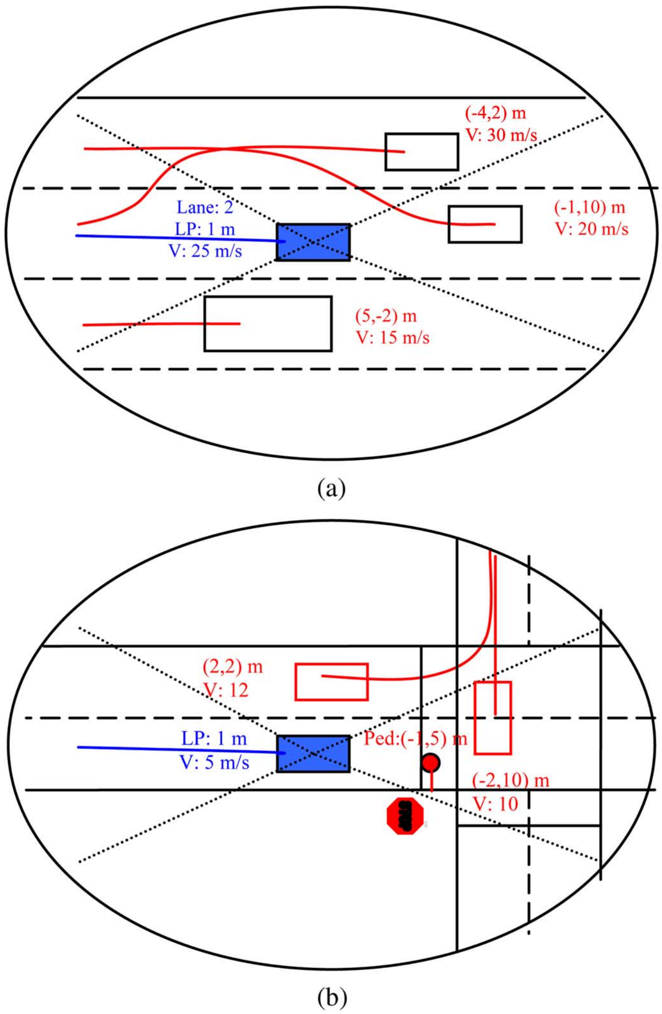


Fig. 5. Illustration of a dynamic surround map for (a) freeway and (b) city streets. (x; y): Coordinates of other objects with respect to own vehicle. V: Velocity with respect to road. LP: Lateral position of own vehicle with respect to the center of the lane.

stereo [25], [26] are often employed for detecting these objects. Pedestrians are the most vulnerable road users. Unlike the vehicles, which are large rigid objects moving fast along well-defined trajectories, the pedestrians are slow moving small thin objects having articulated motion and more complex trajectories. Surveys of pedestrian detection research are presented in [27], [36].

Many of the current driver-assistance systems deal with lanes, front objects, side objects, or back objects in isolation. Such systems are useful for specialized tasks such as lane keeping, adaptive cruise control, blind-spot monitoring, and reverse-collision avoidance. However, a complete driver-support system should be aware of all of the surroundings of the vehicle in a holistic manner. Also, it is not sufficient just to detect the objects but also estimate the danger posed by them. In [28], we have introduced the concept of DPS map that shows the nearby surroundings of the vehicle, as in Fig. 5. We have demonstrated successful generation of DPS in experimental runs on an instrumented vehicle testbed using monocular as well as binocular omni-camera systems.

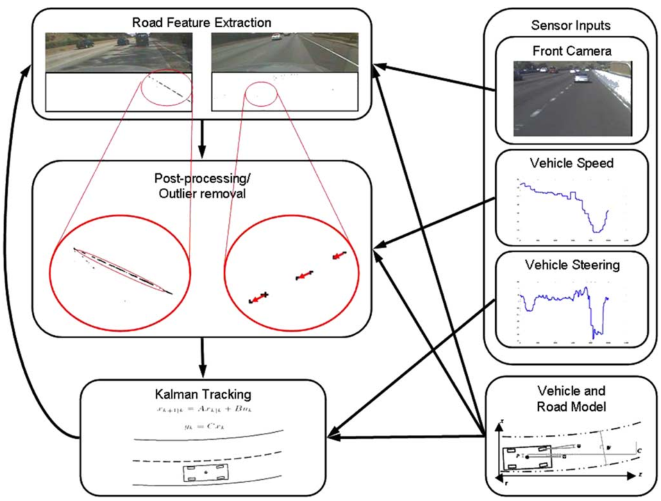


Fig. 6. Block diagram for the VioLET system.

In the following sections, we briefly describe our research on the analysis of the vehicle surroundings.

1) Lane Detection and Tracking: In [23], we have proposed a Video-based Lane-Estimation and Tracking (VioLET) system for driver assistance under a wide variety of environmental conditions such as changes in lighting, shadows, and road textures. The system block diagram is shown in Fig. 6. A constant curvature model is used for the road for good accuracy up to 30–40 m corresponding to look-ahead time of around 1 s. Steerable filters are used for robust detection and localization of multiple types of lane markings such as solid and dashed lines as well as the circular-reflector markers called "Botts-Dots," which are especially found on California highways. The potential candidates for lane markings are detected, and postprocessing is performed to cluster true candidates and reject the false ones. The lane position is tracked using Kalman filter assuming linear vehicle dynamics and vehicle controller-area-network (CAN) bus in order to improve robustness and accuracy of localization.

The performance of the system was thoroughly evaluated using a number of 65-km-long test drives on highways in Southern California. The camera for lane detection was mounted to look in front of the vehicle. The "ground truth" was generated using another camera directed downwards at the road on the side of the vehicle. The lane positions detected from the front camera were compared with the "ground truth" positions. Detailed quantitative results are described in [23]. Fig. 7 shows examples of lane detection at four typical locations in various environmental conditions. The system typically operated at around 15 frames per second on a 3-GHz Pentium PC. Additional performance gains could be made by adjusting the region of interest in the video or adjusting the resolution. In such cases, accuracy versus computational cost evaluations would need to be considered.

2) Omnidirectional Vision for Surround Capture: Omnidirectional or omni cameras which give a 360° panoramic field of view of can be useful for visualizing and analyzing the nearby surroundings of the vehicle and detecting objects such as vehicles, pedestrians, and stationary obstacles. Omni cameras or other wide-field-of-view cameras could be particularly useful

Fig. 7. Scenes from aerial views (row 1), dawn (row 2), daytime (row 3), dusk (row 4), and nighttime (row 5) from four locations on a 65-km-long route for test drive. These scenes show the environmental variability caused by road markings and surfaces, weather, and lighting.

TABLE I
COMPARISON BETWEEN VARIOUS OMNI-CAMERA CONFIGURATIONS FOR VEHICLE SURROUND CAPTURE

Monocular omni camera on roof top [29]: + 360 degree surround view covered with single camera. + Large FOV resolves ambiguities in motion estimation. - Difficult to resolve front objects with motion alone. - Separating independent motion from parallax is problematic.
Binocular omni cameras on sides[28] + Stereo view of front, monocular views of sides, driver, passenger. + Larger disparity difference between vehicle and road makes stereo discrimination easier. - Vehicles have reduced frontal area, hence smaller image size. - Rear view not covered.
Binocular omni cameras in front with short baseline [30]: + Suitable for detecting close objects such as pedestrians. - Rear view is not covered. - Stereo estimation is infeasible for distant objects.
Binocular coaxial omni cameras in back of car [26]: + Effective for monitoring blind spots behind the vehicle. + Epipolar lines along radial lines instead of concentric circles. - Smaller coverage of surroundings.

for increasing pedestrian safety, since a wide field of view could enable timely detection of a pedestrian on a collision course before it is too late to prevent a collision.

Due to the comparatively lower resolution of omni cameras, proper configuration is very important for obtaining good coverage, sensitivity, and foreground–background discrimination. Table I describes the advantages and disadvantages of various configurations that could be used for mounting omni cameras on the vehicle. Here, we briefly describe the detection of vehicles and pedestrians using three of the above configurations.

Fig. 8 shows the generation of 360° surround map using a monocular omni camera mounted on top of the vehicle. The road motion is modeled using a planar-motion model whose parameters are initially obtained using the approximate knowledge about the camera calibration and speed. Using the composition of the motion model and the omni-camera model, a warping transform is generated to compensate road motion between two consecutive frames. To account for the inaccuracy in prior knowledge of ego-motion, the parameters are iteratively updated using the spatial and temporal gradients of the motion-compensated images using Bayesian estimation. The updated model is used to compensate the motion of the road plane between the two frames, leaving features independent motion, or height above the road, uncompensated. These features are aggregated into objects and tracked over frames. Details of the approach are described in [29]. On standard hardware consisting of 2.6-GHz Pentium IV, the algorithm achieves a frame rate of approximately 4 frames per second for 320 $\times$ 240 images and 15 frames per second for 160 $\times$ 120 images. The majority of computational time is spent on estimation of motion parameters from spatio-temporal image gradients. Use of specialized hardware should improve performance.

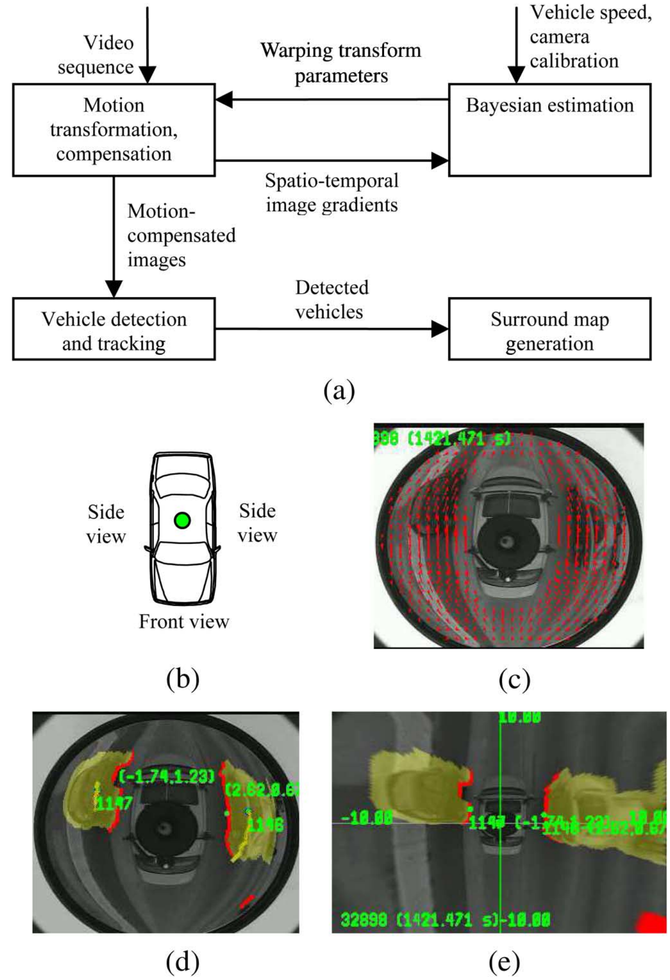


Fig. 8. Motion-based vehicle detection using omni camera. (a) System block diagram. (b) Camera configuration. (c) Video image with estimated parametric motion of road. (d) Detection and tracking of moving vehicles marked with track ID and the coordinates. (e) Surround map formed by transforming the omni image.

Fig. 9 shows the detection of pedestrians and other objects in front of the vehicle using a stereo pair of omni-directional cameras. Video sequences are obtained from a pair of omni cameras mounted on two sides of the vehicle. Camera calibration is performed offline to determine the relationship between the vehicle and pixel coordinates. Using the calibration

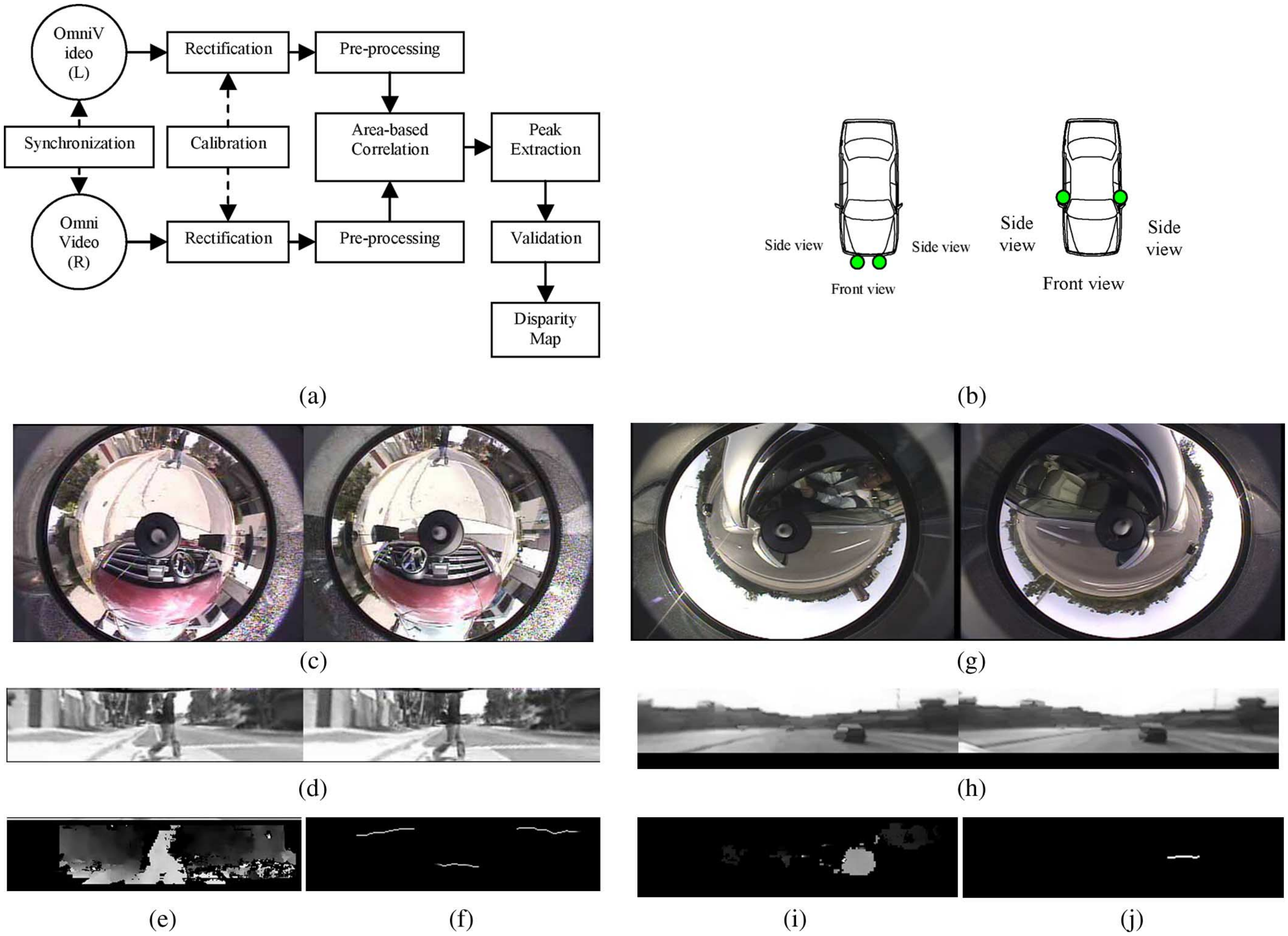


Fig. 9. Detection of pedestrian and other obstacles using a stereo pair of omni cameras. (a) System block diagram. (b) Camera configuration. (c) Original omni images. (d) Rectified virtual-perspective images. (e) Disparity image. (f) Detection of pedestrian and other obstacles. (g-j) Detection of vehicle in front of the host vehicle.

information, the images are transformed to obtain virtual-perspective views looking forward toward the road. This transformation, called rectification, simplifies the stereo geometry, making it easier to match corresponding features between the two images. Area-based correlation is then used to perform stereo matching between features. The result is a disparity map showing the displacement of features from one image to another. Based on the disparity map, the features are grouped into objects, and the distance to the objects is computed. Details of this algorithm are described in [28] and [30]. On standard 2.66-GHz Pentium-IV hardware, the processing achieves a rate of 15 frames per second, which should enable the operation to be performed in real time.

C. Looking in and Looking out of the Vehicle: Driver Intent Analysis

This section gives applications that combine the use of sensors that are looking-in as well as looking-out in order to predict driver's intended actions. Predicting the driver's intent can provide a useful tool in developing driver-assistance systems that work in harmony with the driver. System-response feedback can be better tailored to the driver's immediate situation and attentive state. We will discuss two examples of predicting driver intent and discuss how they can be used to enhance vehicle safety. These two examples are a system for predicting the driver's intent to change lanes and a predictive brake-assistance system.

1) Lane-Change-Intent Analysis: In our first example, we will explore how sensors looking in and looking out of the vehicle can be used to predict the driver's intent to change lanes. In many previous approaches, only sensors detecting the vehicles positioning on the road and internal-vehicle sensors monitoring speed and steering were used to predict lane trajectory [31], [32]. However, lane trajectory can be different than actual intent to change lanes. This is evident in situations where the driver is drifting toward the lane boundary on a curve. In this type of situation, the driver might be purposefully changing lanes, or the driver might be unaware of the road curvature and accidentally depart the lane. Only information about the driver inside the vehicle carries the potential to disambiguate these types of situations. As we will show, overall classifier accuracy can be improved when we fuse information from visual sensors viewing both the inside and outside of the vehicle.

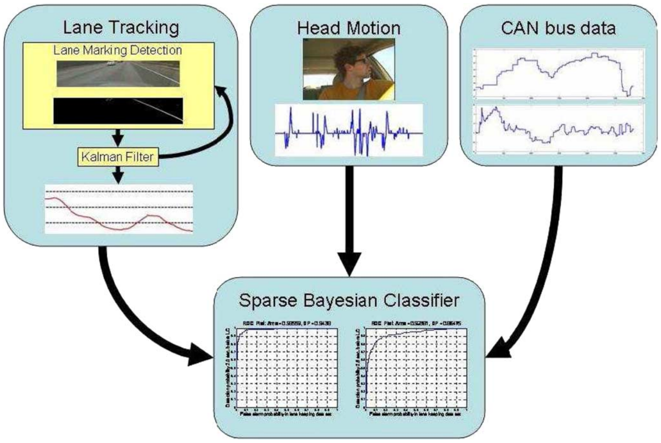


Fig. 10. Block diagram for inference of lane-change intent using information from lane tracking, head motion, and the CAN bus data.

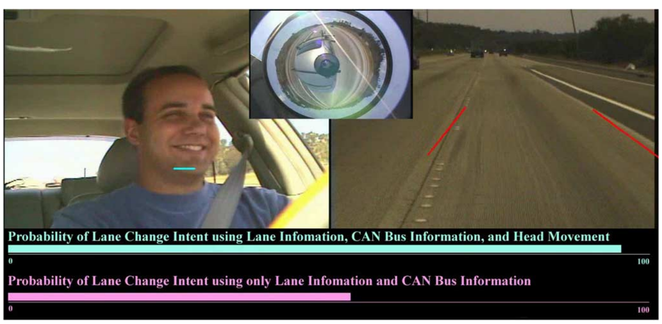


Fig. 11. Example of detection of lane-change intent. The top bar shows the estimated probability of lane change using all the three sources, whereas the bottom bar shows the probability using only lane tracking and CAN bus data. The lane change is detected earlier when head-motion information is used.

The system is composed of four key components, as shown in Fig. 10. These components are

1) lane-position-tracking system that determines the lateral position of the vehicle in the lane at any given time;
2) driver head-motion-estimation module that uses facial features detected from a camera in order to estimate the approximate motion of driver's head;
3) vehicle-parameter-collection system that gives parameters such as vehicle speed, yaw rate, and acceleration;
4) lane-change-intent classifier based on sparse Bayesian learning that combines the features from the above components in order to determine the probability of lane change at any given time.

A more detailed description of the system is given in [33].

Fig. 11 shows an example of lane-change-intent detection. The top bar shows the estimated probability of lane change using lane tracking, vehicle dynamics from the CAN bus, as well as head motion. The bottom bar is derived without using head motion. It is observed that the use of head motion gives an advantage of 0.5 s in detecting lane-change intent, which is critical for preventing accidents. This is shown in Fig. 12, in which receiver-operator-characteristic (ROC) curves compare classifier performance for a classifier predicting 2.5 s into the future without head information and a classifier predicting 3.0 s into the future [33].

2) Predictive-Brake-Assistance System: In our second example of how driver-assistance systems can be improved by examining data from both inside and outside the vehicle, we will look at a predictive-brake-assistance system. One of the most important advantages of systems that predict driver behavior and intent is the potential to improve the user-experience. Working cooperatively with and adapting to the driver, rather than letting the driver adapt to the system, can increase the user's acceptance of the system.

As an example of this, we created a predictive-braking-assistance system that predicts both the need for braking as well as the driver's intent to perform the braking action. The system, after having detected a need to brake, has the potential to warn or otherwise intervene in the situation, if the driver appears to be unaware of the severity of the situation. This is performed by decomposing the probability that the system should warn or alert the driver or otherwise intervene in the situation into the probability describing the need for braking, given the surround situation and the probability of describing the intent of the driver to perform the braking action. This is expressed mathematically, as follows, using Bayes rule and assuming conditional independence:

$$P(C|B_o, B_i) = k \cdot P(B_o|C)P(B_i|C). \tag{1}$$

In this equation, C represents the situation criticality or the need for the system to intervene, B_i represents the intent for the driver to brake, B_o represents the need for braking according to the vehicle's environment, and k is a scale factor based on the prior $P(C)$ and marginal joint probability $P(B_i, B_o)$. Graphically, this can also be shown as a Bayesian network, as in Fig. 13. $P(B_i|C)$ is computed using the signals listed in the driver-behavior block of the diagram, while $P(B_o|C)$ is computed using the signals listed in the vehicle-environment block of the diagram. Sparse Bayesian learning was used to estimate the probability density functions based on the observations. The complete description of this system is given in [35].

Driver behavior is trained and predicted using a camera viewing the driver's foot and pedal movements and a camera for the driver's head movements. The need for braking based on the situation is trained and predicted from information gathered from outside the vehicle, including CAN bus data and LASER-RADAR information (distance and relative speed of the lead vehicle). Using sparse Bayesian learning to train the probabilities on the right-hand side in (1), we can construct a compact classifier with the desirable properties of good generalization and robustness to over fitting. The performance of the driver-braking-intent classifier is shown in Fig. 14. Images depicting the system performance are shown in Fig. 15. The driver-intent classification for both brake assistance and lane-change intent run with minimal computational load because of the sparseness enforced by the learning algorithm. The actual operating time is more dependent on the speed at which the lane tracking and other cue extraction systems run.

V. Concluding Remarks

In the development of a real-time, robust dynamic context-capture system for an automobile, computer vision and

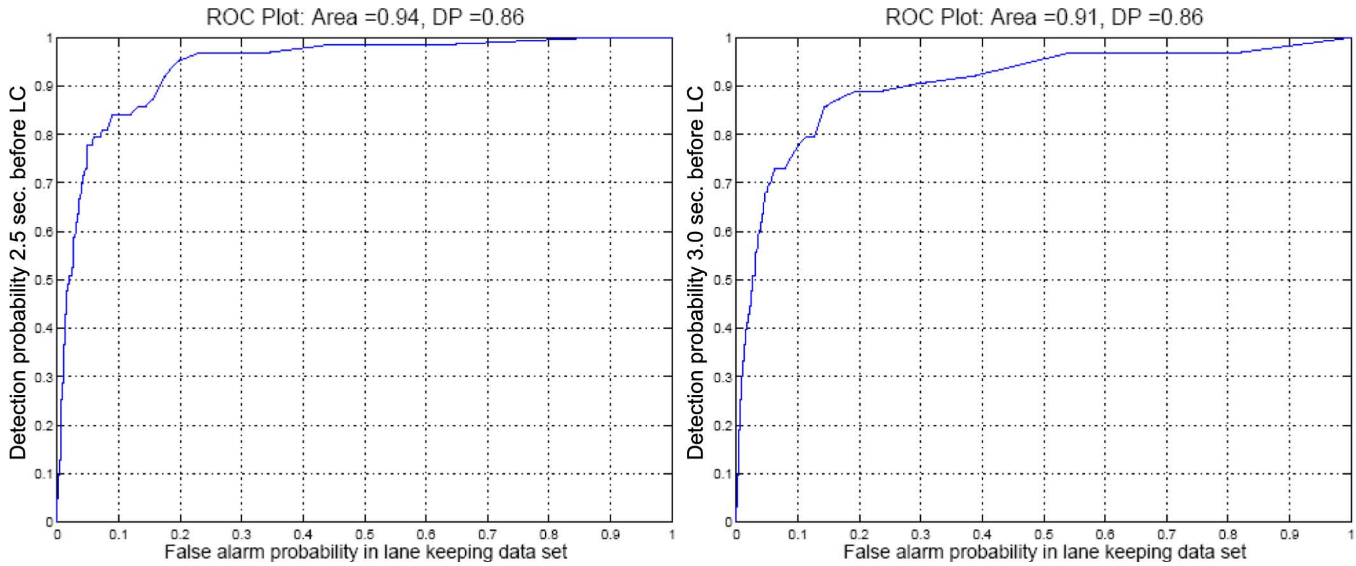


Fig. 12. ROC curves showing the classifier performance at 2.5 s before the lane change without using head information (left) and 3.0 s before the lane change using head information (right).

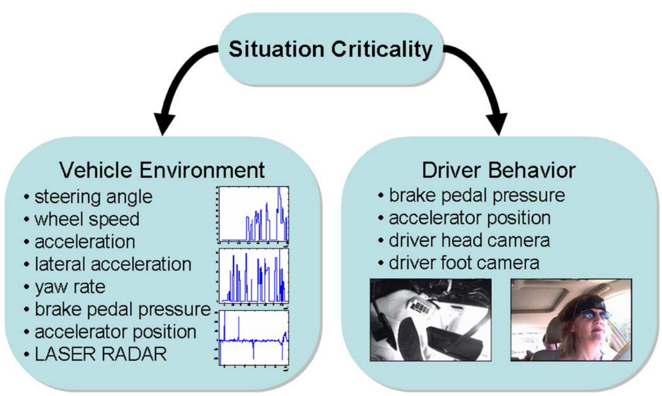


Fig. 13. Bayesian network for modeling the criticality of the situation or the systems' need for intervention.

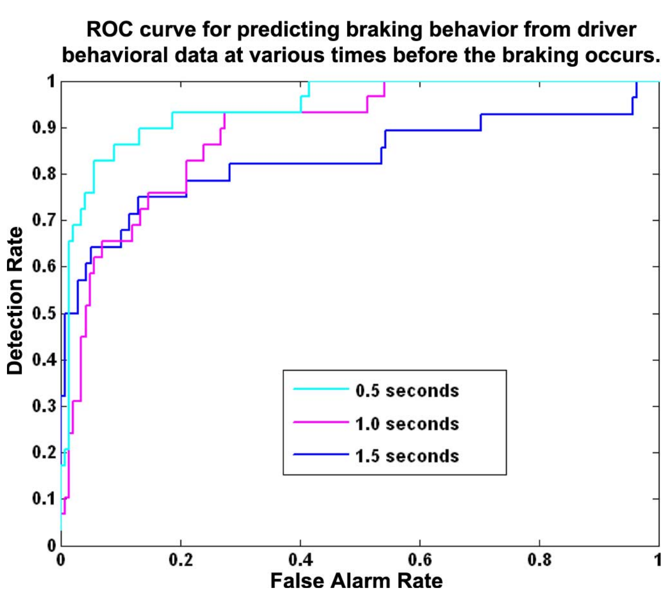


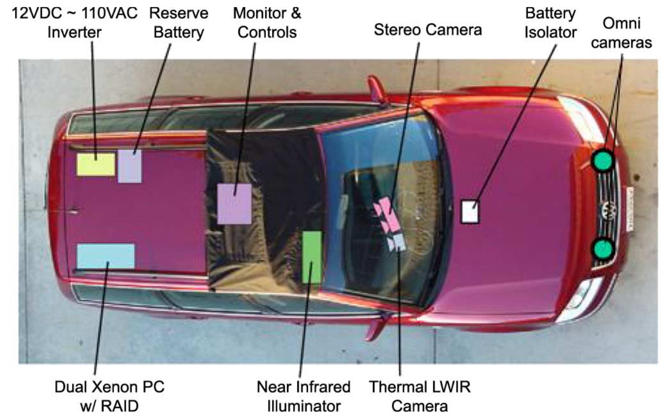
Fig. 14. ROC curve for predicting the driver's intent to brake. The performance is shown at 0.5, 1.0, and 1.5 s before the braking action.

machine-learning techniques play important roles. In this paper, we presented a motivation and experimental support for developing vision systems for looking in and looking out of a vehicle. An "active," multimodal sensory approach for "dynamic context capture and situational awareness" using cameras, radars, audio, etc., allows for the establishment of representations of the state of the environment, the vehicle, and the driver with accurate dynamic uncertainty characterization. At the individual component level, novel techniques for interpreting the driver's surroundings, driver behavior, and the driver's intended actions were demonstrated. These comprise the necessary components for a holistic safety system centered on looking in and out of the car. Not only does each component of this system represent a significant contribution, but also, the combination of these systems and how they are integrated provides advancement in automobile safety. Looking at the fully integrated system can give expected performance and safety margins for a complete driver-assistance system: not just individual components. It is believed that successful integration of such powerful sensory suites in a human-centric decision logic framework will have a significant impact on the safety of new generations of automobiles and telematics devices used for in-car communication, information access, business transactions, and entertainment.

APPENDIX
LABORATORY FOR INTELLIGENT, SAFE AUTOMOBILES (LISA) TESTBEDS

To provide adaptable experimental testbeds for evaluating the performance of various sensing modalities and their combination, two test environments based upon a Volkswagen Passat vehicle [Laboratory for Intelligent, Safe Automobiles-P (LISA-P)] and a Nissan Infinity Q-45 vehicle (LISA-Q) were outfitted with a computer and a multitude of cameras and acquisition systems. Of principal importance in the hardware specification and software architecture was the ability to capture

Fig. 15. Images depicting the performance of the predictive brake-assistance system. The probabilities of the driver's intent to brake, the situational need for braking, and the need for system intervention are shown as bars expanding from left to right proportionaly to the predicted probability.



(a)

(b)

Fig. 16. (a) LISA-P testbed used for occupant- and driver-posture analysis and pedestrian detection. (b) Images obtained from various cameras in the testbed including the visible light images from different perspectives, thermal image (bottom-mid), and stereo disparity map (bottom-right).

and process data from all the sensor subsystems simultaneously and to provide facilities for algorithm development and offline testing. A short description of these testbeds is provided below with references to relevant papers for details.

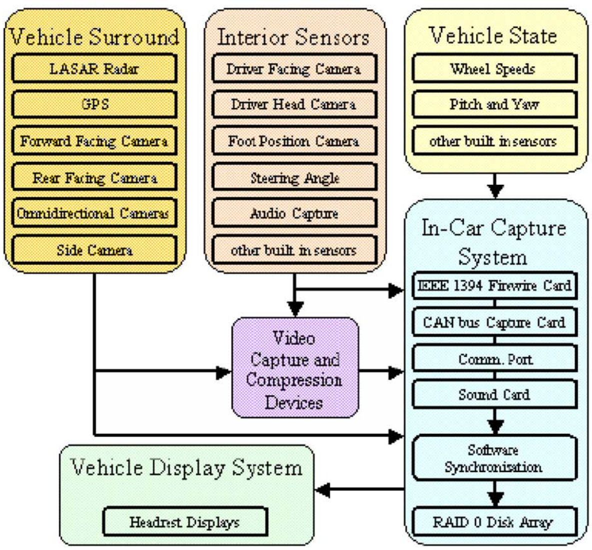
A. LISA-P: Occupant- and Driver-Posture Analysis and Pedestrian Detection

The LISA-P testbed shown in Fig. 17(a) [20] is designed for collecting and processing large amounts of synchronized data from a number of different sensors for monitoring the driver's state and surroundings. The sensory modules used in this testbed include the following.

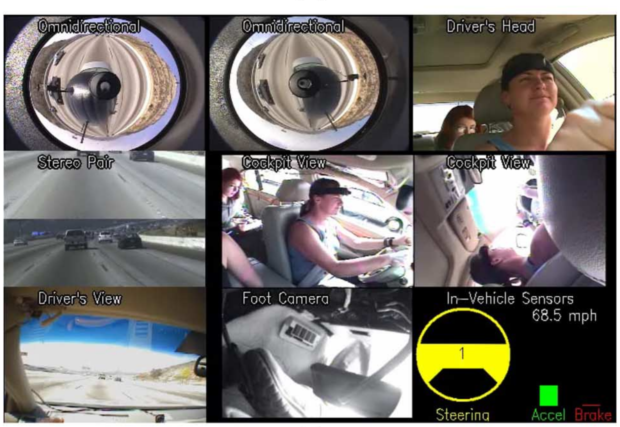
1) A stereo system which provides 2-1/2-D stereo disparity maps is used for estimating the distance of the occupant.
2) A miniature 2-D thermal long-wavelength infrared sensor is mounted on the dashboard to observe the face of the occupant. This device provides video response in the LWIR spectrum (7–14 um).
3) An array of four color CCD cameras provides images used for obtaining 3-D voxel reconstruction through SFS.
4) A pair of omnidirectional cameras in front of the vehicle gives panoramic images used for detection of pedestrians and nearby vehicles.
5) A pair of SICK LASER scanners on two sides of the car can be used for detecting nearby objects and determining accurate distance to them.

The placement of the sensors is shown in Fig. 16(a). These sensors are supported by a synchronized video-stream-capturing hardware, high-volume high-throughput storage, and a powerful computing platform. The computing platform allows for a good deal of processing to be done in real time as well as store data for offline processing. A detailed description of the testbed is provided in [20].

(a)



(b)



(c)

Fig. 17. (a) LISA-Q testbed used for vehicle surroundings, lane tracking, and driver-support system. (b) Information flow between the subsystems in the testbed. (c) Outputs from various cameras and sensors from the vehicle interior as well as the surroundings.

B. LISA-Q: Vehicle Surround, Lane Tracking, and Driver-Support System

The LISA-Q intelligent test bed shown in Fig. 17 [34] is designed to obtain complete coverage of the vehicle surroundings, the vehicle interior, the state of the vehicle for extended periods of time from a variety of modular sensing systems, and processing of the data in order to be fed back to the human occupant. Sensor systems include rectilinear cameras, wide field-of-view camera systems, GPS and navigation systems, and the data from internal-automobile vehicle-state sensors. The system contains an array of computers that serve for data collection as well as real-time processing of information. The hardware capabilities of the LISA-Q intelligent vehicle include

1) eight NTSC hardware video compressors for simultaneous capture;
2) CAN interface for acquiring steering angle, pedals, yaw rate, and other vehicle information;
3) built-in five-beam forward-looking LASER-RADAR range finder;
4) wide area augmentation system (WAAS)-enabled GPS;
5) integration into car audio and after-market video displays for feedback and alerts.

Fig. 17(c) shows the outputs from various cameras and other sensors in the vehicle interior as well as the surroundings.

ACKNOWLEDGMENT

The authors would like to thank their colleagues at the Computer Vision and Robotics Research Laboratory for their cooperation, assistance, and contributions. They are especially thankful to their industrial collaborators, who provided them with invaluable guidance and support to provide a "reality-check" for their research. It is indeed their pleasure to acknowledge the positive impact of their interactions with such collaborators, including Dr. Y. Akatsu, Dr. E. Boer, Dr. A. Jameel, Dr. A. Kinoshita, Dr. S. Kitazaki, Dr. C. Pribe, Dr. K. Schaff, Dr. A. Stoschek, Dr. H. Takahashi, and their research teams.

REFERENCES

[1] M. Peden, R. Scurfield, D. Sleet, D. Mohan, A. A. Hyder, E. Jarawan, C. Mathers. (2004, Apr.). "World report on road traffic injury prevention: Summary" World Health Organization, Geneva, Switzerland, ISBN 92 4 156260. [Online]. Available: http://www.who.int/world-health-day/2004/infomaterials/world_report/en

[2] "Traffic safety facts 2003: A compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system," U.S. Dept. Transp., Nat. Highway Traffic Safety Admin., Washington, DC, Jan. 2005. Tech. Rep. DOT HS 809 775. [Online]. Available: http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSFAnn/2003HTMLTSF/TSF2003.HTM

[3] E. D. Dickmanns, B. Mysliwetz, and T. Christians, "An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles," *IEEE Trans. Syst., Man Cybern.*, vol. 20, no. 6, pp. 1273–1284, Nov./Dec. 1990.

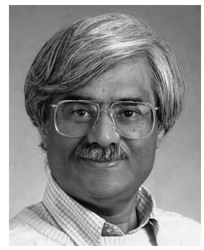
[4] D. Pomerleau and T. Jochem, "Rapidly adapting machine vision for automated vehicle steering," *IEEE Expert—Special Issue Intelligent System and Their Applications*, vol. 11, no. 2, pp. 19–27, Apr. 19–27, 1996.

[5] U. Franke, D. Gavrila, S. Gorzig, F. Lindner, F. Puetzold, and C. Wohler, "Autonomous driving goes downtown," *IEEE Intell. Syst.*, vol. 13, no. 6, pp. 40–48, Nov./Dec. 1998.

[6] M. Bertozzi, A. Broggi, and A. Fascioli, "Vision-based intelligent vehicles: State of the art and perspectives," *Robot. Auton. Syst.*, vol. 32, no. 1, pp. 1–16, Jul. 2000.

[7] 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.

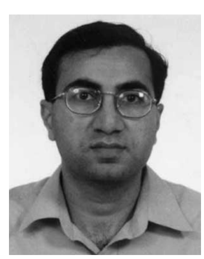
[8] F. Heimes and H.-H. Nagel, "Towards active machine-vision-based driver assistance for urban areas," *Int. J. Comput. Vis.*, vol. 50, no. 1, pp. 5–34, Oct. 2002.
[9] W. Enkelmann, "Video-based driver assistance—From basic functions to applications," *Int. J. Comput. Vis.*, vol. 45, no. 3, pp. 201–221, Dec. 2001.
[10] R. Grace, V. E. Byrne, J. M. Legrand, D. J. Gricourt, R. K. Davis, J. J. Staszewski, and B. Carnahan, "A machine vision based drowsy driver detection system for heavy vehicles," in *Proc. Occular Meas. Driver Alertness Conf.*, Apr. 1999, pp. 75–86. FHWA-MC-99-136.
[11] P. Green, *Driver Distraction, Telematics Design, and Workload Managers: Safety Issues and Solutions*. SAE Paper 2004-21-0022.
[12] V. L. Neale, T. A. Dingus, S. G. Klauer, J. Sudweeks, and M. Goodman, "An overview of the 100-car naturalistic driving study and findings," in *Proc. 19th Int. Tech. Conf. ESV*, Washington, DC, Jun. 2005, pp. 19:1–10.
[13] J. D. Lee and D. L. Strayer, "Preface to the special section on driver distraction," *Hum. Factors: Journal Human Factors and Ergonomics Society*, vol. 46, no. 4, pp. 583–586, Winter 2004.
[14] L. M. Trick, J. T. Enns, J. Mills, and J. Vavrik, "Paying attention behind the wheel: A framework for studying the role of selective attention in driving," *Theor. Issues Ergon. Sci.*, vol. 5, no. 5, pp. 385–424, 2004.
[15] J. Levy, H. Pashler, and E. Boer, "Central interference in driving: Is there any stopping the psychological refractory period?" *Psychological Sci.,* vol.17, no. 3, pp. 228–235, Mar. 2006.
[16] M. A. Goodrich and E. R. Boer, "Model-based human-centered task automation: A case study in ACC system design," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 33, no. 3, pp. 325–336, May 2003.
[17] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Hum. Factors*, vol. 39, no. 2, pp. 230–253, Jun. 1997.
[18] J. D. Lee and N. Moray, "Trust, control strategies and allocation of function in human machine systems," *Ergonomics*, vol. 35, no. 10, pp. 1243–1270, 1992.
[19] K. Huang, M. M. Trivedi, and T. Gandhi, "Driver's view and vehicle surround estimation using omnidirectional video stream," in *Proc. IEEE Intell. Vehicles Symp.*, Columbus, OH, Jun. 2003, pp. 444–449.
[20] M. M. Trivedi, S. Y. Cheng, E. M. C. Childers, and S. J. Krotosky, "Occupant posture analysis with stereo and thermal infrared video: Algorithms and experimental evaluation," *IEEE Trans. Veh. Technol.*, vol. 53, no. 6, pp. 1698–1712, Nov. 2004.
[21] A. Eleftheriadis and A. Jacquin, "Face location detection for model-assisted rate control in H.261-compatible coding of video," *Signal Process.*, vol. 7, no. 4–6, pp. 435–455, Nov. 1995.
[22] A. Laurentini, "How many 2-D silhouettes does it take to reconstruct a 3-D object?" *Comput. Vis. Image Underst.,* vol.67, no. 1, pp. 81–89, Jul. 1997.
[23] J. McCall and M. M. Trivedi, "Video based lane estimation and tracking for driver assistance: Survey, algorithms, and evaluation," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 20–37, Mar. 2006.
[24] W. Krüger, "Robust real-time ground plane motion compensation from a moving vehicle," *Mach. Vis. Appl.*, vol. 11, no. 4, pp. 203–212, Dec. 1999.
[25] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Trans. Image Process.*, vol. 7, no. 1, pp. 62–81, Jan. 1998.
[26] L. Matuszyk, A. Zelinsky, L. Nilsson, and M. Rilbe, "Stereo panoramic vision for monitoring vehicle blind-spots," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2004, pp. 31–36.
[27] D. M. Gavrila, "Sensor based pedestrian protection," *IEEE Intell. Syst.*, vol. 16, no. 6, pp. 77–81, Nov./Dec. 2001.
[28] T. Gandhi and M. M. Trivedi, "Vehicle surround capture: Survey of techniques and a novel omni video based approach for dynamic panoramic surround maps," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 3, pp. 293–308, Sep. 2006.
[29] ——, "Parametric ego-motion estimation for vehicle surround analysis using omni-directional camera," *Mach. Vis. Appl.*, vol. 16, no. 2, pp. 85–95, Feb. 2005.
[30] ——, "Vehicle mounted wide FOV stereo for traffic and pedestrian detection," in *Proc. Int. Conf. Image Process.*, Sep. 2005, pp. II-121–II-124.
[31] N. Kuge, T. Yamamura, and O. Shimoyama, "A driver behavior recognition method based on a driver model framework," *SAE Trans.*, vol. 109, no. 6, pp. 469–476, 2000.
[32] D. Salvucci, "Modeling driver behavior in a cognitive architecture," *Hum. Factors*, vol. 48, no. 2, pp. 362–380, 2006.
[33] J. McCall, D. Wipf, M. M. Trivedi, and B. Rao, "Lane change intent analysis using robust operators and sparse Bayesian learning," *IEEE Trans. Intell. Transp. Syst.*, to be published.
[34] J. McCall, O. Achler, and M. M. Trivedi, "Design of an instrumented vehicle testbed for developing human centered driver support system," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2004, pp. 483–488.
[35] J. McCall and M. M. Trivedi, "Human behavior based predictive brake assistance," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 8–12, 2006, pp. 8–12.
[36] T. Gandhi and M. M. Trivedi, "Pedestrian collision avoidance systems: A survey of computer vision based recent studies," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Sep. 2006, pp. 976–981.



Mohan Manubhai Trivedi received the B.E. degree (with honors) from Birla Institute of Technology and Science, Pilani, India, and the Ph.D. degree from Utah State University, Logan.

He is a Professor of Electrical and Computer Engineering and the founding Director of the Computer Vision and Robotics Research Laboratory, University of California, San Diego (UCSD), La Jolla. In partnership with research laboratories of major automobile companies and the UC Discovery Program, he has established the Laboratory for Intelligent and Safe Automobiles, UCSD, to pursue a multidisciplinary research agenda. He has an extensive publication record with more than 300 archival papers and 12 edited volumes. He served as the Editor-in-Chief of the *Machine Vision and Applications Journal* (1996-2004) and on the editorial boards of several journals. He regularly serves as a Consultant to industry and government agencies in the USA and abroad. His research interests include intelligent systems, computer vision, intelligent ("smart") environments, intelligent vehicles and transportation systems, and human-machine interfaces areas.

Prof. Trivedi is currently an Associate Editor for the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS. He served as a Program Chair for the IEEE International Intelligent Vehicles IV 2006 Symposium. He serves on the executive committees of the University of California Digital Media Innovation Program and of the California Institute for Telecommunication and Information Technologies [Cal-IT2] as the leader of the Intelligent Transportation and Telematics Layer at UCSD. He is the recipient of the Distinguished Alumnus Award from Utah State University, the Pioneer Award (Technical Activities), and the Meritorious Service Award from the IEEE Computer Society.



Tarak Gandhi (S'93–M'99) received the Bachelor of Technology degree in computer science and engineering from the Indian Institute of Technology, Bombay, India, and the M.S. and Ph.D. degrees in computer science and engineering from the Pennsylvania State University, specializing in computer vision.

He worked at Adept Technology Inc., designing algorithms for robotic systems. He is currently an Assistant Project Scientist with the California Institute for Telecommunications and Information Technology, University of California at San Diego, La Jolla. He is one of the key members of the Computer Vision and Robotics Research (CVRR) Laboratory at the university. His interests include computer vision, motion analysis, image processing, robotics, target detection, and pattern recognition. He is working on projects involving intelligent driver assistance, motion-based event detection, traffic-flow analysis, and structural health monitoring of bridges.

Joel McCall received the B.S. degree in electrical engineering and computer science from the University of California, Berkeley, in 1999, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, San Diego, La Jolla, in 2003 and 2006, respectively. His Ph.D. research focused on human-centered driver-assistance and safety systems.

He is currently working at Microsoft Corporation, Redmond, WA.