# *Looking-In* and *Looking-Out* of a Vehicle: Selected Investigations in Computer Vision based Enhanced Vehicle Safety

Mohan M. Trivedi, Tarak Gandhi, and Joel McCall
Computer Vision and Robotics Research Laboratory
University of California San Diego, La Jolla, CA, USA
{mtrivedi,tgandhi,jmccall}@ucsd.edu

*Abstract*— **This paper presents an overview of investigations into the role of computer vision technology in developing safer automobiles. We consider vision systems which can not only look out of the vehicle to detect and track roads, avoid hitting obstacles or pedestrian, 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, driver, vehicle, and vehicle surround. 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. The paper includes 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*— **Intelligent vehicles, Driver support systems, Real-time machine vision systems, Active safety.**

## I. INTRODUCTION: RESEARCH MOTIVATION AND SCOPE

TRAFFIC related accidents are recognized as a serious and growing problem with global dimensions. According to a recent study by World Health Organization (WHO) mentions that annually over 1.2 million fatalities and over 20 million serious injuries occur in the world [1]. Enhancement of traffic safety is pursued with as a high priority item not only by various government agencies such as National Transportation Safety Administration (NTHSA) [2], but also by most major automobile manufactures. University based researchers are also contributing to this important mission. 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.

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][4][5] Such efforts helped to demonstrate the power of camera based systems to support real-time control of vehicles. These and other earlier studies were not focused on the enhancement of automobile safety as the primary objective. However, a new trend started emerging in the late nineties, where research in computer vision got focused on enhancement of safety of automobiles [6][7][8][9][10][11]. 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.

The research pursued in our Laboratory for Intelligent and Safe Automobiles (LISA), considers development of vision based systems with 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. Basically, these investigations have considered three types of viewing perspectives for the cameras.

*(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, and

(3) *Simultaneous Looking in and Looking out of the vehicle:* to correlate the visual contextual information of vehicle interior and vehicle exterior, driver's behavior and intent can be systematically investigated which can lead to derivation of useful feedback mechanisms for managing driver distraction.

A sensor system that is capable of maintaining dynamic representations of the external world surrounding the vehicle, the state of the vehicle itself, and of the driver. Dynamic context capture for the Human-Centered Intelligent Driving Support System (HC-IDSS) requires analysis of multimodal sensory information and their fusion at multiple levels of abstraction. To develop a robust dynamic context capture system, computer vision and machine learning techniques play an important role. In our research we have pursed development and evaluation of an active, multi-modal sensory approach for "*dynamic context capture and situational awareness*" using cameras, radars, audio, etc for establishing representations of the state of the environment, the vehicle, and the driver with accurate dynamic uncertainty characterization.

Manuscript received October 2, 2005. This work was supported in part by theUC Discovery Grant's Digital Media Innovations Program, Nissan Motor Corporation, and the Volkswagen Corporation.

Figure 1 Framework for Multifunctional, "active" computer vision based dynamic context capture system.

For dynamic context capture, vehicle-based state-of-the-art integrated sensor suites are pursued. We propose a hierarchical structure, as shown in Figure 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 multi-modal sensor network.

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

- *What sets of sensors are robust enough under a wide variety of environmental conditions?*
- *What contexts can support a sufficiently complete representation of the environment, the vehicle state, and the driver state?*
- *What is the best computational framework to extract contexts from sensor networks that is compatible with human perception?*
- *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?*
- *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?*

## II. LISA TESTBEDS

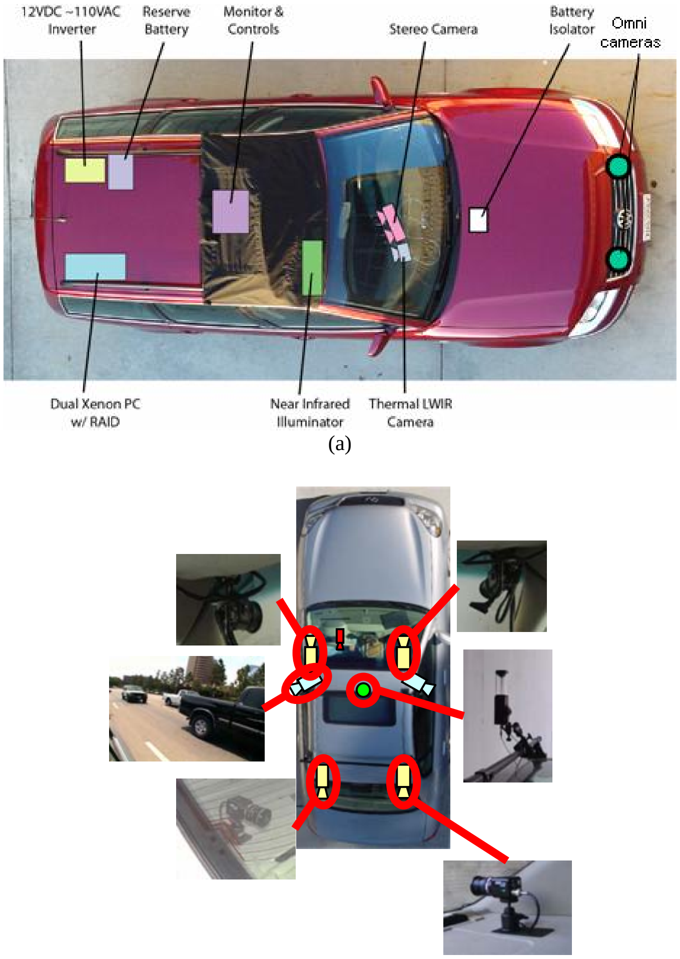
To provide adaptable experimental testbeds for evaluating the performance of various sensing modalities and their combination, two test environments, based upon a Volkswagen Passat vehicle [Laboratory for Intelligent, Safe Automobiles-P (LISA-P)] vehicle, and a Nissan Infinity Q-45 vehicle (LISA-Q) were 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 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 test bed shown in Figure 2 (a) is designed for collecting and processing large amount of synchronized data from a number of different sensors, especially for monitoring driver's state. Various sensory and computing modules used in this testbed include:

(1) A trinocular stereo system from Pt-Grey Research, 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, Raytheon model 2000 AS is mounted on the dashboard to observe the face of the occupant. This device provides video response in the LWIR spectrum (7–14 m).
(3) An array of four color CCD cameras providing images used for obtaining 3-D voxel reconstruction through shape-from-silhouette (SFS).
(4) A pair of omnidirectional cameras in front of the vehicle giving panoramic images used for detection of pedestrians and nearby vehicles.
(5) A pair of SICK LIDAR sensors on two sides of the car which can be used for detecting nearby objects and determining accurate distance to them.

The placement of the sensors is shown in Figs. 2 (a). These sensors are supported by a synchronized video-stream-capturing hardware, and high-volume storage. The computing platform consists of a commercial Xeon PC with a high-throughput disk subsystem for streaming video data. The computing platform allows for a good deal of processing to be done in real time as well as store data for off-line processing. LISA-P is outfitted with a power inverter to supply 120 volts AC power. A detailed description of the testbed is provided in [8].



(a)

(b)

Fig. 2 (a) LISA-P testbed used for occupant and driver posture analysis, and pedestrian detection (b) LISA-Q testbed used for vehicle surround, lane tracking, and driver support system.

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

The LISA-Q intelligent test bed shown in Figure 2 (b) [12] is designed as a system capable of collecting large amounts of data from a variety of modular sensing systems and processing that 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 key capabilities of the LISA-Q intelligent vehicle include:

(1) Eight NTSC hardware video compressors for simultaneous capture.
(2) Controller-Area-Network (CAN) interface for acquiring steering angle, pedals, yaw rate, and other vehicle information.
(3) Built-in 5-beam forward looking LASER RADAR range finder.
(4) WAAS enabled GPS.
(5) Integration into car audio and after-market video displays for feedback and alerts.

Detailed information about this test bed is described in [12].

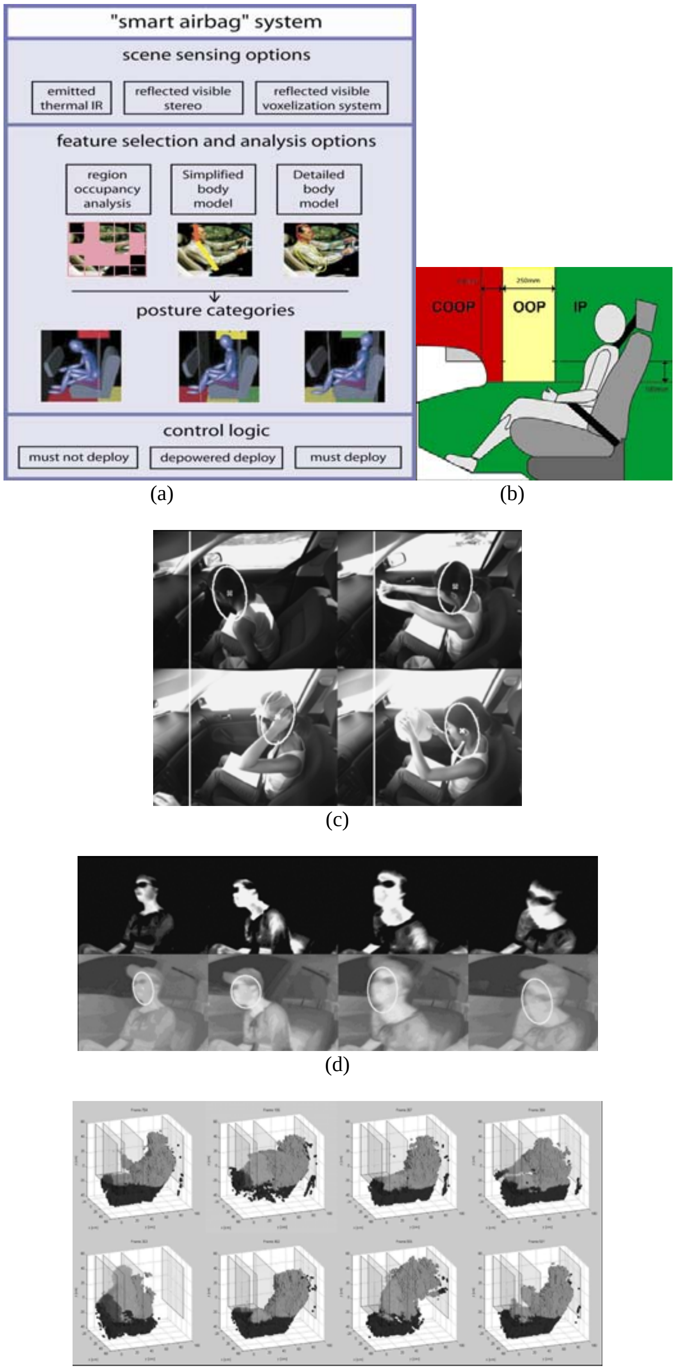
## III. VISION SYSTEMS FOR ENHANCED SAFETY: ILLUSTRATIONS

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

The objective of this research is 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 three parts: 1) real-time scene sensing; 2) feature selection; and 3) body size, posture, and movement analysis, followed by decision logic for various levels of airbag deployment as shown in Figure 3 (a). We propose using video cameras for their unobtrusiveness and potential for other purposes beyond “smart airbags.”

For scene sensing, we consider emitted LWIR imaging and stereo depth imaging. For feature selection and analysis, we consider both simple region occupancy features to detailed human body model pose estimation. Using stereo or multi-camera systems with high-level human body modeling would also provide information useful for other applications with minimal extra effort. High-quality input data and detailed analysis of body pose can also be used to enhance safety by analyzing driver alertness and could also be used to build intelligent interfaces to different in-car devices, such as the mobile phone or radio.

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. A diagram of the in-position (IP), out-of-position (OOP), and critically out-of-position (COOP) areas in the passenger seat is shown in Figure 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, and knees on the dashboard or the edge of the seat.



(a) (b)

(c)

(d)

(e)

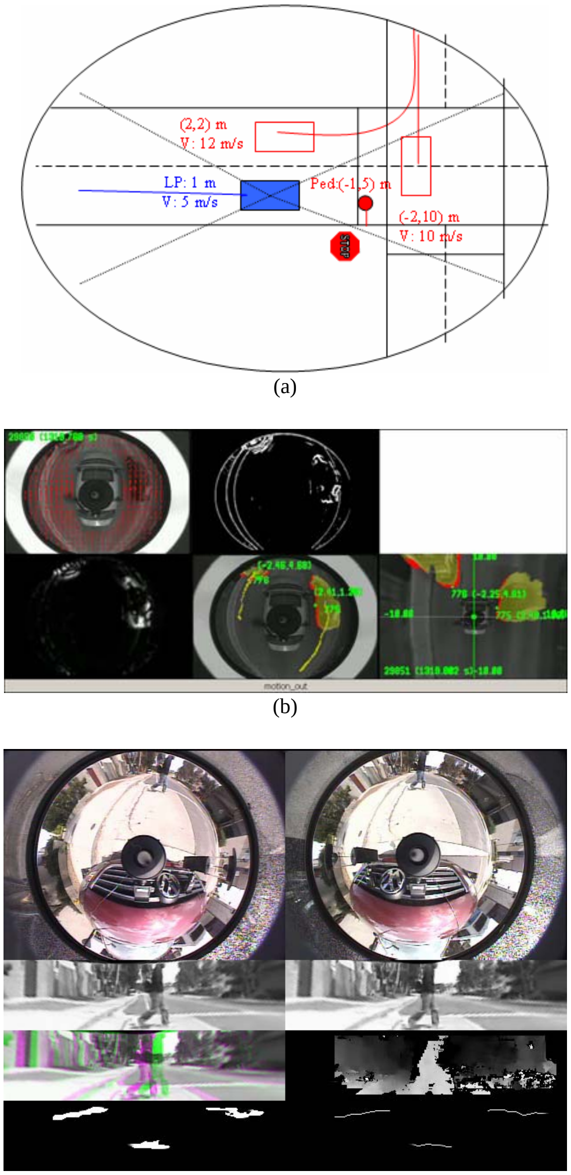
Figure 3 Smart airbag system that makes airbag deployment decision based on position of the occupant. (a) Block diagram (b) Areas for In-position (IP), Out-of-position (OOP) and Critically out-of-position (COOP) regions (c) Head detection using stereo cameras (d) Head detection using thermal cameras (e) 3-D voxel reconstruction of upper part of body using multiple cameras.

### B. Looking-out of the vehicle: Dynamic Panoramic Surround

Awareness of what surrounds a vehicle directly affects the safe driving and maneuvering of an automobile. Surround information or maps can help in studies of driver behavior as well as provide critical input in the development of effective driver assistance systems. Omnidirectional cameras which give a panoramic view of the surroundings such as Figure 4 (a) can be useful for visualizing and analyzing the nearby surroundings of the vehicle. In [13], we have introduced the concept of Dynamic Panoramic Surround (DPS) map that shows the nearby surroundings of the vehicle, and detects the objects of importance on the road. We have demonstrated successful generation of DPS in experimental runs on an instrumented vehicle testbed using monocular as well as binocular omni camera systems. These experiments prove the basic feasibility and show promise of omni video based DPS capture algorithm to provide useful semantic descriptors of the state of moving vehicles and obstacles in the vicinity of a vehicle.

Figure 4 (b) shows the generation of 360 degree surround map using a monocular omni camera mounted on top of the vehicle. The motion of the road is modeled using a parametric planar motion model whose parameters are estimated by optimally combining camera calibration, vehicle speed information from CAN bus, and the motion of features on the ground. The features that do not satisfy the model are separated as outliers. Using the model, the road motion between two frames is compensated, and normalized image difference is used to detect objects that are above the road or have independent motion. The omnidirectional image with object positions is unwarped to give the surround map. Details of the approach are described in [14].

Figure 4 (c) shows the detection of pedestrians 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 off-line to determine the relationship between the vehicle and pixel coordinates. Using the calibration information, the images are transformed to obtain virtual perspective views looking forward towards 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 distance to the objects is computed. Details of this algorithm are described in [15].



(a)

(b)

(c)

Figure 4 (a) Illustration of a dynamic surround map. (x; y): Coordinates of other objects w.r.t. own vehicle. V : Velocity w.r.t. road. LP: Lateral position of own vehicle w.r.t center of the lane (b) Generation of surround map using motion-based detection using a single omni camera on top of the vehicle (c) Pedestrian detection using stereo pair of omni cameras.

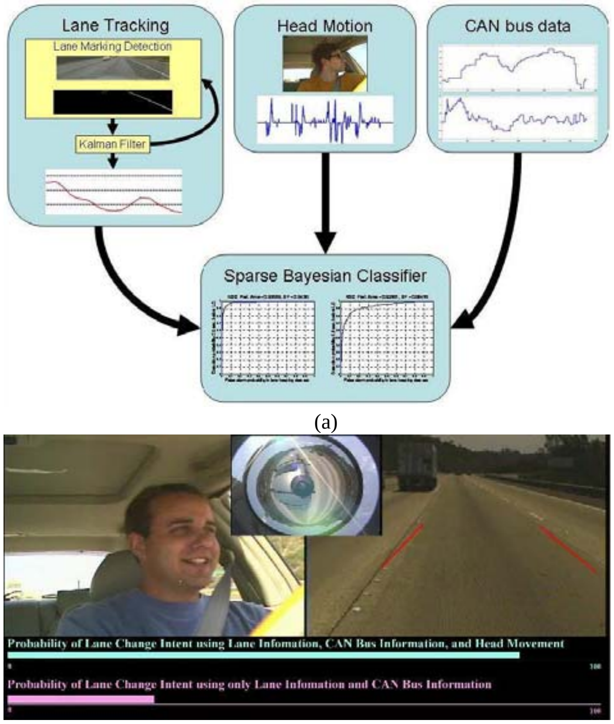
### *C. Looking-in and Looking-out of the vehicle: Lane change intent analysis*

This section gives an application that combines the use of sensors that are looking-in as well as looking-out in order to predict driver's intended actions. Driver intent inference presents a challenging classification problem; namely, given a diverse array of multi-modal features, how to infer or classify driver intentions.

This vision system estimates driver intentions in the specific area of lane changes, arguably one of the most important actions relevant to intelligent support systems. For this purpose, it uses the information not only from the camera looking outside to detect and track lanes, but also from camera inside the vehicle that monitors driver's facial motion. Since lane change is usually preceded by turning of head, the head motion gives an advance indication of the lane change intent. The system composed of a the following key components as shown in Figure 5 (a):

(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.

System details are described in [16]. Figure 5 (b) shows an example of lane change intent detection. The top bar shows the estimated probability of lane change using lane tracking, vehicle dynamics from CAN bus, as well as head motion. The bottom bar is derived without using the head motion. It is observed that the use of head motion gives an advantage of 0.5 seconds in detecting the lane change intent, which is critical for preventing accidents.



(a)

(b)

Figure 5 (a) Block diagram for inference of lane change intent using information from lane tracking, head motion, and the CAN bus data (b) 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 the head movement is used.

## IV. CONCLUDING REMARKS

Development of a real-time, a robust dynamic context capture system for an automobile, computer vision and 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" (LILO) of a vehicle. Such an "active", multi-modal sensory approach for *"dynamic context capture and situational awareness"* using cameras, radars, audio, etc. allows for establishing representations of the state of the environment, the vehicle, and the driver with accurate dynamic uncertainty characterization. It is believed that successful integration of such powerful sensory suits in 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.

## ACKNOWLEDGMENT

We want to thank our colleagues at the Computer Vision and Robotics Research Laboratory for their cooperation, assistance, and contributions. We are especially grateful to our industrial collaborators, who provided us with invaluable guidance and support to provide a "reality-check" for our research. It is indeed our pleasure to acknowledge the positive impact of our interactions with such collaborators including, Dr. Yousuke Akatsu, Dr. Erwin Boer, Dr. Akhtar Jameel, Dr. Akio Kinoshita, Dr. Satoshi Kitazaki, Dr. Chris Pribe, Dr. Klaus Schaff, Dr. Arne Stoschek, Dr. Hiroshi Takahashi, and their research teams.

## REFERENCES

[1] "World report on road traffic injury prevention: Summary," *World Health Organization*, Tech. Rep., April 2004, http://www.who.int/world-health-day/2004/infomaterials/world report/en

[2] "Traffic safety facts 2003 FARS/GES annual report (Final Edition)," *US Dept. of Transportation, National Highway Traffic Safety Administration*, Tech. Rep., January 2005, http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSFAnn/2003HTMLTSF/TSF2003%.HTM.

[3] E.D. Dickmanns, B. Mysliwetz, T. Christians, "An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles," *IEEE Transactions on Systems, Man and Cybernetics*, 20 (6), 1273-1284, Nov.-Dec. 1990.

[4] D. Pomerleau and T. Jochem, "Rapidly adapting machine vision for automated vehicle steering," *IEEE Expert: Special Issue on Intelligent System and their Applications*, 11 (2), 19-27, April 1996.

[5] U. Franke, D. Gavrila, S. Gorzig, F. Lindner, F. Puetzold, C. Wohler, "Autonomous driving goes downtown," *IEEE Intelligent Systems and Their Applications*, 13 (6), 40-48, Nov.-Dec. 1998.

[6] F. Heimes and H.-H. Nagel, "Towards active machine-vision-based driver assistance for urban areas," *International Journal of Computer Vision*, 50 (1), 5-34, 2002.

[7] Grace, R., Byrne, V. E., Legrand, J. M., Gricourt, D. J., Davis, R. K., Staszewski, J. J., Carnahan, B., A Machine Vision Based Drowsy Driver Detection System for Heavy Vehicles, Proceedings of The Ocular Measures of Driver Alertness Conference, pp. 75-86, April 1999, FHWA-MC-99-136.

[8] M. M. Trivedi, S. Y. Cheng, E. M. C. Childers, S. J. Krotosky, "Occupant Posture Analysis with Stereo and Thermal Infrared Video: Algorithms and Experimental Evaluation", IEEE Trans. Vehicular Technology, 53 (6), 1698-1712, Nov. 2004.

[9] M. Bertozzi and A. Broggi, "GOLD: a Parallel Real-Time Stereo Vision System for Generic Obstacle and Lane Detection," *IEEE Trans. on Image Processing*, 7 (1), 62-81, Jan. 1998.

[10] W. Enkelmann, "Video-based driver assistance - from basic functions to applications," *International Journal of Computer Vision*, 45 (3), 201-221, December 2001.

[11] L. Matuszyk, A. Zelinsky, L. Nilsson, and M. Rilbe, "Stereo panoramic vision for monitoring vehicle blind-spots," in *Proc. IEEE Intelligent Vehicles Symposium*, 31-36, June 2004.

[12] J. McCall, O. Achler and M. M. Trivedi, "Design of an Instrumented Vehicle Testbed for Developing Human Centered Driver Support System," In *Proc. IEEE Intelligent Vehicles Symposium*, 483-488, June 2004.

[13] T. Gandhi and M. M. Trivedi, "Dynamic Panoramic Surround Map: Motivation and Omni Video Based Approach," In *Proc. IEEE International Workshop on Machine Vision for Intelligent Vehicles*, June 2005.

[14] T. Gandhi, M. M. Trivedi, "Parametric Ego-Motion Estimation for Vehicle Surround Analysis Using Omni-Directional Camera," *Machine Vision and Applications*, January 2005.

[15] T. Gandhi and M. M. Trivedi, "Vehicle Mounted Wide FOV Stereo for Traffic and Pedestrian Detection," In *Proc. International Conference on Image Processing*, September 2005.

[16] J. McCall, D. Wipf, M. M. Trivedi, B. Rao, "Lane Change Intent Analysis Using Robust Operators and Sparse Bayesian Learning," In *Proc. IEEE International Workshop on Machine Vision for Intelligent Vehicles*, June 2005.

[17] J. McCall, M. M. Trivedi, "Visual Context Capture and Analysis for Driver Attention Monitoring", In *Proc. 7th IEEE Conf. on Intelligent Transportation Systems*, October 2004.