Vehicle Surround Capture: Survey of Techniques and a Novel Omni-Video-Based Approach for Dynamic Panoramic Surround Maps

Tarak Gandhi and Mohan Manubhai Trivedi

***Abstract*—Awareness of what surrounds a vehicle directly affects the safe driving and maneuvering of an automobile. This paper focuses on the capture of vehicle surroundings using video inputs. 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. A survey of literature related to surround analysis is presented, emphasizing detecting objects such as vehicles, pedestrians, and other obstacles. Omni cameras, which give a panoramic view of the surroundings, can be useful for visualizing and analyzing the nearby surroundings of the vehicle. 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 is introduced. A novel approach for synthesizing the DPS using stereo and motion analysis of video images from a pair of omni cameras on the vehicle is developed. Successful generation of the DPS in experimental runs on an instrumented vehicle test bed is demonstrated. These experiments prove the basic feasibility and show promise of omni-camera-based DPS capture algorithm to provide useful semantic descriptors of the state of moving vehicles and obstacles in the vicinity of a vehicle.**

***Index Terms*—Active safety, collision avoidance, driver information systems, image motion analysis, panoramic vision, stereo vision.**

I. Introduction and Motivation

SAFER automobile travel and smoother traffic conditions are universally sought. The universality and gravity of road hazards was recognized when the World Health Organization issued a comprehensive report on Road Accident Prevention [1]. According to the studies conducted by the U.S. National Highway Traffic Safety Administration (NHTSA), traffic accidents are the leading cause of death for age group 3–33 and the eighth leading cause among all age groups. The NHTSA report Traffic Safety Facts 2003 [2] states that there have been 38 252 fatal crashes with 42 643 fatalities due to accidents on U.S. roads in 2003. As shown in Table I, 40.4% were due to collision with other vehicles, 31.1% were due to collision with a fixed object, and 11.5% were due to collision with pedestrians. In fact, for passenger cars, accounting for 44.7% of total fatalities, around one third of fatal accidents are with other vehicles in front, one sixth are with vehicles on sides, and one sixth are with fixed objects. According to crash data from NHTSA, approximately 88% of rear-end collisions are caused by driver inattention and following too closely [3].

Safety systems for automobiles are typically classified in two categories, namely 1) passive and 2) active. Over the past three to four decades, passive safety approaches such as improved design and structure of the vehicle and restraint systems including safety belts and airbags [4] have indeed saved countless lives and have minimized the extent of serious injuries. Active systems are relatively new, and unlike passive systems, active systems are employed to prevent a vehicle from an accident or collision. Examples of these include antilock breaks, electronic stability control, and improved visibility systems. The research presented in this paper is about an active safety system for accurate and robust characterization of various dynamic events surrounding a vehicle that can be made available to the driver.

Intelligent driver support systems, which warn the driver of a possible collision, would allow the driver to take timely action to avoid or at least reduce the impact of collisions. In fact, NHTSA countermeasure effectiveness modeling predicts that "head-way detection systems can theoretically prevent 37% to 74% of all police reported rear-end crashes" [3]. In recent years, considerable research has been performed for developing such driver support systems to enhance safety by reducing the accidents. These systems use a sensor suite, which may contain the video cameras mounted in various positions, thermal infrared imagers, active sensors such as RADAR, LIDAR, LASER scanners, and ultrasonic sensors, and data from vehicle dynamic sensors. Lane detection helps to determine the lateral position of the vehicle and warn the driver in case of lane departure. This in turn helps to prevent scenarios such as collision with other vehicles or fixed objects, or running off the road. Detecting objects in front of the vehicle is useful in preventing accidents due to sudden braking of the other vehicle. Monitoring sides of the vehicle containing blind spots is especially useful when the driver intends to change lanes. Detecting nonmotorists such as pedestrians and bicyclists is of particular interest for nonhighway driving since pedestrians are harder to detect than vehicles and are more vulnerable. In addition to visible sensors, thermal infrared sensors are being explored for pedestrian detection.

Manuscript received March 4, 2005; revised June 30, 2005, January 14, 2006, February 3, 2006, and February 27, 2006. This work was supported by a UC Discovery Program Digital Media Grant in collaboration with the Nissan Research Center. The Associate Editor for this paper was C. Stiller.

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

Color versions of Figs. 1, 3, 5, and 7–13 are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2006.880635

TABLE I
(a) CRASHES BY FIRST HARMFUL EVENT, MANNER OF COLLISION, AND CRASH SEVERITY. (b) PASSENGER CARS INVOLVED IN CRASHES BY MOST HARMFUL EVENT AND CRASH SEVERITY (BASED ON [2])

(a)

	Crash Severity							
	Fatal		Injury		Property Damage Only		Total	
First Harmful Event	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Collision with Motor Vehicle in Transport	15,458	40.4	1,339,000	69.5	2,962,000	67.9	4,316,000	68.2
Collision with Fixed Object	11,879	31.1	332,000	17.2	637,000	14.6	980,000	15.5
Collision with Object Not Fixed	6,160	16.1	163,000	8.5	665,000	15.2	834,000	13.2
Noncollision	4,727	12.4	92,000	4.8	101,000	2.3	198,000	3.1
Total	38,252	100.0	1,925,000	100.0	4,365,000	100.0	6,328,000	100.0

(b)

	Crash Severity							
	Fatal		Injury		Property Damage Only		Total	
First Harmful Event	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Collision with Motor Vehicle in Transport by Initial Point of Impact								
Front	8,775	33.5	857,000	40.2	1,597,000	36.7	2,463,000	37.8
Left Side	2,733	10.4	247,000	11.6	611,000	14.0	861,000	13.2
Right Side	2,209	8.4	219,000	10.3	550,000	12.6	772,000	11.9
Rear	1,377	5.3	447,000	21.0	829,000	19.0	1,277,000	19.6
Other/Unknown	170	0.6	*	*	*	*	1,000	*
Subtotal	15,264	58.3	1,771,000	83.2	3,587,000	82.4	5,373,000	82.5
Collision with Fixed Object	4,831	18.5	193,000	9.1	382,000	8.8	580,000	8.9
Collision with Object Not Fixed								
Nonmotorist	2,512	9.6	70,000	3.3	3,000	0.1	76,000	1.2
Other	430	1.6	34,000	1.6	347,000	8.0	382,000	5.9
Subtotal	2,942	11.2	105,000	4.9	350,000	8.0	457,000	7.0
Noncollision	3,119	11.9	61,000	2.8	37,000	0.8	101,000	1.5
Total	**26,169	100.0	2,129,000	100.0	4,356,000	100.0	6,511,000	100.0

*Less than 500 or less than 0.05 percent. **Includes 13 passenger cars involved in fatal crashes with unknown most harmful event.

Thus, it can be seen that an effective driver assistance system should have a perception of complete surroundings including the events taking place in front, back, and sides of the car. A dynamic surround map illustrated in Fig. 1 could be very useful for visualizing and analyzing the situation. The surround map would contain lanes, vehicles, pedestrians, and other objects along with their attributes such as position, size, and velocity. Surround map generation would also be useful for offline analysis of interesting events, especially in driver behavioral studies [5]. Based on driver behavioral studies and the expertise of specialists on cognitive science, psychology, and human–machine interface, one would be able to develop and evaluate methods for conveying the information obtained from Dynamic Panoramic Surround (DPS) to the driver in reliable and nondistracting manner using modalities such as visual, haptic, auditory, or their combination.

In the following, we present research dealing with the derivation of DPS maps. These maps would be useful not only for systematic studies of driving behavior but also in the development of a new generation of driver assistance systems. This paper brings out the following contributions: (1) motivation and justification for surround map in driver assistance systems and driving behavioral studies; (2) presentation of a brief survey of research relevant to surround detection; (3) introduction of the DPS map concept and development of an integrated stereo- and motion-based panoramic surround map using omni video streams; and (4) experimental studies using LISA-Q [6], a novel instrumented test bed in generating DPS maps.

II. SURVEY OF RELATED RESEARCH

Autonomous driving and mobile robotics were the main drivers in the development of some of the early video-based obstacle detection and navigation systems. A number of good survey papers present some of those earlier efforts. Bertozzi *et al.* [7], [8], give a comprehensive survey of the use of computer vision in intelligent vehicles. Approaches for lane, pedestrian, and obstacle detection are described and analyzed. Kastrinaki *et al.* [9] present another survey of vision techniques used for traffic analysis from stationary platforms as well as moving vehicles. In addition, conferences such as the IEEE Intelligent Vehicles Symposium [10], [11] and the IEEE Intelligent Transportation Systems Conference [12] deal with the recent research in these topics. In [13], a survey of lane detection techniques and their characteristics is performed, and a novel method using steerable filters [14] is proposed to detect lanes as well as the Botts' dots markers especially found on California highways. In [15], Fang *et al.* present a system for road sign recognition based on the way humans perform recognition. The system consists of sensory, perceptual, and conceptual layers analyzing information at their respective levels and passing it to the next higher level. This concept is also used in [16] to detect changes in driving environment such as entering or exiting tunnel or freeway, lane changes, and overpasses.

Recently, there has been a shift toward the development of video-based systems for driver assistance rather than autonomous driving. A number of research groups have considered lane detection and tracking as the main driver for their

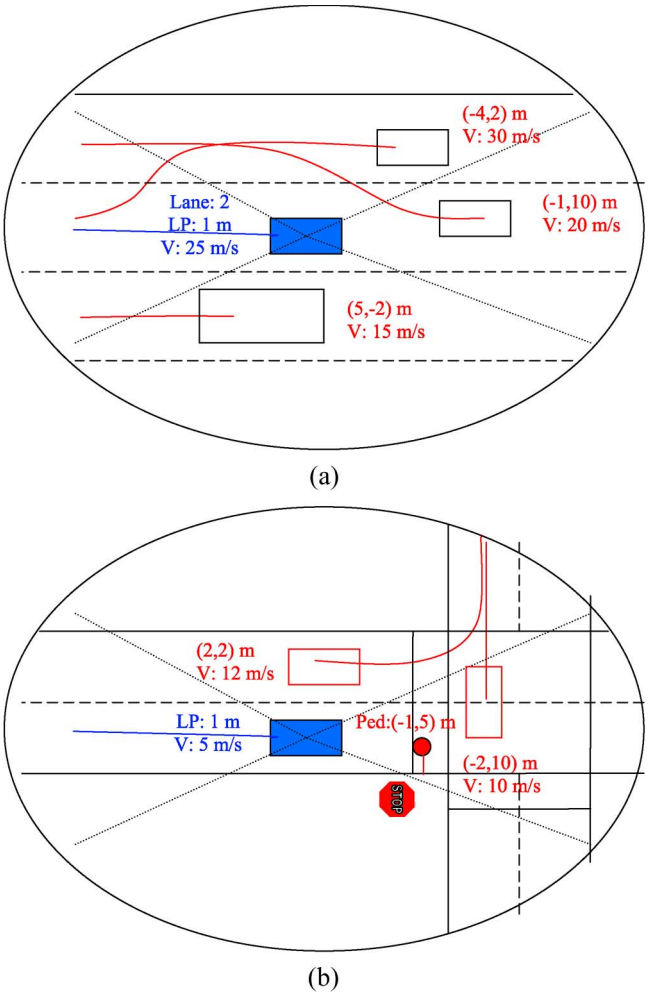


Fig. 1. Dynamic surround map for (a) freeway and (b) city streets. (x, y) coordinates of other objects w.r.t. ego vehicle; V velocity w.r.t. road; LP lateral position of ego vehicle w.r.t. center of the lane. The cross lines denote typical boundaries between the front and the side surroundings.

vision-based systems. This is indeed a very important and challenging problem, especially when the constraints of robustness, accuracy, and real-time operation are critical. In this paper, our focus is mainly on detection, localization, and mapping of objects (moving or stationary) in the immediate vicinity of a moving vehicle with video sensors. Table II shows a number of studies that consider related issues and are indeed quite relevant to the topic considered in this paper. A key differentiator of the research discussed in this paper is the desire to develop a DPS map in a "holistic" manner rather than considering independent sensors and modules for detecting objects in the front or side of the vehicle.

A. Front-Object Detection

The objects of interest in front of the vehicle include other vehicles, stationary obstacles, and pedestrians. For monocular cameras, motion is an important cue used for detection. For motion-based detection, the ego motion of background induced by camera motion is separated from the motion of independently moving objects in [17]. This approach is used in [18], which describes a system for video-based driver assistance involving lane and obstacle detection. In addition, three-dimensional (3-D) model-based detection [19], symmetry features [20], and Gabor filter banks with genetic algorithm [21] have also been used for vehicle detection.

Stereo cameras are particularly useful for separating front objects from road features. If the road is assumed to be planar, the disparities of road features can be determined using the camera calibration and compensated by appropriate transformation of the images, and the regions with residual disparity are processed to detect obstacles [22]–[24]. For dealing with nonplanar roads, Labayrade *et al.* [25] developed a novel concept of the V-disparity image. Each row of the V-disparity image contains a histogram of disparities on that row. Road and vehicles form distinctive patterns such as curves and vertical lines, which can be used to extract the vehicle positions. Rebut *et al.* [42] generates 3-D edge maps using stereo matching of image edges. They perform object detection by grouping and validating 3-D edges, and removing road edges using V-disparity.

Pedestrians are the most vulnerable objects on the road. For the PROTECTOR system [26], [27], pedestrian detection is performed with stereo images using hierarchical templates and neural networks. In [28], 3-D stereo detection is followed by support vector machine (SVM) classification to discriminate pedestrians from nonpedestrians. Thermal infrared images are particularly useful for pedestrian detection in cold and dark conditions by searching for hotspots and validating them using cues such as motion [29] or 3-D human models [30]. In [31], SVM-based classification is used to validate the pedestrians.

There is a considerable interest in active vision systems in which the camera parameters are controlled based on processing of external stimuli. These systems attempt to solve the problems due to large illumination changes and tradeoffs such as resolution versus field-of-view (FOV). In [32], an active vision system mounted on transit bus is used for real-time monitoring of traffic parameters.

Sensor fusion has been used to combine relatively accurate depth information from radar or laser scanner and the higher angular resolution of video to get more reliable detection. Kato *et al.* [33] use the depth output from a millimeter radar and model the obstacle as a planar surface at the distance given by the radar. Fang *et al.* [34] use coarse depth information of multiple targets from radar or stereo to split the image into layers corresponding to different depths to facilitate robust detection of objects.

B. Blind Spot Monitoring and Omnidirectional Vision

In addition to objects in front, monitoring of the rear view and side views including blind spots is also important. In [35], a biologically inspired motion detection approach is used to monitor the blind spots and warn the driver about overtaking vehicles. Researchers at the University of Minnesota have been working on lane keeping for the Minnesota Department of Transportation (Mn/DOT) Bus Rapid Transit program to enable buses to run safely in narrow bus-only shoulders and merge into normal traffic when required. They have developed a concept of "virtual mirror" [36], which uses lidar to detect objects around the bus, register them to the geographic information given by differential Global Positioning System (GPS), and display the

TABLE II
RELATED RESEARCH ON OBJECT DETECTION

System	Cameras/Sensors	Comments
Monocular vehicle detection		
Kruger MVA'99 [17], Enkelman IJCV'01 [18]	Monocular camera, navigation data	Performs motion based detection by modeling the road as a planar surface, and computing motion parameters by combining the data from vehicle dynamic sensors with information from spatio-temporal image gradients. The approach is used for video-based driver assistance system incorporating lane and obstacle detection.
Fleischer IV'02 [19]	Monocular camera	Detects vehicles and road signs using 3D models of objects and prior knowledge of areas in which these objects can be found. The 3D models are projected onto the image in various poses and compared with image features to identify the objects.
Sun ITS'05 [21]	Monocular camera	Uses a set of Gabor filters for vehicle detection. Genetic algorithm is used to perform global optimization of the filter set for best discrimination between vehicle and non-vehicle samples. An incremental clustering approach is then used to eliminate redundant filters.
Binocular vehicle detection		
Luong ICCV'95 [22]	Stereo cameras	Assumes planar road surface, compensates road plane disparity by transforming one image onto another, performs obstacle detection from the difference image.
Bertozzi TIP'98 [23]	Stereo cameras	Assumes planar road surface, compensates disparity of road using inverse perspective mapping of the two images. The obstacles form pairs of inverted triangles in the difference image and are detected using polar histograms.
Labayrade IV'02 [25]	Stereo cameras	Detects obstacles on planar and non-planar roads. Developed the concept of V-disparity image obtained by clustering similar disparities on each image row. A line or curve in this image corresponds to straight or curved road respectively. A vehicle on the road forms a vertical line in the this image, whose intersection with the road curve gives the point of contact.
Bensrhair IV'02 [24]	Stereo cameras	Vehicle detection using cooperative processing of monocular and binocular methods.
Rebut IV'04 [42]	Stereo cameras	Generates 3D edge maps using stereo matching of image edges. Object detection is performed by grouping and validating 3D edges, and removing road edges using V-disparity.
Pedestrian detection		
Gavrila ICCV'99 [26], IV'04 [27]	Stereo cameras	Pedestrian detection is performed using coarse-to-fine matching with hierarchical pedestrian templates on images segmented using stereo depth map. Neural network is used to verify the detected pedestrians and reduce false alarms. A methodology for experimental evaluation of pedestrian detection is developed and extensive experimentation is performed to validate the system.
Grubb IV'04 [28]	Stereo cameras	The scene is segmented into 3D objects using passive stereo vision with V-disparity. The objects are classified into pedestrians and non-pedestrians using a Support Vector Machine. Kalman filter is used to track the pedestrians and Bayesian probabilities are integrated over the track to obtain confidence measure.
Broggi IV'04 [30]	Monocular IR cameras	Pedestrians are detected in infrared images by winding warm objects with suitable aspect ratio, and validated using 3D human models that encode thermal and morphological characteristics of humans.
Liu TVT'04 [29]	Stereo IR cameras	Detects pedestrians by identifying moving objects whose motion is not consistent with background. Suitable for detecting pedestrians moving sideways across the road. Shows examples of pedestrian detection in dark conditions.
Xu ITS'05 [31]	Monocular IR camera	Performs pedestrian detection by identification of hot-spots, classification using SVM, and tracking using Kalman filter and mean-shift algorithm. Information from road-detection module is used for validating the pedestrians.
Active vision, Sensor fusion		
Rabie CVIC'01 [32]	Active vision system with stereo PTZ cameras	System for application in traffic monitoring from a moving bus. Detects independent motion using optical flow. Fastest moving object of significant size is detected, and the gaze of active camera is focused on it. Stereo tracking is used to obtain relative speed and distance between the camera and the tracked object.
Kato ITS'02 [33]	Millimeter Wave radar and monocular camera	Uses depth output from millimeter radar to model the obstacle as a planar surface. Vision is then used to accurately determine the lateral position. Pairs of image features are detected and those with relative motion consistent with the distance predicted by radar are classified as obstacles. Experiments show the localization of moving vehicles and pedestrians, successfully eliminating false alarms.
Fang ITS'02 [34]	Radar and stereo cameras	Uses coarse depth information of multiple targets from radar or stereo to split the image into layers corresponding to different depths. Edge features with similar depth are grouped to determine the target sizes and locations. Experiments show detection of other vehicles, pedestrians, and a bouncing ball.
Stiller IVC'00 [37]	Digital map, stereo cameras, radar, LASER scanners	Obtains reliable information about obstacles by fusing information from multiple sensors. Stereo vision is used to detect front objects. Multiple laser scanners look for obstacles in all directions. The outputs from these sensors are used to robustly track objects using Kalman filter. However, video images are not obtained for side and rear views.
Sergi 03 [36] (UMN Tech. Rep.)	Differential GPS, LIDAR mounted on sides of bus	LIDAR detects objects around the bus, which are registered to the geographic information given by a differential GPS. The objects and the environment in a part of the surroundings are displayed to the driver in form of a `virtual mirror'.
System	**Cameras/Sensors**	**Comments**
Blind spot monitoring, omni vision		
Mota IJRA'04 [35]	Monocular camera	A biologically inspired motion detection approach is used to monitor the blind spots and warn the driver about overtaking vehicles.
Matuszyk IV'04 [40]	Omni stereo cameras mounted coaxially	Use the V-disparity approach with coaxially mounted omni cameras to monitor blind spots behind the vehicle. Stereo matching is performed between the virtual panoramic images generated from the cameras. However, a large part of the FOV is occupied by the car itself and is therefore unused.
Huang IV'03 [41]	Monocular omni camera	An omni camera obtains a panoramic view of surroundings as well as the driver's face. The system synthesizes a virtual view of the surroundings towards the viewing direction of the driver by estimating driver's head orientation.
Achler IV'04 [38]	Monocular omni camera	Uses steerable filters to detect wheels of vehicles on side. Vehicle hypotheses are formed based on wheel detection and the vehicles are tracked using Kalman filter.
Gandhi MVA'05 [39]	Monocular omni camera, CAN bus data	Performs surround monitoring using motion based object detection on omni video. The ego-motion of the road is compensated using gradient-based method. Features having residual motion are grouped to detect vehicles and generate a complete surround view showing the position and tracks of the vehicles.
Names of Journals/Conference Proceedings		
MVA	Machine Vision and Applications	
IV	IEEE Intelligent Vehicle Symposium	
ITS	IEEE Transactions on Intelligent Transportation Systems	
ICCV	IEEE International Conference on Computer Vision	
TIP	IEEE Transactions on Image Processing	
TVT	IEEE Transactions on Vehicular Technology	
CVIC	Canadian Vision Interface Conference	
IVC	Image and Vision Computing	
IJRA	International Journal of Robotics and Automation	

objects and the environment to the driver. Stiller *et al.* [37] combine the information from digital map, stereo cameras, radar, and laser scanners for robust obstacle detection.

Omnidirectional cameras or omni cameras give a 360° view of the surroundings and have recently gained popularity in intelligent vehicle applications. In [38], vehicles on both sides are captured using virtual views from a single omni camera and detected based on wheel features. In [39], a monocular omni camera is used to estimate and compensate vehicle ego motion, and features having residual motion are grouped to detect vehicles. In [40], the V-disparity approach developed by Labayrade *et al.* [25] is applied to a pair of coaxially mounted omni cameras for monitoring blind spots behind the vehicle especially during reversing. However, a large part of the FOV is occupied by the vehicle itself and is therefore unused. In [41], an omni camera mounted in a car obtains a panoramic view consisting of the surroundings as well as the driver's face.

C. Role of Surround Generation

Many of the above systems deal with lanes, front objects, or side objects in isolation. For driver assistance systems to be effective, the full surround scenario should be considered in an integrated manner. Omni cameras with their panoramic view are naturally suited for complete surround monitoring of vehicles. A virtual view looking toward any specified direction can be electronically generated. However, due to large FOV of the camera, the resolution of the virtual views becomes low. Hence, the omni cameras are appropriate for detecting nearby objects using a small number of sensors.

In this paper, we present an approach for generating a surround map using a pair of omni cameras mounted on the sides of the vehicle. Video images from both omni cameras are used to detect objects in front of the vehicle using binocular stereo. For each side of the vehicle, a monocular view is available from the respective omni camera. "Motion stereo," which uses two consecutive frames from the moving vehicle, is applied to detect objects on the sides. Note that although this particular implementation uses omni cameras with hyperbolic mirrors, the approach can be easily adapted to other types of optics, such as fish-eye lenses with wide FOV.

In our previous work [39], motion analysis with a monocular omni camera was used to generate a surround map. It was observed that the monocular analysis gave good results on the sides of the vehicle, where independent motion of other vehicles gives large motion disparity. However, for front objects at larger distances, binocular stereo-based methods are more effective since the stereo disparity between laterally placed cameras is usually greater than the motion disparity due to the longitudinal motion of a single camera. Therefore, this system combines the use of binocular stereo analysis using two cameras for front view and motion analysis using individual cameras for each side view.

III. Vehicle Surround Capture With Omni Video

The block diagram of the video analysis framework is shown in Fig. 2. Video sequences are obtained from a pair of omni cameras mounted on the two sides of the vehicle. Camera

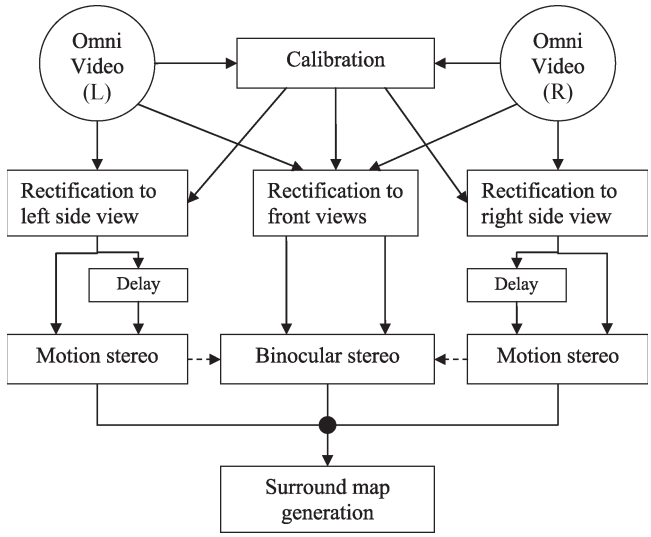


Fig. 2. Block diagram of video analysis from vehicle-mounted omni cameras.

calibration is performed offline to determine the relationship between the vehicle coordinates and pixel coordinates. Using the calibration information, the images are transformed to obtain virtual perspective views looking in front of the vehicle. This transformation called rectification simplifies the binocular 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.

Each omni camera also gives a monocular view of its side of the road. To detect side objects, stereo processing is performed between two consecutive frames from the "same" moving camera, so that the camera is displaced between the frames. This process is known as motion stereo. Rectification is performed by projecting a part of the image on the vertical plane parallel to the direction of motion. In this case, the disparity map has components corresponding to the distance of the object as well as the independent motion of the object. However, if the moving object is rigid with the exposed surface almost vertical, its disparity is approximately constant. On the other hand, the road below the horizon has the disparity varying with image position. This difference can be used to segment at least the part of the object below the horizon from the road. The approximate object position can be obtained by assuming that the bottom of the object touches the road.

The object positions obtained from binocular stereo and motion stereo are converted to the vehicle coordinates to insert them into the surround map.

The following coordinate systems are used, as shown in Fig. 3.

1) Vehicle Coordinate System: The origin is on the ground, the Z-axis points toward the front of the vehicle, the Y-axis points vertically downward, and the X-axis points toward right side. The coordinates in the vehicle system are denoted by $P_0 = (X_0, Y_0, Z_0)^T$.

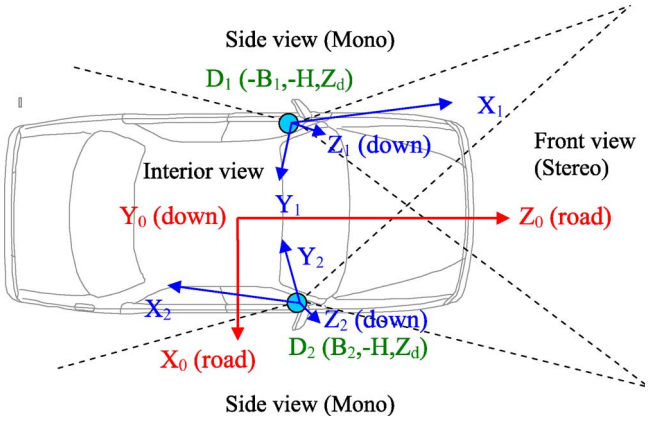


Fig. 3. Camera configuration for surround capture and the relevant coordinate systems. Dashed lines denote the approximate boundaries of the front and the side virtual views obtained from the omni cameras.

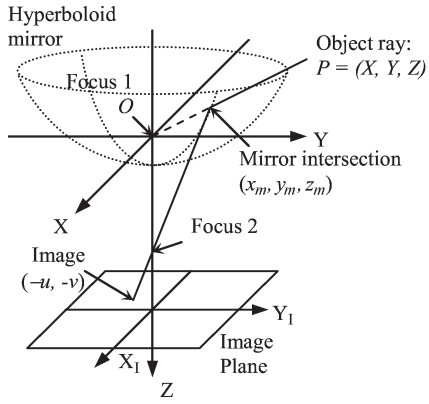


Fig. 4. Omni camera geometry. The ray from the object passing through the first focus is reflected toward the second focus and forms an image in the charge-coupled device (CCD) camera.

2) Coordinate System of Camera i: The origin of a camera system is at its optical center, the Z-axis is aligned with the optical axis, and the X-axis and Y-axis form the image plane. The transformation between the vehicle and the camera coordinate frame $i = 1, \ldots, 2$ is represented as

$$P_0 = R_i P_i + D_i, \quad P_i = R_i^T (P_0 - D_i) \tag{1}$$

where P_i and P_0 are coordinates in camera and vehicle systems, and R_i and D_i are the rotation matrix and translation vectors of camera i, respectively.

3) Omni Pixel Coordinates: Each point P_i in the camera i's coordinate system is projected onto the omni camera pixels $w_i = (u_i, v_i)^T$ by a many-to-one mapping f_i as

$$w_i = f_i(P_i), \quad P_i(\lambda) = \lambda g_i(w_i) + h_i(w_i) \simeq \lambda g_i(w_i) \tag{2}$$

where f_i denotes a many-to-one transform from camera to pixel, and g_i and h_i constitute the inverse transformation mapping every pixel w_i back to a 3-D line parameterized by λ in the camera coordinate system. For central panoramic cameras, all projection lines pass through a single point on the Z-axis, which can be taken as origin, i.e., $h_i = 0$. Even if the camera is not central, one can use the central approximation provided the points observed from the camera are at a large distance compared with the camera's dimensions, which is the case in this application.

The omni cameras used in this work are central panoramic cameras, each consisting of a hyperbolic mirror and a camera placed on its axis, with the center of projection of the camera on one of the focal points of the hyperbola as shown in Fig. 4. It can be shown [43] that a point $P_i = (X_i, Y_i, Z_i)^T$ in the camera coordinate system is mapped to pixel coordinates $w_i = (u_i, v_i)^T$ according to

$$\begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} = K_i \begin{pmatrix} X_i \\ Y_i \\ a_i Z_i + b_i \sqrt{X_i^2 + Y_i^2 + Z_i^2} \end{pmatrix} \tag{3}$$

where a_i and b_i depend on the omni mirror parameters, and K_i is the 3×3 calibration matrix for camera i. To convert an omni pixel (u_i, v_i) back to camera coordinates P_i (or p_i with a scale factor), the inverse transformation is given by

$$P_i = \lambda p_i = \lambda \begin{pmatrix} u_i' \\ v_i' \\ a_i - b_i \sqrt{u_i'^2 + v_i'^2 + 1} \end{pmatrix}$$
$$\begin{pmatrix} u_i' \\ v_i' \\ 1 \end{pmatrix} = K_i^{-1} \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} \tag{4}$$

where λ is the scale factor signifying that the omni pixel maps to a 3-D line passing through camera origin.

Although this work uses hyperbolic omni cameras, other types of optics such fish-eye lenses could also be used. In that case, one only needs to change the functions f, g, and h in (2) based on the geometry of the optical system.

IV. Camera Configuration and Calibration

Due to the comparatively lower resolution of omni cameras, proper configuration is very important for obtaining good coverage, sensitivity, and foreground–background discrimination. Consider the simplified situation of a pair of stereo cameras with focal length f and baseline B at height H above the road as shown in Fig. 5(a). Let an object at distance D be modeled using a rectangle perpendicular to the camera axes of width w and height h. In such case, the whole rectangle will have a constant disparity, i.e.,

$$d = \frac{Bf}{D} \tag{5}$$

where f is the focal length. If $h < H$ and if the object were removed, the points at the top of the original object would be in the same line of sight as the points on the road at distance D_{bg} given by

$$D_{\text{bg}} = D \frac{H}{H - h}. \tag{6}$$

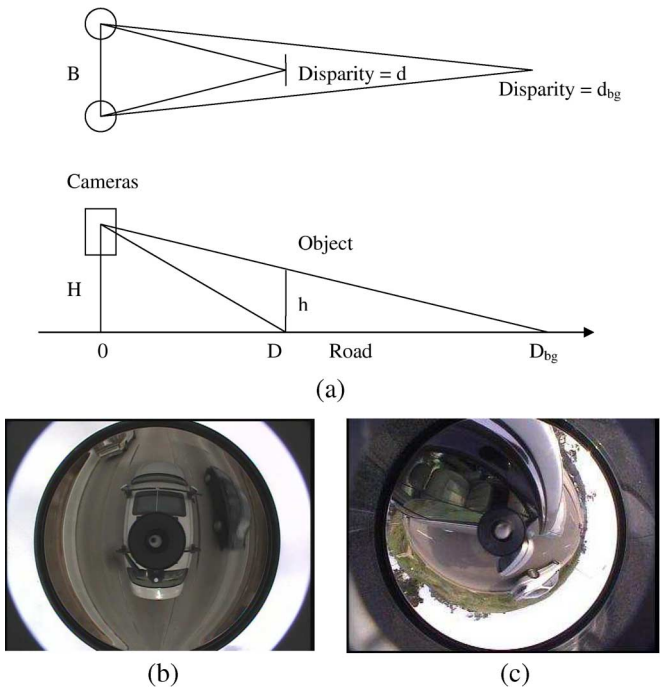


Fig. 5. (a) Geometry of object detection from a vehicle-mounted camera. (b) Image from a top-mounted camera. The object size is large, but the disparity difference $d_h - d_{\rm bg}$ between object and projected point on the road is small. (c) Images from the side-mounted camera. The object size is small, but the disparity difference $d_h - d_{\rm bg}$ between object and projected point on the ground is large.

Hence, the disparity $d_{\rm bg}$ of the road at that point would be

$$d_{\rm bg} = d\frac{H-h}{H}. \tag{7}$$

The difference in disparities between the object top and ground on the same line of sight would be

$$d - d_{\rm bg} = d\frac{h}{H}. \tag{8}$$

Thus, a smaller camera height H gives a larger disparity difference making it easier to discriminate between objects and road.

The above analysis is helpful in determining the advantages and limitations of different camera configurations. For example, if a camera is mounted higher, such as on the top of the roof as in [39], the resulting images contain the top and sides of other vehicles that have a larger surface area as shown in Fig. 5(b). However, the disparity difference between vehicle and road is small, making it difficult to isolate the vehicles purely by stereo. Furthermore, such a configuration is not suitable for standard cars. A pair of coaxially mounted omni cameras was used by Matuszyk *et al.* [40] to monitor blind spots behind the car effectively. However, a large part of the FOV is occupied by the car itself and is therefore unused.

Here, we use a pair of cameras near each side-view mirror as shown in Fig. 3 (see also Fig. 7 for the actual car photos and omni images). The view in front of the vehicle overlaps in the two cameras and is used for binocular stereo. A monocular view on each side is obtained from the corresponding camera and is used for motion stereo. The disparity difference between the vehicle and the ground is larger than that for a top-mounted camera, making binocular stereo discrimination easier. As a tradeoff, other vehicles have a smaller frontal area with the window-mounted camera, reducing their image size as seen in Fig. 5(c). This makes it somewhat more difficult to detect vehicles that are farther away.

In addition, the views of the driver and passenger are also obtained in this camera configuration and can be used to analyze the driver behavior. In [41], for example, the driver's face is detected in a similar setup by fitting an ellipse to edges using the randomized Hough transform and tracked using a Kalman filter. The face orientation is estimated using hidden Markov models and used to generate the view that the driver observes. This approach has been shown to be robust to illumination changes, shadows, and other imaging problems.

To match points between multiple cameras and map them to the 3-D space, it is necessary to calibrate the intrinsic and extrinsic parameters of the cameras. The intrinsic parameters relating the pixel and camera coordinates can be precomputed before the cameras are installed using a setup in [39] or [40]. The intrinsic parameters are used to transform the pixel coordinates w_i to a ray in 3-D space through P_i. The extrinsic parameters relate the vehicle coordinates to each of the camera coordinates and need to be calibrated when the cameras are mounted on the vehicle. In particular, the rotation matrices R_i of the cameras are used to obtain rectified perspective views of the overlapping FOVs of the omni cameras.

Calibration of extrinsic parameters is currently performed by taking the vehicle into a scene with a number of parallel lines in the directions of the vehicle axes. Sample points on the lines are manually marked. Using (4), the pixel coordinates $w_1, w_2, \ldots, w_K$ of sample points on each line are transformed to camera projective coordinates $p_1, p_2, \ldots, p_K$ within a scale factor, each corresponding to a ray from the camera origin in the direction of the line of sight. The coordinates are normalized, so that $\|p_i\| = 1$ and therefore correspond to projections on a virtual sphere centered at the camera origin. Singular value decomposition (SVD) of the following matrix is used to fit a line to all the points:

$$L = (p_1 \quad p_2 \quad \cdots \quad p_K). \tag{9}$$

The left singular vector corresponding to the smallest singular value of L gives the vector l corresponding to line equation $l^T p = 0$. The procedure is repeated for a number of lines in the image.

Images of parallel lines intersect at the vanishing point corresponding to the point at infinity in the direction of the line. If a number of parallel lines have images represented by line vectors $l_1, l_2, \ldots, l_N$, their vanishing point m of these lines can be similarly obtained using SVD of

$$M = (l_1 \quad l_2 \quad \cdots \quad l_N). \tag{10}$$

For lines along the length of the car (vehicle Z-axis), the direction of the vanishing point is $(0,0,1)^T$ in the vehicle coordinate system. In the camera coordinate system, this is transformed to $R^T(0,0,1)^T = r_3$, where r_3 is the third column of R^T (or third row of R). Similarly, the vanishing point of the lines across the length of the car along vehicle X-axis is

$(1,0,0)^T$ in the vehicle coordinate system, transforming to r_1 in the camera coordinate system. This way, the vanishing point in each direction is equal to the respective row of the rotation matrix R.

For the vertical Y-direction, one can use vertical objects such as poles in the image and find their intersection. If such vertical objects are not available, one can assume orthogonality and put $r_2 = r_3 \times r_1$. Moreover, due to errors in the determination of vanishing points, it may be possible that the rows r_1, r_2, r_3 of the matrix R may not be exactly orthonormal to each other. In such case, one can take the SVD $R = USV^T$ and replace R by UV^T, which is orthonormal. In fact, using an argument similar to [44, Th. 5.9], it can be shown that this matrix is an orthonormal matrix closest to the original R.

However, calibration obtained using the above procedure is sensitive to camera drift and vibrations, which should be dynamically corrected. For this purpose, one can use the epipolar geometry [44] of the scene. Point and line features can be extracted from the omni images, and knowing the prior calibration would help to narrow down the search region for correspondences. The RANSAC algorithm [44] can then be used to estimate the calibration parameters more accurately. It may also be possible to use fixed fiducial marks on the car's body in the FOVs of the cameras to partly compensate the camera rotations. Also, for initial calibration, the manual marking of lines could be replaced by automatic line finding methods such as Hough transform, which is applied in the projective coordinate space.

V. Stereo- and Motion-Based Panoramic Surround Generation

The calibration obtained above is used to obtain a virtual perspective view in front of the car from both the omni camera images. If the cameras are at the same height and longitudinal position, rectified images are obtained, in which the viewing directions of both images are parallel, and the baseline is perpendicular to the viewing direction. Suppose a point (X, Y, Z) forms images at (x_l, y_l) and (x_r, y_r) in the respective virtual perspective planes of the two cameras with focal length f. Then, these are related by

$$x_l - x_r = fB/Z, \quad y_l = y_r. \tag{11}$$

Stereo matching is simplified in this configuration where for every feature at (x_l, y_l), one tries to find the corresponding feature (x_r, y_r) that should lie along the same row. If the difference between the camera heights and positions is small, the vertical disparity $y_l - y_r$ can be neglected for objects far from the cameras.

The basic problem in stereo analysis is to find such correspondences. Stereo methods are classified into sparse and dense stereo. Considerable research has been done on stereo matching, with [45] and [46] surveying the state of art. Implementations such as [47] are commercially available for computing disparity maps between two stereo images. In case of sparse stereo, the depths are obtained only at significant features such as corners or edges. On the other hand, dense stereo attempts to find depths at all pixels. However, due to aperture problem, the depth computations in areas with low texture may not be reliable. Such pixels should either be removed or the depths interpolated using neighboring pixels. In [48], a hierarchical approach based on using multiple primitives such as edge pixels, edge segments, and regions is used to produce depth maps at different levels of detail. Results of stereo analysis at higher levels are used to guide matching at lower levels to produce reliable depth maps for wide variety of scenes.

In this work, area-based correlation is used for matching to obtain dense depth map. The implementation in [47] was used, which consists of the following steps.

1) *Preprocessing:* To effectively perform stereo matching, bandpass filtering is applied to the images by convolving with a Laplacian of Gaussian filter, i.e.,

$$G(x,y) = \sum_{(\delta x, \delta y) \in W_l} F(x - \delta x, y - \delta y) M(\delta x, \delta y) \tag{12}$$

where F is the original image, M is the filter mask with region of support W_l, and G is the resulting image.

2) *Area correlation:* To find correspondences between the image regions, rectangular patches from one image centered around (x_l, y) are compared with neighboring patches (x_r, y) displaced in horizontal direction using sum of absolute differences (SAD), i.e.,

$$S = \sum_{(\delta x, \delta y) \in W} \|G(x_l + \delta x, y + \delta y) - G(x_r + \delta x, y + \delta y)\| \tag{13}$$

where disparity $d = x_l - x_r \in \{0, \ldots, D_{\max}\}$, W is the region of support of the patch window mask, and $D_{\max}$ is the maximum permissible disparity. The processing is performed at two levels—coarse and fine—to fill the areas with low texture and allow for a larger range of disparities.

3) *Peak extraction:* For each patch centered at (x_l, y) in the first image, the corresponding patch (x_r, y) in the second image with minimum value of SAD is extracted. The displacement between the patches $x_l - x_r = d$ is the disparity for that pixel. A disparity map is formed by assigning a value proportional to d to each pixel (x_l, y) in the left image.

4) *Post filtering:* This step cleans up the noise in the disparity image. An interest operator is used to reject uniform areas where the SAD would not have a sharp extremum and disparity computations would therefore be unreliable. A left–right check is used to eliminate errors due to depth discontinuities. Pixels with invalid disparity are given a null value in the disparity map.

A temporal consistency check is then performed on the disparity map by computing the per-pixel running variance and removing pixels with variance larger than the threshold. This process suppresses pixels with noisy disparities. However, some pixels at the disparity edges may also get suppressed,

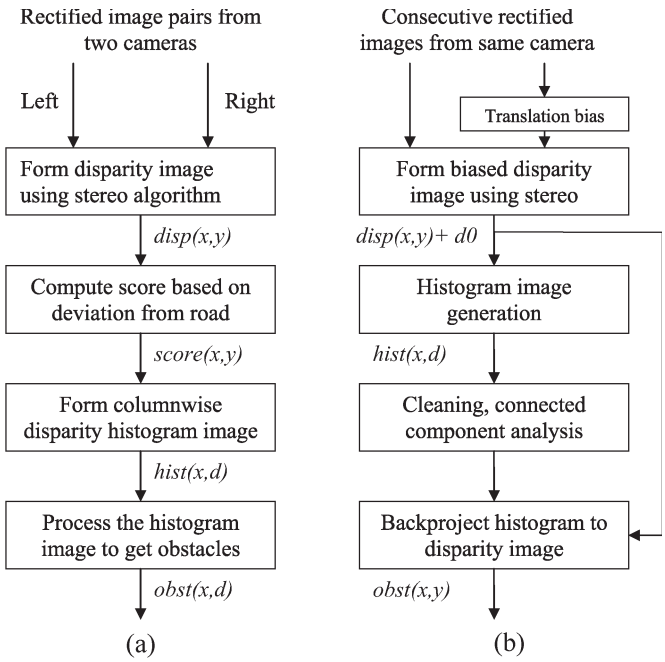


Fig. 6. Flow charts for (a) front-object detection using binocular stereo and (b) side-object detection using motion stereo.

although the pixels somewhat inside the object should pass the test. For better discrimination, one may need to adaptively adjust the threshold according to the presence of disparity edges.

A. Front-Object Detection

The road is modeled as a planar surface. Therefore, the disparity in rectified images is zero at the horizon and linearly increases with y. The theoretical road disparity is computed using the information from the calibration module. The pixels having disparities smaller than or equal to this disparity plus a small threshold are assumed to lie on the road and are suppressed. Moreover, pixels on a vehicle object are likely to have the same disparity. To facilitate detection, pixels with the same disparity in each column are clustered in a similar manner as in [25]. The steps in the above process shown in Fig. 6(a) are as follows:

1) Form an image $(\text{score}(x, y))$ based on the difference between the disparity d of the pixel and the corresponding disparity d_{bg} if it were on the road such that

$$\text{score}(x,y) = \begin{cases} 0 & \text{if } \text{disp}(x,y) \le d_{\text{bg}}(y) \\ [\text{disp}(x,y) - d_{\text{bg}}(y)]/T & \\ \quad \text{if } d_{\text{bg}}(y) \le \text{disp}(x,y) \le d_{\text{bg}}(y) + T & \\ 1 & \text{if } \text{disp}(x,y) \ge d_{\text{bg}}(y) + T \end{cases} \tag{14}$$

2) Form a columnwise disparity histogram image $(\text{hist}(x, d))$ such that

$$\text{hist}(x,d) = \sum_{\text{disp}(x,y)=d} \text{score}(x,y). \tag{15}$$

Each column of this image is a histogram of the disparities in that column weighted by the score. Pixels in an object at a particular distance would have nearly the same disparity and therefore form a horizontal ridge in the disparity histogram. Even if disparities of individual object pixels are inaccurate, the histogram image clusters the disparities and makes it easier to isolate the objects.

3) Smooth the histogram image using Gaussian filtering and morphological closing. Threshold the image and find connected components.
4) Form a front obstacle image $(\text{obst}(x, d))$ by replacing pixels in each column of connected component in histogram image by the centroid. This averages the error in the disparity and gives a subpixel disparity value that is stored as the gray level in this image.

B. Side-Object Detection

Each omni camera obtains a monocular view on the respective side of the car. The above stereo algorithm is applied to consecutive frames from the same camera, so that the camera is displaced between the two frames creating a stereo pair. This process is known as motion stereo. In the absence of independent motion, the disparity of each feature is directly proportional to the camera velocity and inversely proportional to the distance to the object. However, in the case of independently moving objects, the object motion adds to the disparity; hence, depth estimates cannot be derived directly from the disparity. However, if the moving object is rigid with the exposed area almost vertical, its disparity is approximately constant. On the other hand, for the road below the horizon, the disparity is a function of the image position. This difference can be used to segment the object from the road. Note that the distant features above the horizon have zero disparity. Hence, if the object has the same speed as the camera, one would not be able to separate parts of it above the horizon. However, if they have different speeds, one can separate the object even from the distant background above the horizon.

Note that the disparity can be negative in case of overtaking vehicles, whereas the stereo algorithm implementation [47] searches only positive disparities. To overcome this, the right image is displaced by a constant offset toward left, hence adding a fixed value to the disparity, making it positive in most cases. The object extraction procedure is a modification of that for binocular stereo [Fig. 6(b)] as described as follows.

1) Displace the right image by a fixed offset toward left to make disparities positive.
2) Use the implementation for binocular stereo [47] to form disparity image.
3) Form the disparity histogram image $(\text{hist}(x, d))$ using steps similar to (1)–(3) in the previous algorithm.
4) For each component in the histogram image, find the pixels in disparity image that contributed to the component. Form a side obstacle image $\text{obst}(x, y)$ by replacing each column of these pixels in disparity image with the bottommost pixel.
5) Assume that the bottom of the object lies on the ground and compute its vehicle coordinates using the calibration parameters.

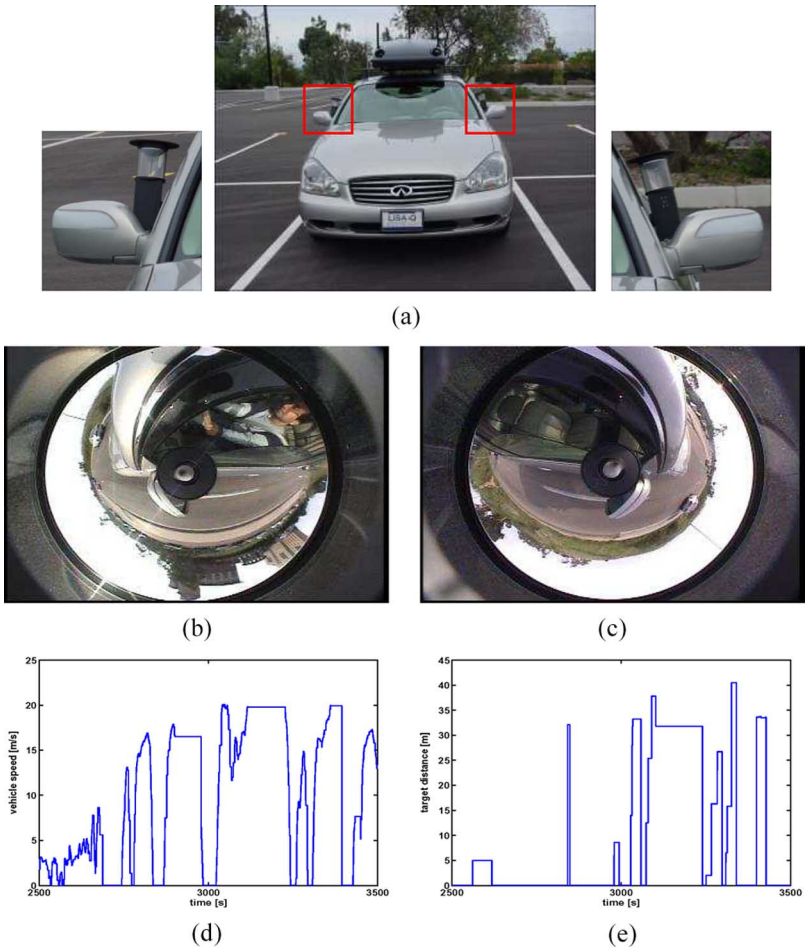


Fig. 7. (a) LISA-Q intelligent vehicle test bed. (b) Insets are close-up views of the omni cameras on two sides of the vehicle. (c) Video images from the omni cameras on the left and right sides of the vehicle. (d) Vehicle velocity retrieved from the CAN bus plotted against time. (e) Plot of the distance to the front object obtained from the radar.

C. Panoramic Surround Generation

The front and side obstacle maps obtained above are used to generate the surround map. Suppose the two omni cameras are situated at $D_1 = (-B_1, -H, Z_d)$ and $D_2 = (+B_2, -H, Z_d)$ in the vehicle coordinate system, where $B = B_1 + B_2$ is the baseline, H is the camera height, and Z_d is the longitudinal distance between the camera and the vehicle center. Assume without loss of generality that the centers of the rectified virtual images are at (0, 0), and the focal length is f, corresponding to the pixel scale. Then, for each pixel (x_1, y_1) in the front obstacle map (with respect to (w.r.t.) the left image) with disparity d, the projections of the 3-D vehicle coordinates on ground plane (X_0, Z_0) in the vehicle system are given by

$$Z_0 = \frac{f}{d}B + Z_d, \quad X_0 = \frac{x_1}{d}B - B_1. \tag{16}$$

In case of side obstacle map, the disparities do not give distance information for moving objects. Hence, it is assumed that the points (x_b, y_b) on the bottom of obstacle lie on the ground, and the y coordinate of the pixel is used to find the object distance. Thus, for the left camera

$$X_0 = -\left[\frac{y_b}{f}H + B_1\right], \quad Z_0 = \frac{x_b}{y_b}H + Z_d \tag{17}$$

and for the right camera

$$X_0 = \left[\frac{y_b}{f}H + B_2\right], \quad Z_0 = -\frac{x_b}{y_b}H + Z_d. \tag{18}$$

The vehicle coordinates (X_0, Z_0) of each obstacle pixel in front and side are projected on the panoramic surround map forming contours corresponding to each component. To smooth the contours and fill the gaps, a morphological opening operation is applied to the contours to keep errors on the side of caution by assigning a closer object distance when in doubt.

The above analysis applies in the case that the own car is travelling in straight line parallel to its axis. We propose the following approach when the own car is turning and the axes of the vehicle and omni camera are continuously changing. The yaw rate and speed information available from the Controller-Area-Network (CAN) bus of the vehicle could be used to compute the position of the car in the current vehicle coordinate system at the time of the two sample images used for motion stereo. The virtual images should be created in a vertical plane parallel to the displacement vector between the two positions. The disparity analysis should then be applied as before, and the computed object position should be transformed back to the vehicle's current coordinate system. The sensitivity of this

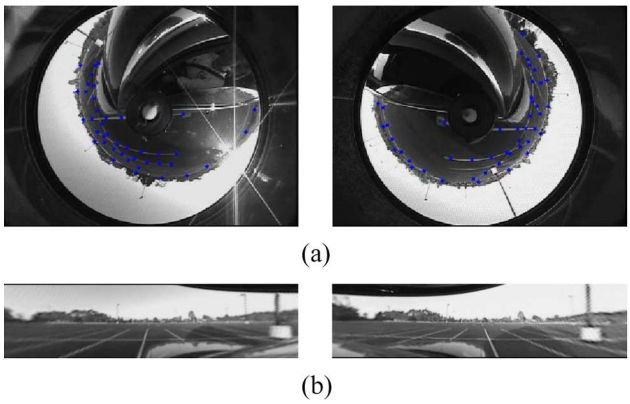


Fig. 8. (a) Images used for calibration of the cameras. Points on the lines in parking lot as well as on the horizon line are used to obtain the calibration parameters. (b) Virtual perspective images from the cameras corresponding to view in front of the car.

approach to yaw rate accuracy needs to be studied to check the robustness of the approach.

VI. DPS Maps: Experimental Studies

The LISA-Q intelligent vehicle test bed shown in Fig. 7(a) [6] is designed as a system capable of collecting large amounts of data from a variety of modular sensing systems and processing that data to be fed back to the human occupant. Sensor systems include rectilinear cameras, wide FOV 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. Fig. 7(b)–(e) shows the samples of acquired video and other sensor data. The key capabilities of the LISA-Q intelligent vehicle include

- Eight National Television Standards Committee (NTSC) hardware video compressors for simultaneous capture;
- CAN interface for acquiring steering angle, pedals, yaw rate, and other vehicle information;
- Built-in five-beam forward-looking laser radar range finder;
- Wide Area Augmentation System (WAAS)-enabled GPS;
- Integration into car audio and after-market video displays for feedback and alerts.

Detailed information about this test bed is described in [6].

For this work, a pair of omni cameras was mounted on the windows of the test-bed car. The data from the CAN bus was also acquired. The car was first moved around in a parking lot with parallel lines used for calibration. Note that the parking lines are mapped to curves in the omni image. The points on the lines were manually marked as shown in Fig. 8(a), and calibration was computed using the method in Section IV. Rectification using this calibration generates virtual perspective views in front of the car as shown in Fig. 8(b). Due to error in estimating the calibration, a small correction in the form of a vertical translation was manually applied to one of the rectified images, so that the disparities lie exactly along horizontal lines.

After calibration, the car was driven on city roads. Video sequences from the omni cameras were captured and stored. The processing was performed offline on a Pentium IV. Individually, the binocular stereo processing of both images or motion stereo processing of each of the side images could be performed with a frame rate of approximately 15 frames/s. This would result in effective speed of 5 frames/s if all the processing was performed on a single processor. However, since the processing for binocular stereo and motion stereo are independent, the use of three processors could give a rate of 15 frames/s, which would be suitable for real-time performance.

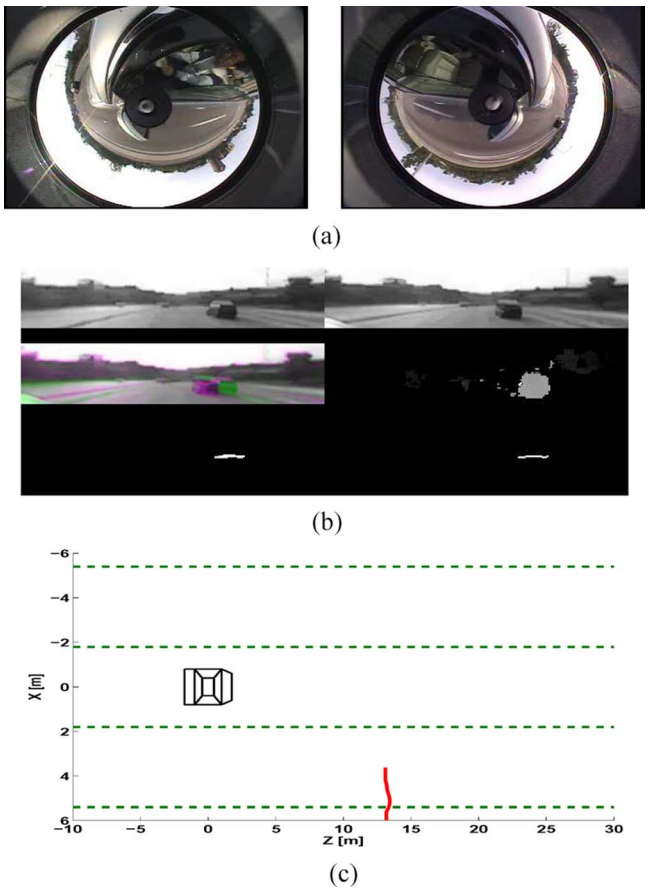


Fig. 9. Binocular stereo for front surround generation. (a) Left and right image frames from omni videos. (b) Processing of the omni images in six parts. Top: Virtual front views generated from omni cameras showing a car in front. Mid left: Composite color image showing displacement of the car between left and right images. Mid right: Disparity map showing nearer objects with lighter shades. Bottom left: Disparity histogram image. The disparities corresponding to the car in front cluster to form a line segment. Bottom right: Obstacle image obtained by processing the disparity histogram image. (c) Surround map showing own car (black) and the projected approximate position of the car in front (red).

Fig. 9 shows the application of the binocular stereo on the surroundings in front of the car. The raw images obtained from the left and the right omni cameras are shown in Fig. 9(a). The processing of these images is shown in the six parts of Fig. 9(b). The top two images in this figure show the virtual front views obtained by transforming the images using the calibration parameters. The superimposition of these two images shown in mid left clearly shows the disparity between the images for the object car if viewed in color. The stereo implementation of [47] was used to compute the disparity map shown in mid right. In this image, brighter shades correspond to nearer objects and darker shades to farther objects. Pixels where disparity computations were unreliable due to uniform

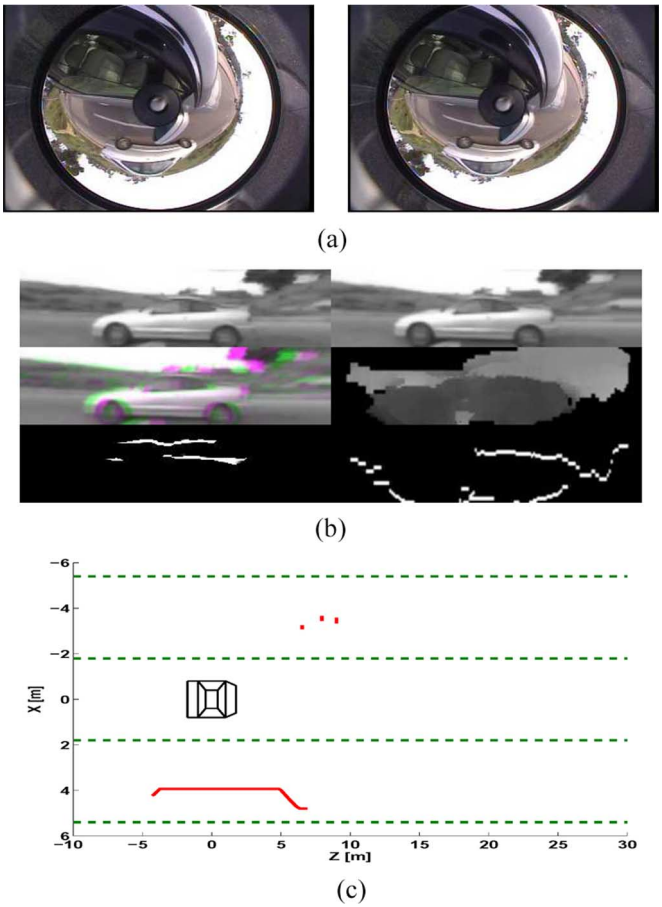


Fig. 10. Motion stereo for side surround generation. (a) Consecutive image frames from right omni video. (b) Processing of the omni images in six parts. Top: Virtual side views generated from omni cameras showing a car on side. Mid left: Composite color image showing displacement of the car between left and right images. One of the images is translated to make the disparities positive. Mid right: Disparity map with larger disparity in lighter shades. Bottom left: Disparity histogram image. The disparities corresponding to the car and the background cluster to form horizontal line segments. Bottom right: Obstacle image obtained by backprojecting these clusters to the disparity image, and for each cluster, marking the lowest pixel in every column. (c) Surround map showing own car (black) and the projected approximate position of the car on side (red).

texture are also shaded dark. The histogram image of the columns of the disparity map was computed as shown in the bottom left part with each row in this image corresponding to a particular disparity. Each pixel in this image has the brightness proportional to the weighted sum of the pixels in that column having the corresponding disparity. Since the front features of the car have nearly the same disparity, they accumulate in the form of a horizontal segment in the disparity image and can be easily isolated. The histogram image is processed to give the obstacle image showing the column position and the disparity of the object in the bottom right part. Using this information along with the camera calibration parameters, the segment corresponding to the detected car is projected to the surround map as shown in Fig. 9(c).

Fig. 10 shows the application of motion stereo on the surroundings on the side of the car. The raw images of successive frames from the right omni camera video containing an overtaking car are shown in Fig. 10(a). The processing of these images is shown in the six parts of Fig. 10(b). The top two images in this figure show the virtual right side views obtained by transforming the images using the calibration parameters. Note that the disparity in this case consists of the components due to the depth of the object as well as the independent motion, which could be positive or negative. To use the stereo implementation of [47], which only searches positive disparities, one of the images was translated. The color superimposition of images is shown in mid left, and the disparity map is shown in mid right. Although the car is nearer to the camera than to the background, the independent motion of the overtaking car in the forward direction produces negative overall disparity, which, when biased, becomes a positive value smaller than that from the distant background. The histogram image containing the distribution of disparities of each columns of the image was formed. The histogram image clusters the disparities as shown in the bottom left, with the vehicle and background layers concentrating as line segments. However, regions having no significant features do not get disparity estimates; hence, there are breaks in the segments. Each cluster is backprojected onto the disparity image by selecting the pixels that contributed to the cluster, therefore separating the disparity image into layers. For every layer, the lowest pixel in each column is selected and marked in the obstacle image shown in the bottom right part. The projection of the detected car on the surround map is shown in Fig. 10(c).

Fig. 11 shows the results of surround map generation on snapshots from the video sequence in which vehicles in front approach and then recede from the test-bed vehicle. The analysis of the stereo images is shown in the image sets [Fig. 11(a)], where the top images are the rectified left and right images, the mid left is the composite image, the mid right is the disparity image, the bottom left is the disparity histogram image, and the bottom right is the obstacle image as described before. This and the motion stereo from side views (not shown for brevity) are projected onto the surround map in [Fig. 11(b)], showing the detected cars. The approaching and receding of the detected vehicles is clearly seen. Fig. 12 shows a car overtaking from left. In the first three snapshots, the car is detected in side images using motion stereo. In the last snapshot, the car goes in front view, and binocular stereo detects it. The surround map shows the transition of the detected car from side view to front view.

These experiments show the detection of vehicles in front as well as on the sides. The size of the objects in the image determines the distance at which they can be detected reliably. It was observed that at the distance of 10–15 m, a typical car had size of approximately 16 $\times$ 10 pixels in omni image and 25 $\times$ 16 pixels in the virtual perspective image. The disparity was about 20 pixels, and the object could be detected reliably. However, at distances of 40–50 m, the omni resolution decreased to 5 $\times$ 5 pixels, the virtual image resolution to 10 $\times$ 4 pixels, and the disparity to approximately 6 pixels. The detection at this distance was somewhat sporadic, and the estimates of range and therefore lateral position were less accurate. A higher resolution camera, such as $1K \times 1K$, could be used to improve the accuracy and resolution.

There were some inaccuracies in localization of the vehicles due to the errors in the disparity for front images and error in locating the object cluster in the side images. Fig. 13 shows

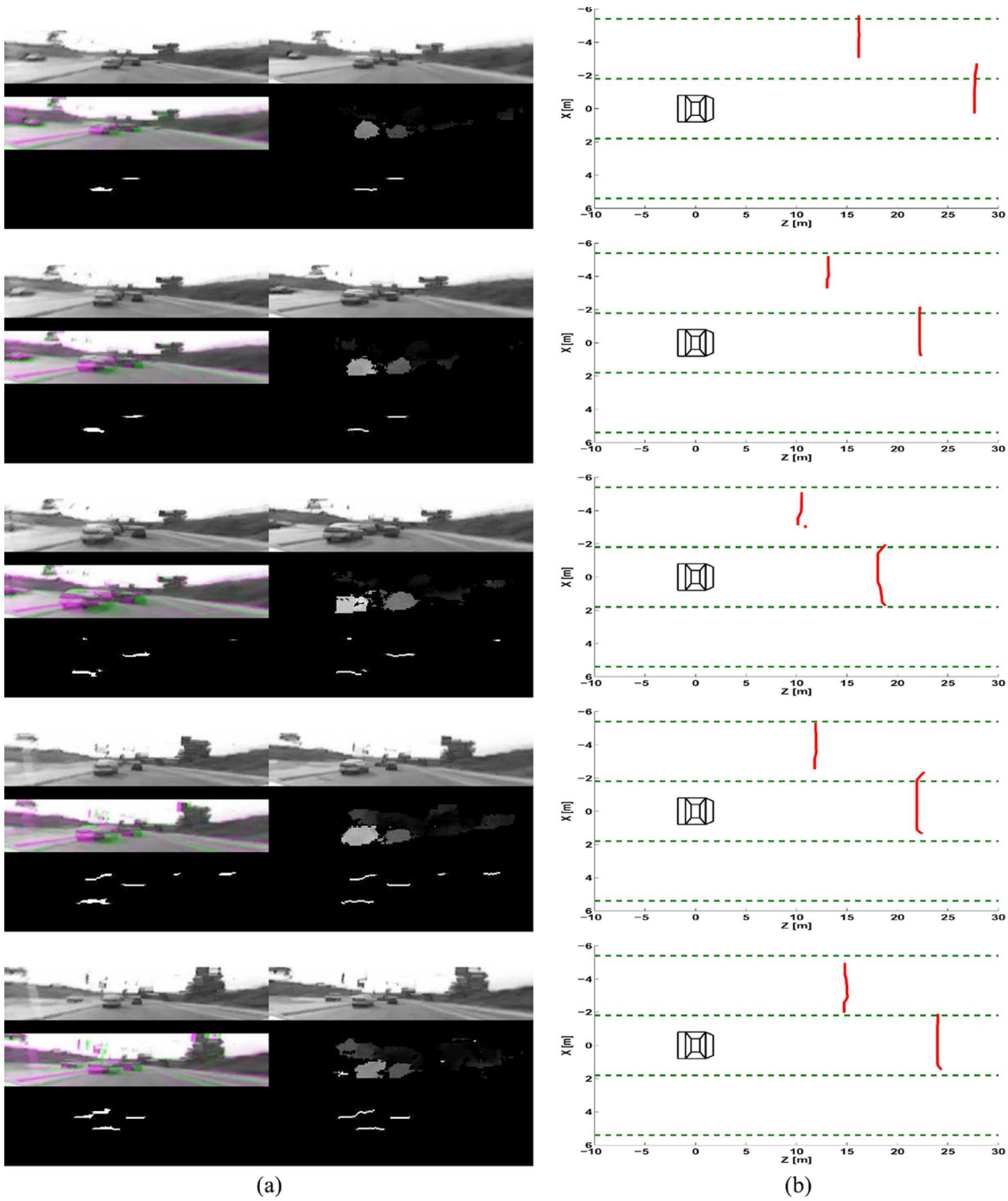


Fig. 11. Results of surround generation for vehicles in front of the car that are approaching and then receding. The virtual front views generated from omni cameras are analyzed using binocular stereo. (a) Steps in detection. Top: Rectified left and right images. Mid left: Composite image. Mid right: Disparity image. Bottom left: Disparity histogram image. Bottom right: Obstacle image. (b) Surround map showing own car (black) and approximate positions of other vehicles (red).

cases where the detection and localization of the objects are degraded. Improvements in the disparity analysis is likely to give better results, which would enable accurate tracking and velocity estimation of the surrounding vehicles.

VII. Concluding Remarks

The objective of active safety systems is to prevent collisions of automobiles. Camera-based systems have an important role in advancing the active safety technology. It is believed that future driver assistance systems would rely on camera-based technology to capture and monitor the immediate surroundings of the vehicle. Such a DPS map may prove to be a crucial element in the development of appropriate warnings about potential collisions with obstacles or vehicles that may even be in the blind spots of the driver. The primary contributions of this paper are (1) motivating the need for dynamic surround capture for driver assistance systems and studies of driver behavior, (2) video-based approaches having relevance to surround capture, and (3) introduction of an omni-video-based

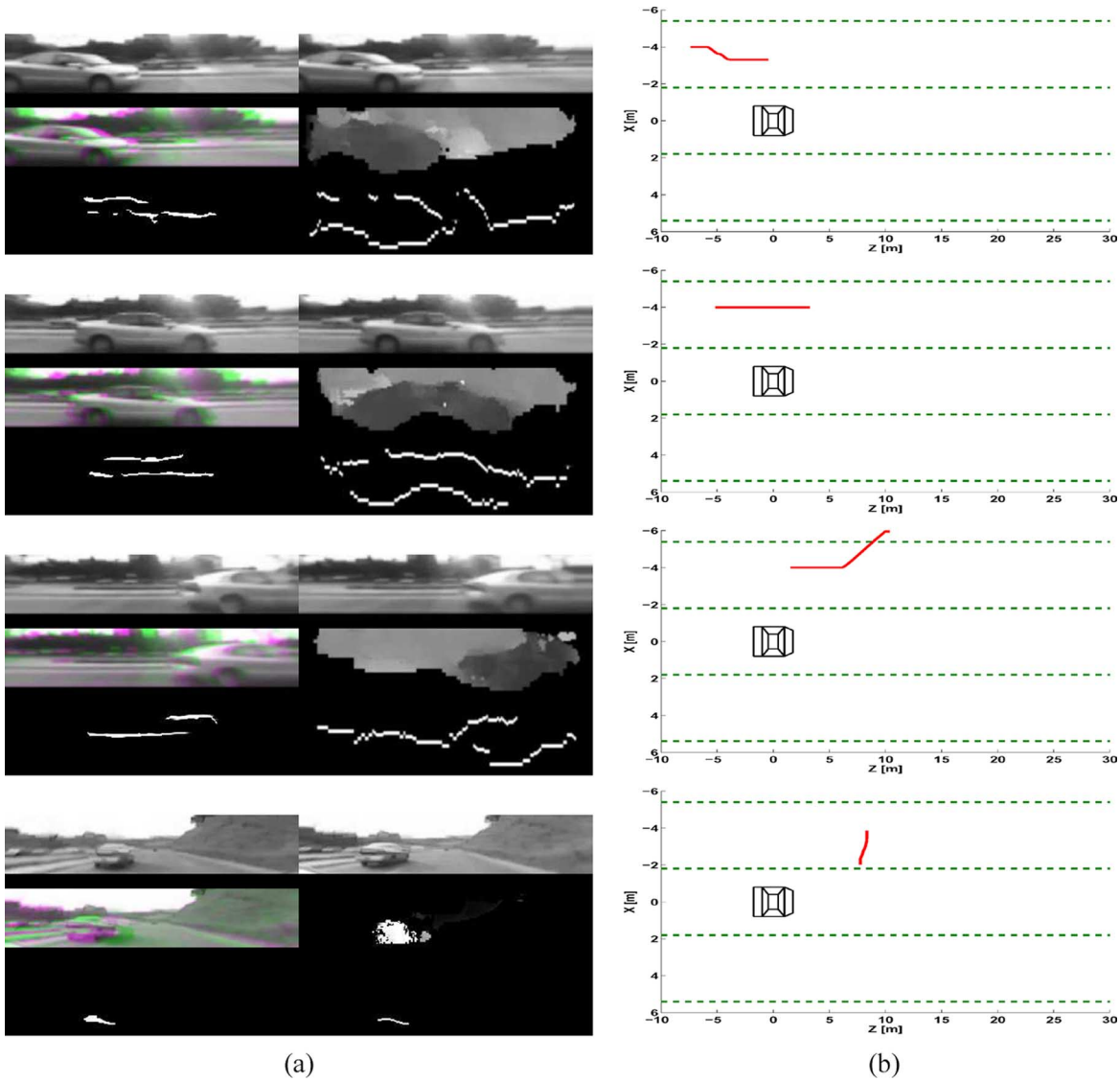


Fig. 12. Results of surround generation detecting a vehicle passing from left. The first three image sets are the left side views analyzed using motion, and the last has the front views generated from the omni cameras and analyzed using stereo. (a) Steps in detection. Top: Rectified left and right images. Mid left: Composite image. Mid right: Disparity image. Bottom left: Disparity histogram image. Bottom right: Obstacle image. (b) Surround map showing own car (black) and approximate positions of other vehicles (red).

approach using stereo and motion for dynamic panoramic map generation.

This paper described the importance of the vehicle surround and discussed the recent research in this field. A novel framework for synthesizing a panoramic map of the vehicle surroundings using a pair of omni cameras mounted on the sides of the vehicle was described. Using the omni camera, virtual perspective images with FOV in any direction can be synthesized. For vehicle detection, binocular stereo was used on the overlapping views of the front of the car, and motion stereo was used for monocular views of the sides of the car. The detected vehicles were inserted into a panoramic surround map. Experimental runs with an instrumented test vehicle were conducted to show the basic feasibility of omni-camera-based surround capture algorithm. It was observed that the resolution of omni cameras is low for the front parts of the image. Due to this, segmentation and depth estimation becomes more challenging. However, the use of only two sensors for analyzing front and side views makes the approach attractive for designers.

In future work, we plan to explore methods for improving the robustness of detection and localization of surrounding objects especially by better handling of featureless regions where disparity estimates are not available. We also intend to integrate motion and stereo for front surround analysis by performing disparity analysis on feature tracks obtained from the two cameras. The calibration is currently performed by manually marking points on the parking lot lines. To automate this procedure, we can use Hough transform or other methods to detect lines. We also plan to explore methods for dynamic correction of calibration using scene epipolar geometry [44] or fiducial marks on car body.

Finally, it is important to acknowledge that an effective and useful approach for surround-based driver assistance system needs systematic and careful "human-factors"-oriented investigations in addition to the development of the novel surround

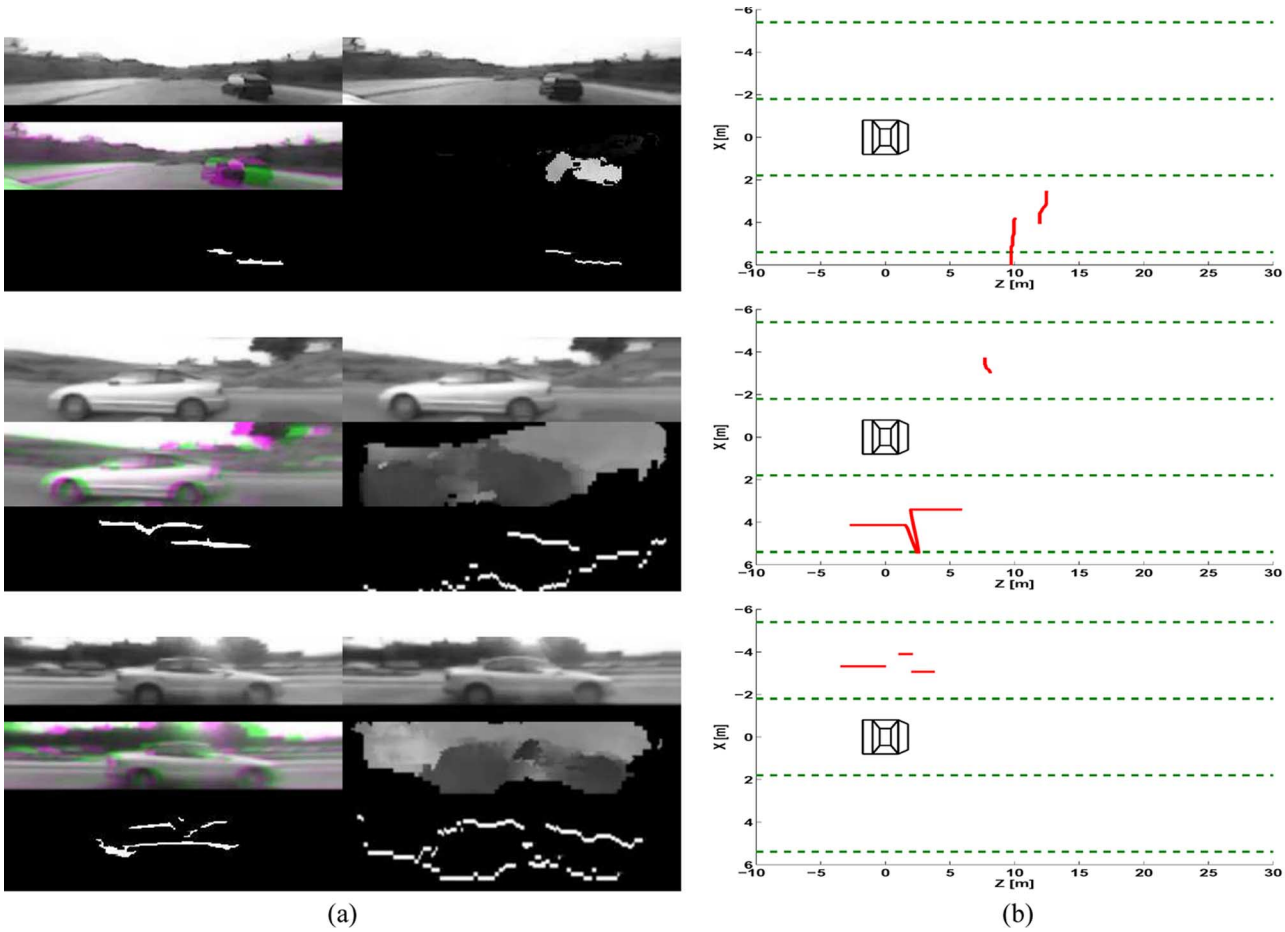


Fig. 13. Detection and localization problems. (a) Splitting of front object into two parts at different distances. Possible reason: The left side of the object car, unlike the front, has variable depth. (b) Gaps and sharp curvature in side-object detection. Possible reason: Featureless parts in the middle of the car give unreliable disparity values.

capture technology. Such multidisciplinary studies consider optimum means for presenting information about potential dangers to the driver in a nondistracting and reliable manner using multiple modes of communication, such as visual, auditory, and haptic. Collaborations with experts from human–machine interactions, cognitive science, and psychology are essential to make progress in this area [5], [49]–[52].

ACKNOWLEDGMENT

The authors would like to thank the Editor and the reviewers for providing insightful comments that helped improve the quality of this paper; their colleagues from the Computer Vision and Robotics Research laboratory for contributions and support; in particular, J. McCall for the help on the hardware of the car test bed and the final proofreading of this paper, as well as S. Cheng and S. Krotosky for the help on the stereo software.

REFERENCES

[1] M. Peden, R. Scurfield, D. Sleet, D. Mohan, A. A. Hyder, E. Jarawan, and C. Mathers, "World report on road traffic injury prevention," World Health Organization (WHO) Press, Tech. Rep. ISBN 92 4 156260 9, 2004. [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., National Highway Traffic Safety Administration, 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] "Automotive collision avoidance systems (ACAS) program: Final report," U.S. Dept. Transp., National Highway Traffic Safety Administration, Washington, DC, Tech. Rep. DOT HS 809 080, May 10, 1998. [Online]. Available: http://www.nhtsa.dot.gov/people/injury/research/pub/ACAS/ACAS_index.htm

[4] 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.

[5] J. McCall, O. Achler, M. M. Trivedi, P. Fastrez, D. Forster, J. B. Haue, and E. B. J. Hollan, "A collaborative approach for human-centered driver assistance systems," in *Proc. 7th IEEE Conf. Intell. Transp. Syst.*, Oct. 3–6, 2004, pp. 663–667.

[6] 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. Veh. Symp.*, Jun. 2004, pp. 483–488.

[7] 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.

[8] 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.

[9] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A survey of video processing techniques for traffic applications," *Image Vis. Comput.*, vol. 21, no. 4, pp. 359–381, Apr. 2003.

[10] *Proc. IEEE Intell. Veh. Symp.*, Las Vegas, NV, Jun. 6–8, 2005.

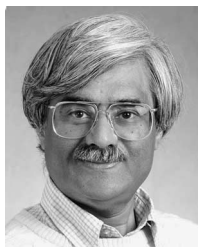
[11] *Proc. IEEE Intell. Veh. Symp.*, Tokyo, Japan, Jun. 13–15, 2006.

[12] *Proc. IEEE Intell. Transp. Syst. Conf.*, Vienna, Austria, Sep. 13–16, 2005.

[13] 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.
[14] ——, "An integrated, robust approach to lane marking detection and lane tracking," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2004, pp. 533–537.
[15] C. Y. Fang, C. S. Fuh, P. S. Yen, S. Cherng, and S. W. Chen, "An automatic road sign recognition system based on a computational model of human recognition processing," *Comput. Vis. Image Underst.*, vol. 96, no. 2, pp. 237–268, Nov. 2004.
[16] C. Y. Fang, S. W. Chen, and C. S. Fuh, "Automatic change detection of driving environments in a vision-based driver assistance system," *IEEE Trans. Neural Netw.*, vol. 14, no. 3, pp. 646–657, May 2003.
[17] W. Kruger, "Robust real time ground plane motion compensation from a moving vehicle," *Mach. Vis. Appl.*, vol. 11, no. 4, pp. 203–212, 1999.
[18] W. Enkelmann, "Video-based driver assistance: From basic functions to applications," *Int. J. Comput. Vis.*, vol. 45, no. 3, pp. 201–221, Dec. 2001.
[19] K. Fleischer, H. H. Nagel, and T. Rath, "3-D-Model-based-vision for innercity driving scenes," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2002, vol. 2, pp. 477–482.
[20] A. Broggi, P. Cerri, and P. Antonello, "Multi-resolution vehicle detection using artificial vision," in *Proc. IEEE Intell. Veh. Symp.*, Parma, Italy, Jun. 2004, pp. 310–314.
[21] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using evolutionary Gabor filter optimization," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 125–137, Jun. 2005.
[22] Q. T. Luong, J. Weber, D. Koller, and J. Malik, "An integrated stereo-based approach to automatic vehicle guidance," in *Proc. Int. Conf. Comput. Vis.*, Jun. 1995, pp. 52–57.
[23] 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.
[24] A. Bensrhair, A. Bertozzi, A. Broggi, A. Fascioli, S. Mousset, and G. Toulminet, "Stereo vision-based feature extraction for vehicle detection," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2002, vol. 2, pp. 465–470.
[25] R. Labayrade, D. Aubert, and J. P. Tarel, "Real time obstacle detection in stereo vision on non flat road geometry through V-disparity representation," in *Proc. IEEE Intell. Veh. Symp.*, vol. II, 2002, pp. 646–651.
[26] D. M. Gavrila and V. Philomin, "Real-time object detection for smart vehicles," in *Proc. Int. Conf. Comput. Vis.*, Kerkyra, Greece, 1999, pp. 87–93.
[27] D. Gavrila, J. Giebel, and S. Munder, "Vision-based pedestrian detection: The PROTECTOR system," in *Proc. IEEE Intell. Veh. Symp.*, Parma, Italy, Jun. 2004, pp. 13–18.
[28] G. Grubb, A. Zelinsky, L. Nilsson, and M. Rilbe, "3-D vision sensing for improved pedestrian safety," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2004, pp. 19–24.
[29] X. Liu and K. Fujimura, "Pedestrian detection using stereo night vision," *IEEE Trans. Veh. Technol.*, vol. 53, no. 6, pp. 1657–1665, Nov. 2004.
[30] A. Broggi, A. Fascioli, M. Carletti, T. Graf, and M. Meinecke, "A multi-resolution approach for infrared vision-based pedestrian detection," in *Proc. IEEE Intell. Veh. Symp.*, Parma, Italy, Jun. 2004, pp. 7–12.
[31] F. Xu, X. Liu, and K. Fujimura, "Pedestrian detection and tracking with night vision," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 1, pp. 63–71, Mar. 2005.
[32] T. Rabie, G. Auda, A. El-Rabbany, A. Shalaby, and B. Abdulhai, "Active-vision-based traffic surveillance and control," in *Proc. 14th Can. Vis. Interface Conf.*, Jun. 2001, pp. 87–93.
[33] T. Kato, Y. Ninomiya, and I. Masaki, "An obstacle detection method by fusion of radar and motion stereo," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 3, pp. 182–188, Sep. 2002.
[34] Y. Fang, I. Masaki, and B. Horn, "Depth-based target segmentation for intelligent vehicles: Fusion of radar and binocular stereo," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 3, pp. 196–202, Sep. 2002.
[35] S. Mota, E. Ros, E. M. Ortigosa, and F. J. Pelayo, "Bio-inspired motion detection for blind spot overtaking monitor," *Int. J. Robot. Autom.—Special Issue of Neuromorphic Systems*, vol. 19, no. 4, p. 2713, 2004.
[36] M. K. Sergi, "Bus rapid transit technologies: A virtual mirror for eliminating vehicle blind zones," Intell. Veh. Lab, Dept. Mech. Eng., Univ. Minnesota, Minneapolis, Tech. Rep. CS 04-12, Nov. 2003.
[37] C. Stiller, J. Hipp, C. Rossig, and A. Ewald, "Multisensor obstacle detection and tracking," *Image Vis. Comput.*, vol. 18, no. 5, pp. 389–396, Apr. 2000.
[38] O. Achler and M. M. Trivedi, "Vehicle wheel detector using 2-D filter banks," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2004, pp. 25–30.
[39] T. Gandhi and M. M. Trivedi, "Parametric ego-motion estimation for vehicle surround analysis using an omnidirectional camera," *Mach. Vis. Appl.*, vol. 16, no. 2, pp. 85–95, Feb. 2005.
[40] L. Matuszyk, A. Zelinsky, L. Nilsson, and M. Rilbe, "Stereo panoramic vision for monitoring vehicle blind-spots," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2004, pp. 31–36.
[41] K. Huang, M. M. Trivedi, and T. Gandhi, "Driver's view and vehicle surround estimation using omnidirectional video stream," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2003, pp. 444–449.
[42] J. Rebut, G. Toulminet, and A. Bensrhair, "Road obstacles detection using a self-adaptive stereo vision sensor: A contribution to the ARCOS French project," in *Proc. IEEE Intell. Veh. Symp.*, Parma, Italy, Jun. 2004, pp. 738–743.
[43] R. Benosman and S. B. Kang, *Panoramic Vision: Sensors, Theory, and Applications*. New York: Springer-Verlag, 2001.
[44] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision*, 1st ed. ser. Interdisciplinary Applied Mathematics, vol. 26. New York: Springer-Verlag, 2004.
[45] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, no. 1–3, pp. 7–42, Apr./Jun. 2002.
[46] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 993–1008, Aug. 2003.
[47] K. Konolige, "Small vision system: Hardware and implementation," in *Proc. 8th Int. Symp. Robot. Res.*, Oct. 1997, pp. 111–116. [Online]. Available: http://www.ai.sri.com/ konolige/papers
[48] S. B. Marapane and M. M. Trivedi, "Multi-primitive hierarchical (MPH) stereo analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 3, pp. 227–240, Mar. 1994.
[49] J. Levy and H. Pashler, "Is dual-task slowing instruction dependent?," *J. Exp. Psychol., Human Percept. Perform.*, vol. 27, no. 4, pp. 862–869, Aug. 2001.
[50] M. A. Recarte and L. M. Nunes, "Mental workload while driving: Effects on visual search, discrimination, and decision making," *J. Exp. Psych., Applied*, vol. 9, no. 2, pp. 119–137, Jun. 2003.
[51] P. Green, "Driver distraction, telematics design, and workload managers: Safety issues and solutions," in *Proc. Int. Congr. Transp. Electron*. Warrendale, PA: Soc. Automotive Eng., 2004, pp. 165–180. ser. P-387.
[52] N. Kuge, T. Yamamura, O. Shimoyama, and A. Liu, "A driver behavior recognition method based on a driver model framework," presented at the SAE World Congr., Session: Human Centered Driver Assistance Systems, Detroit, MI, Mar. 2000, Paper No. 2001-01-0349.

Tarak Gandhi received the B.Tech. 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, specializing in computer vision, from Pennsylvania State University, University Park.

He was with Adept Technology, Inc., where he worked on designing algorithms for robotic systems. Currently, he is a Postdoctoral Scholar with the Computer Vision and Robotics Research Laboratory, University of California, San Diego, La Jolla. 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.



Mohan Manubhai Trivedi 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. He serves regularly as a consultant to industry and government agencies in the USA and abroad. His research interests are in the areas of intelligent systems, computer vision systems, intelligent ("smart") environments, intelligent vehicles, transportation systems, and human–machine interfaces.

Prof. Trivedi serves on the Executive Committees of the University of California Digital Media Innovation Program and the California Institute for Telecommunication and Information Technologies [Cal-IT2] as the leader of the Intelligent Transportation and Telematics Layer, UCSD. He has received the Distinguished Alumnus Award from the Utah State University and the Pioneer Award (Technical Activities) and the Meritorious Service Award from the IEEE Computer Society.