Collision Sensing by Stereo Vision and Radar Sensor Fusion

Article *in* IEEE Transactions on Intelligent Transportation Systems · January 2010

DOI: 10.1109/TITS.2009.2032769 · Source: IEEE Xplore

5 authors, including:

Shunguang Wu
Johns Hopkins University
51 PUBLICATIONS **307** CITATIONS
SEE PROFILE

Stephen Decker
General Motors Company
4 PUBLICATIONS **69** CITATIONS
SEE PROFILE

Peng Chang
Princeton vision
9 PUBLICATIONS **329** CITATIONS
SEE PROFILE

Jayan Eledath
Amazon.com
31 PUBLICATIONS **203** CITATIONS
SEE PROFILE

Collision Sensing by Stereo Vision and Radar Sensor Fusion

Shunguang Wu, *Member, IEEE*, Stephen Decker, *Member, IEEE*, Peng Chang, *Member, IEEE*, Theodore Camus, *Senior Member, IEEE*, and Jayan Eledath, *Member, IEEE*

***Abstract*—To take advantage of both stereo cameras and radar, this paper proposes a fusion approach to accurately estimate the location, size, pose, and motion information of a threat vehicle with respect to a host one from observations that are obtained by both sensors. To do that, we first fit the contour of a threat vehicle from stereo depth information and find the closest point on the contour from the vision sensor. Then, the fused closest point is obtained by fusing radar observations and the vision closest point. Next, by translating the fitted contour to the fused closest point, the fused contour is obtained. Finally, the fused contour is tracked by using rigid body constraints to estimate the location, size, pose, and motion of the threat vehicle. Experimental results from both synthetic data and real-world road test data demonstrate the success of the proposed algorithm.**

***Index Terms*—Collision sensing, extended target tracking, vision–radar fusion.**

I. Introduction

ADVANCED driving-assistant systems (ADASs) have drawn much attention in the past two decades [1]. These systems include lateral guidance assistance, adaptive cruise control (ACC), collision sensing/avoidance, urban driving and stop-and-go situation detection, lane-change assistance, traffic sign recognition, high-beam automation, and fully autonomous driving. Basically, the success of these systems depends on accurately sensing the spatial and temporal environment information of a host vehicle with a low false-alarm (FA) rate. This information includes present and future road/lane status, such as curvatures and boundaries, and the location and motion information of on-road/off-road obstacles, including vehicles, pedestrians, and the background.

For some systems, *point-object*-based modeling and tracking approaches can be employed, whereas for others, an *extended object* detection and tracking algorithm must be considered. To find the location, orientation, and motion information of an extended object, it is necessary to both measure the accurate distance from the host vehicle to the object and estimate the boundary positions and pose of the object. To achieve this goal, it is insufficient to use only one available sensing method such as radar [2], [3], mono vision [4], stereo vision [5], [6], or laser radar (ladar) [7].

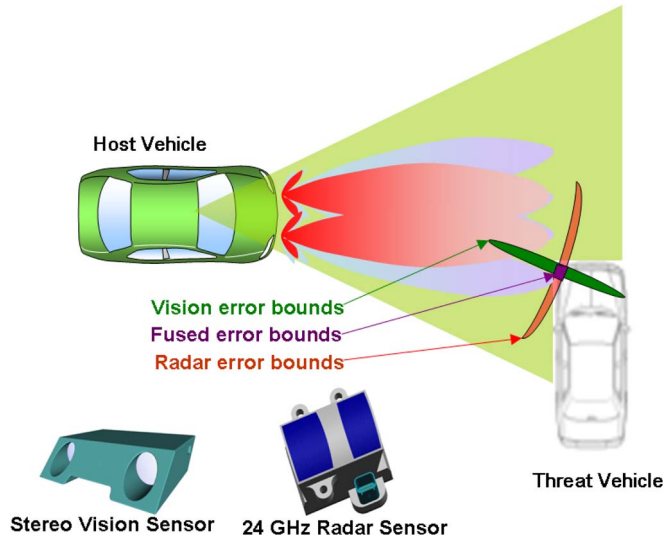

Fig. 1. Sketch showing the pros and cons of both stereo vision and radar sensors, as well as the expected fusion results in the estimation of the threat vehicle location, size, pose, and motion information.

It is well known that radar has the ability to measure the accurate distance to an object and is robust in bad weather, but it does not have high lateral resolution. (The reasons for noise radar azimuth measurement are radar sensor limitation and the fact of having a nonfixed reflection point on the rear part of the threat vehicle.) On the contrary, a vision sensor has sufficient lateral resolution to find the boundaries of an object. However, the distance measurement by a stereo-vision system is less accurate than radar. Although expensive laser-scanning radar can detect the occupying area of an object, the affordable automotive ladar can only reliably detect reflectors. As a result, it has no ability to find all occupying areas. Therefore, it seems that the only way to give a better solution to an ADAS problem is to fuse multimodality sensors.

This paper addresses the problem of accurately estimating the location, size, pose, and motion information of a threat vehicle from both stereo vision and radar sensors located in a host one. As shown in Fig. 1, a stereo-vision sensor can provide extended target detection with high azimuth resolution but has noisy range information, whereas a radar sensor reports

Manuscript received August 31, 2008; revised May 12, 2009 and July 16, 2009. First published October 30, 2009; current version published December 3, 2009. This work was supported by the National Institute of Standards and Technology under Cooperative Agreement 70NANB4H3044. The Associate Editors for this paper were B. De Schutter and S. Shladover.

S. Wu, T. Camus, and J. Eledath are with Vision Technologies, Sarnoff Corporation, Princeton, NJ 08540 USA (e-mail: swu@sarnoff.com; tcamus@sarnoff.com; jeledath@sarnoff.com).

S. Decker was with the Advanced Sensor Development, Autoliv Inc., Southfield, MI 48034 USA (e-mail: stephen.decker@sbcglobal.net).

P. Chang was with Vision Technologies, Sarnoff Corporation, Princeton, NJ 08540 USA. He is now with General Vision LLC, Princeton, NJ 08540 USA (e-mail: pengchang03@yahoo.com).

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

Digital Object Identifier 10.1109/TITS.2009.2032769

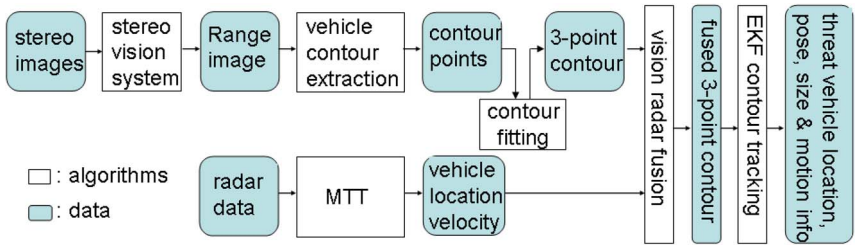

Fig. 2. Block diagram of the vision–radar fusion system.

accurate range and relative speed but noisy azimuth angle observations. By taking into account the advantages of both sensors, the proposed fusion approach can accurately estimate the location, size, pose, and motion information of the threat vehicle up to 20 m.

A. Related Work

Based on the classification of multisensor data fusion [8], most radar and vision fusion algorithms in the literature are of level 1, which includes sublevels of alignment, association, correlation, correlator tracker, and classification. Reference [9] presented a correlator-tracker-level fusion on blobs that are detected in both a single radar image and a single night-vision image; the fusion is performed in polar coordinates and based on an angular position. Reference [10] is another example of a correlator-tracker-level fusion in which, for each target, the fusion system fuses the results of four different image processing algorithms and radar information by automatically combining 12 different features and generating many possible target position proposals. A belief network is generated to organize features and position proposals. An inference algorithm is then used to find out the actual position of the target by deducing which observations are wrong.

For most of the classification-level fusion approaches, a radar system reports a list of targets with range and rough azimuth, and the image information from a monocular vision system is then used for precise object localization, validation, and/or classification. For example, [11] and [12] reported a fusion approach combining radar and monocular image processing to improve both longitudinal and lateral position estimations in an ACC system; it can track vehicles up to a distance of 130 m as well as reliably assign them to specific lanes. Reference [13] introduced a fusion system for detecting, validating, and tracking the preceding vehicle by visually processing the images from a single camera. Reference [14] presented a system that uses a radar-cue-guided vision classification process to validate, characterize, and predict the in-lane primary target. The validation process decides if the target is indeed in the host vehicle lane and represents an obstacle in the host vehicle path. The characterization task aims to predict if the primary target is going to leave the host lane. The system was applied on the forward-collision problem. Reference [15] presented a method to estimate the distance and left and right boundaries of an object by fusing radar and motion stereo observations on an urban road. Since the method does not depend on the appearance of objects, it is capable of detecting an automobile and other objects up to 50 m. For collision warning, [16] described a fusion system that combines vision data from a single camera with radar data for real-time forward-collision warning. This approach fuses radar data with "vision-based object detection" and "overhead sign and structure detection" by using a probabilistic framework with coregistered radar data to reliably obtain the vehicle azimuth and depth by minimizing FAs.

Only a few papers propose systems for fusing the observations from radar and stereo cameras. Reference [17] proposed a sensor fusion method that can make use of radar coarse target depth information to segment target locations in video images. This segmentation method splits an edge map of a binocular image into several layers corresponding to the given radar depth information; hence, different layers contain the edge pixels of targets at different depth ranges. As a result, the original multiple-target segmentation task is decomposed into several simpler and easier single-target segmentation tasks on each depth-based target feature layer, thus improving the segmentation performance.

In addition, from an implementation point of view, most of the fusion approaches were implemented under a Kalman filtering or Bayesian estimation framework, but the modular Bayesian networks were also taken into account by fusing radar and single-camera images to detect threat vehicles [18].

Compared with other fusion algorithms, instead of taking a threat vehicle as a point target to track up to 130 m, we model it as an extended rigid body object and propose a method to simultaneously precisely estimate the location, size, pose, and motion features of the threat vehicles within ranges up to 20 m.

II. System Overview

Fig. 2 presents the block diagram of our vision–radar fusion system. The inputs of the system are the left- and right-stereo images and the corresponding ranges and azimuths of the radar targets at that frame. The contours of the threat vehicles can be extracted from the range image [19]. By using the contour-fitting algorithm, a contour is represented by three points, i.e., the left, middle, and right points of two perpendicular line segments for a two-side-view scenario, whereas for the one-side-view case, they stand for the left, middle, and right points of a line segment. On the other hand, the raw radar data are sent into a multitarget tracking (MTT) algorithm to estimate the location and the velocities of each radar target. During the fusion process, three-point contours and MTT outputs are combined to give more accurate three-point contours. Finally, an extended Kalman filter is designed to track the contours to estimate their location, size, pose, and motion parameters.

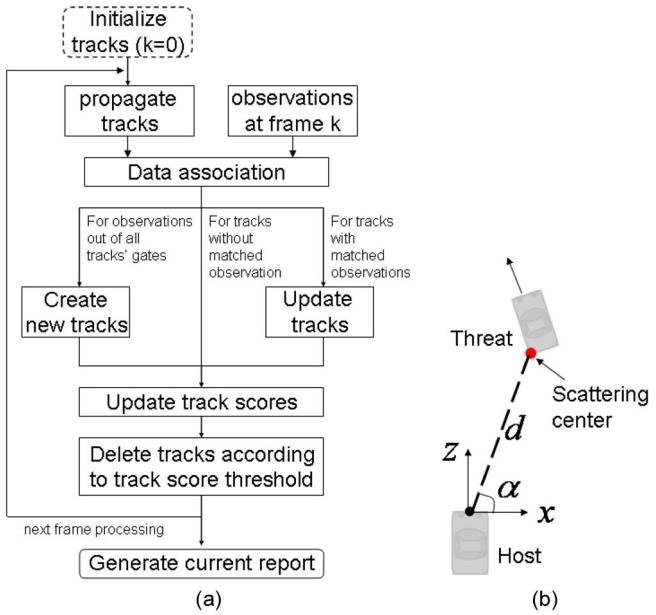

Fig. 3. (a) Block diagram of the radar multiple-target tracking algorithm. (b) Radar coordinate system in the bird's-eye view and definitions of the measurements in (4).

III. Main Algorithms

A. Radar Target Tracking

The radar used in our experiment can report as many as 15 pairs of range–azimuth observations with ID information at each scan. Each range–azimuth pair represents the location of a scattering center (SC). The SCs may or may not be from the same threat car. An MTT algorithm has been developed to estimate the locations and the velocities of the SCs, and the algorithm has the ability to dynamically maintain (create/delete) the tracked SCs by evaluating their track scores.

Fig. 3(a) shows the block diagram of the MTT algorithm. It is composed of the blocks of initialization, prediction, data association, track update, track health monitoring, and track management.

The state vector of one SC/track at time t_k is defined by

$$\mathbf{x}_k^{(r)} = [x, \dot{x}, z, \dot{z}]_k^T \tag{1}$$

where T stands for the transpose operation, and (x, z) and $(\dot{x}, \dot{z})$ are the location and the velocity, respectively, of the SC in the radar coordinate system, which is mounted on the host vehicle and illustrated in Fig. 3(b). The constant-velocity model is used to describe the kinematics of the SC, i.e.,

$$\mathbf{x}_{k+1}^{(r)} = \mathbf{F}_k^{(r)} \mathbf{x}_k^{(r)} + \mathbf{v}_k^{(r)} \tag{2}$$

where $\mathbf{v}_k^{(r)} \sim N(0, \mathbf{Q}_k^{(r)})$, $\mathbf{F}_k^{(r)} = \text{diag}\{\mathbf{F}_0, \mathbf{F}_0\}$, and $\mathbf{Q}_k^{(r)} = \text{diag}\{(\sigma_x^{(r)})^2 \mathbf{Q}_0, (\sigma_z^{(r)})^2 \mathbf{Q}_0\}$, with

$$\mathbf{F}_0 = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix}$$
$$\mathbf{Q}_0 = \begin{pmatrix} \frac{(\Delta t)^4}{4} & \frac{(\Delta t)^3}{2} \\ \frac{(\Delta t)^3}{2} & (\Delta t)^2 \end{pmatrix} \tag{3}$$

where $\Delta t = t_{k+1} - t_k$, and $(\sigma_x^{(r)}, \sigma_z^{(r)})$ are the standard deviations (STDs) of noise intensities in the x- and z-directions, respectively.

With the illustration in Fig. 3(b), the measurement state vector is defined by

$$\mathbf{z}_k^{(r)} = [d, \alpha]_k^T \tag{4}$$

and the measurement equation is

$$\mathbf{z}_k^{(r)} = \mathbf{h}^{(r)} \left(\mathbf{x}_k^{(r)} \right) + \mathbf{w}_k^{(r)} \tag{5}$$

with

$$\mathbf{h}^{(r)} \left(\mathbf{x}_k^{(r)} \right) = \begin{cases} h_1^{(r)} = \sqrt{x_k^2 + z_k^2} \\ h_2^{(r)} = \tan^{-1}(z_k / x_k) \end{cases}$$

and $\mathbf{w}_k^{(r)} \sim N(0, \mathbf{R}_k^{(r)})$. Assume that the measurement noise STDs are $(\sigma_d, \sigma_\alpha)$; then, $\mathbf{R}_k^{(r)} = \text{diag}(\sigma_d^2, \sigma_\alpha^2)$.

In general, any nonlinear filtering algorithm can be employed to estimate the state vector $\mathbf{x}_k^{(r)}$ and its covariance by giving the observation history up to time t_k. In this paper, the standard extended Kalman filtering (EKF) algorithm was implemented. Additionally, by giving an observation of distance (d_0) and azimuth (α_0) at frame zero, the initial EKF state vector can be set as $\mathbf{x}_0^{(r)} = [d_0 \cos(\alpha_0), 0, d_0 \sin(\alpha_0), 0]^T$, and its covariance is set as a diagonal matrix with relatively large variances to compensate the zero-velocity initial values.

To evaluate the health status of each track, the track score of each SC is monitored. Assume that M is the measurement vector dimension, P_d is the detection probability, P_{FA} is the FA probability, β_{NT} is the new target density, and V_c is the measurement volume element such that independent true-target detection and FA events occur within each volume. The track score can be initialized as [20]

$$L(k=0) = \ln(\beta_{\text{NT}} V_c) + \ln \frac{P_d}{P_{\text{FA}}} \tag{6}$$

and it can be updated by

$$L(k) = L(k-1) + \Delta L(k) \tag{7}$$

where

$$\Delta L(k) = \begin{cases} \ln(1 - P_d), & \text{if track is not updated on scan } k \\ \Delta L_k + \Delta L_s, & \text{otherwise} \end{cases}$$
$$\Delta L_k = \ln \left(\frac{V_c}{\sqrt{\det(\mathbf{S})}} - \frac{1}{2} \left(M \ln(2\pi) + \tilde{\mathbf{z}}^T \mathbf{S}^{-1} \tilde{\mathbf{z}} \right) \right)$$
$$\Delta L_s = \ln \left(\frac{P_d}{P_{\text{FA}}} \right). \tag{8}$$

$\tilde{\mathbf{z}}$ and $\mathbf{S}$ are the measurement innovation and its covariance, respectively.

Once we have the evolution curve of the track score, a track can be deleted if $L(k) - L_{\max} < \text{THD}$, where $L_{\max}$ is the maximum track score until t_k, and THD is a track deletion threshold.

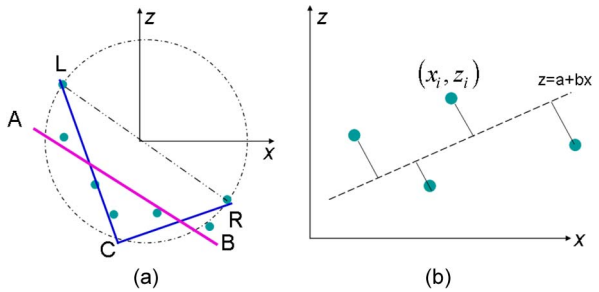

Fig. 4. (a) Two candidate contours, i.e., the line segment AB and the two perpendicular line segments LCR, can be fitted from the given points. (b) Perpendicular line fitting algorithm: Find line $z = a + bx$ from the sample points (x_i, z_i) to minimize the sum of their distances to the line.

B. Contour Point Extraction and Contour Fitting

The contour of a threat vehicle is composed of the points on the boundary of the vehicle in the host coordinate system. As shown in Fig. 3(b), we model both the host and threat vehicles as 2-D rectangular boxes in the bird's-eye-viewing plane, and by considering the fact that only one or two sides of the threat vehicle can be seen by the cameras on the host one, we can use either one line segment or two perpendicular line segments to represent the contour.

The contour points are reported from the *threat detection and segmentation* module [19]. Basically, the input depth image from the stereo cameras is represented with a grid of planar patches. Each patch is labeled as predefined types according to its position and normal vector; then, a grouping algorithm is used to group the small patches together to form the representation for the foreground object. Finally, the x and z values of the contour points are calculated from the centers of the patches that are located on the boundary of the group.

As shown in Fig. 4(a), the contour of a threat vehicle can be represented by either one line segment or two perpendicular line segments (depending on the pose of a threat vehicle in the host vehicle reference system). The contour fitting algorithm fits the line segments from a set of vision contour points such that the sum of perpendicular distances to the line(s) is minimized.

1) Fit One Line: Assume that a set of points (x_i, z_i) $(i = 1, \ldots, n)$ is given; the fitting algorithm estimates line $z = a + bx$ such that the sum of perpendicular distances to the line is minimized [see Fig. 4(b)], i.e.,

$$D = \sum_{i=1}^{n} \frac{|z_i - a - bx_i|}{\sqrt{1 + b^2}}. \tag{9}$$

By taking a square for both sides of (9) and letting $\partial D^2/\partial a = 0$ and $\partial D^2/\partial b = 0$, we have [21]

$$a = \bar{z} - b\bar{x} \quad b = -B \pm \sqrt{B^2 + 1}$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$$

$$B = \frac{\left(\sum_{i=1}^{n} z_i^2 - n\bar{z}^2\right) - \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)}{2\left(n\bar{x}\bar{z} - \sum_{i=1}^{n} x_i z_i\right)}.$$

2) Fit Two Lines: Once we have the component for fitting one line segment, the two-line-segment fitting can be done by the following two steps: First, find the leftmost and rightmost points, i.e., L and R, and make a circle in which LR is its diameter; then, discretely search for the best point C under which the sum of the perpendicular errors to lines LC and RC is the smallest.

With the above-fitted two candidate contours, one can choose the final one by comparing the weighted fitting errors. In this paper, the weights for one-line and two-line selections are 0.35 and 0.65, respectively. Moreover, to conveniently show the fusion algorithm in the next section, we always use two line segments (or three points, i.e., p_L, p_C, and p_R) to represent the contour, regardless if it is from one or two line segment(s). If it is from one line segment, p_c is its middle point.

C. Vision–Radar Fusion

Once we have the fitted contour of a threat vehicle and filtered radar objects, as shown in Fig. 5, the fusion algorithm tries to adjust the location of the contour by using radar information. It can be done by the following steps.

1) Find the Vision Closest Point: Since the contour is represented by two line segments defined by three points p_L, p_C, and p_R, the closest point, say, p_v, can be chosen by comparing the two candidate closest points from the origin with line segments $p_L p_C$ and $p_C p_R$, respectively.

2) Select the Candidate Radar Object: The candidate radar object, say, p_r, can be selected from all radar objects by comparing the Mahalanobis distances from the vision closest point to the radar objects by the nearest neighbor method.

3) Fuse the Closest Point: Assume that the ranges and the azimuth angles of the vision closest point and radar object can be respectively expressed as $(d_v + \delta_{d_v}, \alpha_v + \delta_{\alpha_v})$ and $(d_r + \delta_{d_r}, \alpha_r + \delta_{\alpha_r})$, with $\delta_{d_v} \sim N(0, \sigma_{d_v})$, $\delta_{\alpha_v} \sim N(0, \sigma_{\alpha_v})$, $\delta_{d_r} \sim N(0, \sigma_{d_r})$, and $\delta_{\alpha_r} \sim N(0, \sigma_{\alpha_r})$. The fused range and its uncertainty are expressed as

$$d_f = \frac{d_v \sigma_{d_r} + d_r \sigma_{d_v}}{\sigma_{d_r} + \sigma_{d_v}}$$
$$\sigma_{d_f} = \frac{\sigma_{d_r} \sigma_{d_v}}{\sigma_{d_r} + \sigma_{d_v}}. \tag{10}$$

In our system, the values of σ_{d_r} and σ_{α_r} are directly from the radar sensor, whereas σ_{d_v} and σ_{α_v} can be estimated from vision-estimation errors σ_x and σ_z. The fused azimuth angle and its uncertainty can be calculated in a similar way.

4) Fuse the Contour: Once we have the fused closest point p_f, the fused contour can be obtained by translating the fitted contour from p_v to p_f.

D. Contour Tracking

Since FAs and outliers may exist in radar and vision processes, the fused contour needs to be filtered before being reported to the collision-avoidance algorithm. To this end, an EKF is employed to track the fused contour of a threat vehicle. As shown in Fig. 5(b), the state vector of a contour is defined as

$$\mathbf{x}_k^{(c)} = [x_c, \dot{x}_c, z_c, \dot{z}_c, r_L, r_R, \theta, \dot{\theta}]_k^T \tag{11}$$

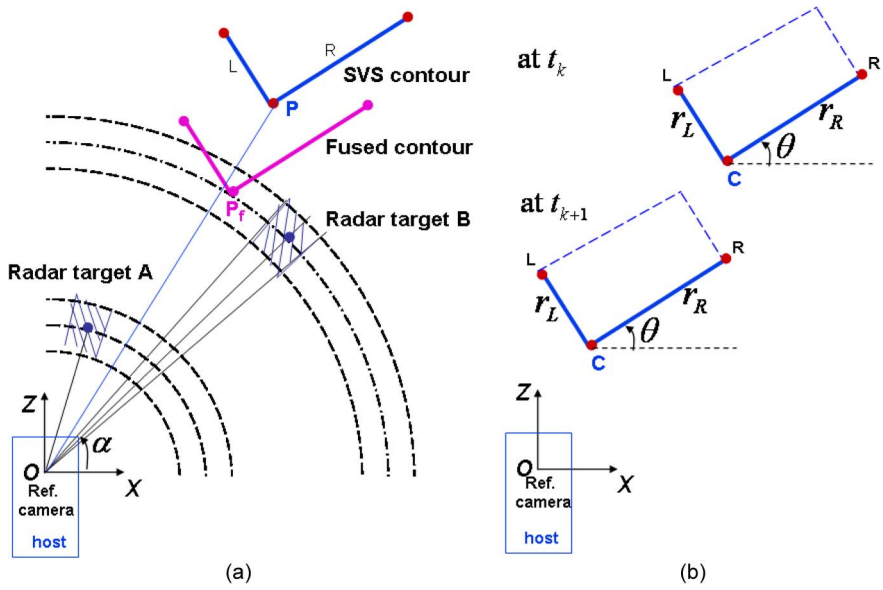

Fig. 5. (a) Sketches of the vision–radar fusion idea. (b) Contour tracking state vector and its modeling.

where c is the intersection point of the two perpendicular line segments if the contour is represented by two perpendicular lines; otherwise, it stands for the middle of the one line segment. $[x_c, z_c]$ and $[\dot{x}_c, \dot{z}_c]$ are the location and the velocity of point c in the host reference system, respectively; r_L and r_R are the left- and right-side lengths of the vehicle, respectively; θ is the pose of the threat vehicle w.r.t. the x-direction of the host reference system; and $\dot{\theta}$ stands for the pose rate.

By considering the rigid body constraint, the motion of the threat vehicle in the host reference system can be modeled as a translation of point c in the xz plane and a rotation w.r.t. the y-axis, which is defined down to the ground in a bird's-eye view. In addition, assuming that the constant velocity model holds between two consecutive frames for both translation and rotation motion, the kinematic equation of the system can be expressed as

$$\mathbf{x}_{k+1}^{(c)} = \mathbf{F}_k^{(c)} \mathbf{x}_k^{(c)} + \mathbf{v}_k^{(c)} \tag{12}$$

where $\mathbf{v}_k^{(c)} \sim N(0, \mathbf{Q}_k^{(c)})$, and

$$\mathbf{F}_k^{(c)} = \text{diag}\{\mathbf{F}_0, \mathbf{F}_0, \mathbf{I}_2, \mathbf{F}_0\} \tag{13}$$

$$\mathbf{Q}_k^{(c)} = \text{diag}\left\{\left(\sigma_x^{(c)}\right)^2 \mathbf{Q}_0, \left(\sigma_z^{(c)}\right)^2 \mathbf{Q}_0, \sigma_r^2 \mathbf{I}_2, \sigma_\theta^2 \mathbf{Q}_0\right\}. \tag{14}$$

In (13) and (14), I_2 is a 2×2 identity matrix, $\mathbf{F}_0$ and $\mathbf{Q}_0$ are defined in (3), and $\sigma_x^{(c)}$, $\sigma_z^{(c)}$, σ_r, and σ_θ are the contour state uncertainty parameters, which can be tuned in during the system performance-evaluation process.

On the other hand, since the positions of the three points L, C, and R can be measured from the fusion block, the observation state vector is

$$\mathbf{z}_k^{(c)} = [x_L, z_L, x_C, z_C, x_R, z_R]_k. \tag{15}$$

According to the geometry, the measurement equation can explicitly be written as

$$\mathbf{z}_k^{(c)} = \mathbf{h}^{(c)}\left(\mathbf{x}_k^{(c)}\right) + \mathbf{w}_k^{(c)} \tag{16}$$

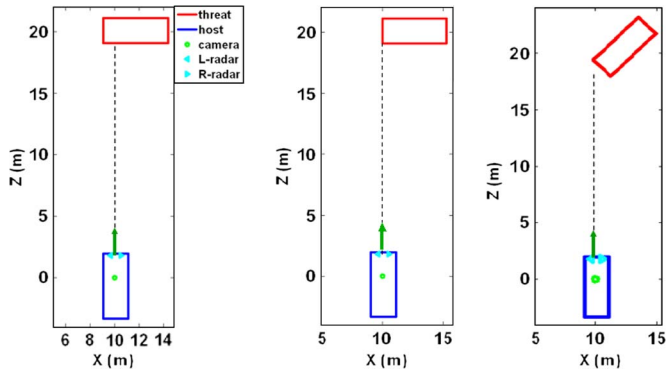

Fig. 6. Sketches of three simulation scenarios.

with

$$\mathbf{h}^{(c)}\left(\mathbf{x}_k^{(c)}\right) = \begin{cases} h_1^{(c)} = x_c - r_L \sin(\theta) \\ h_2^{(c)} = z_c + r_L \cos(\theta) \\ h_3^{(c)} = x_c \\ h_4^{(c)} = z_c \\ h_5^{(c)} = x_c + r_R \cos(\theta) \\ h_6^{(c)} = z_c + r_R \sin(\theta) \end{cases}$$

and $\mathbf{w}_k^{(c)} \sim N(0, \mathbf{R}_k^{(c)})$. Assume that the measurement noise STDs in the x- and z-directions are (σ_x, σ_z); then, $\mathbf{R}_k^{(c)} = \text{diag}(\sigma_x^2, \sigma_z^2, \sigma_x^2, \sigma_z^2, \sigma_x^2, \sigma_z^2)$ if we simply assume that the error distributions of these x's and z's are independent. Theoretically, this is not true for the following two reasons. 1) They are from radar and vision observations in which, although the radar range and angle could be independently distributed, the transformed x and y are not. 2) The C point is derived from L and R points; its error could be correlated with them.

Once we have the system and observation equations, the EKF is employed to estimate the contour state vector and its covariance at each frame.

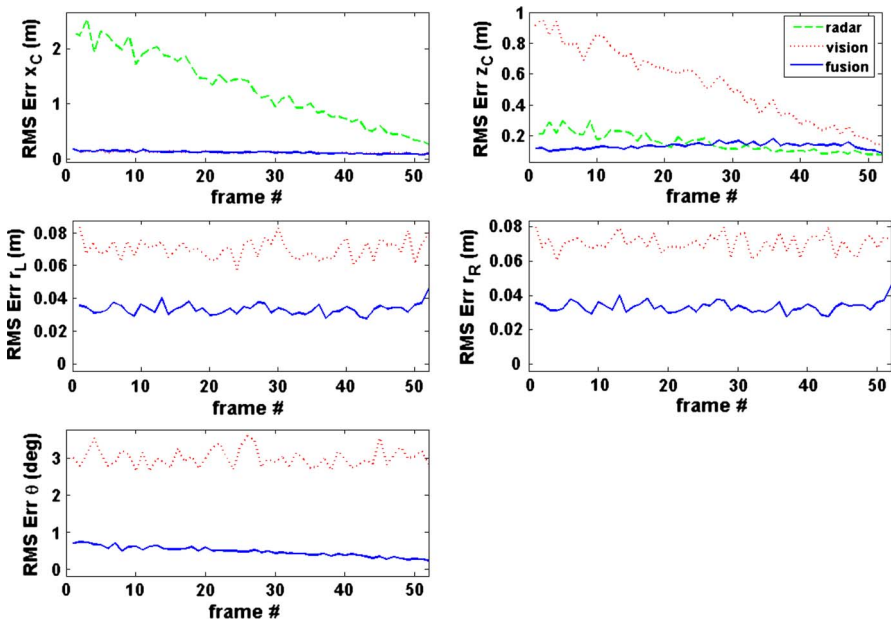

Fig. 7. Comparison of RMS errors from different sensors with 500 MCRs. Note that since the shape and the orientation of the threat vehicle cannot be estimated from the radar sensor, we only plot the radar estimated closest point errors.

IV. Experimental Results

A. Simulations

As shown in Fig. 6, three scenarios are considered in our simulations, where the host vehicle moves toward the threat vehicle at a constant velocity $v_z = 10$ m/s, and the threat vehicle is stationary. These scenarios cover both one-side and two-side views, but there are at different locations.

The following parameters are used to generate synthetic radar and vision data. The radar range and azimuth noise STDs are $\sigma_r = 0.1$ m and $\sigma_\theta = 5^\circ$, respectively, whereas the vision noise STDs in the x- and z-directions are calculated by $\sigma_x = 2(z/f_x) + 0.05|x|$ and $\sigma_z = 0.1z$, respectively. Here, f_x is the focal length in the x-direction. The sampling frequencies for both radar and stereo vision systems are chosen as 30 Hz.

The synthetic observations for the radar range and azimuth are generated by $r_k = \bar{r}_k + \xi_k$ and $\theta_k = \bar{\theta}_k + \zeta_k$, where $\xi \sim N(0, \sigma_r)$ and $\zeta_k \sim N(0, \sigma_\theta)$. The synthetic stereo vision observations are generated as follows.

1) Given the ground truth of the left-, central-, and right-edge points, which are denoted as p_L, p_C, and p_R, respectively.
2) Uniformly sample n points on the two line segments $p_L p_C$ and $p_C p_R$.
3) Add Gaussian noise on each sampling point with local STDs of (0.05, 0.1) m.
4) Add the same Gaussian noise with vision STDs on all points generated by step 3.

To evaluate the simulation results, first, we compare the RMS errors from radar, vision, and fusion by

$$\epsilon_j(k) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[\hat{x}_j(k) - \bar{x}_j(k)]^2}$$

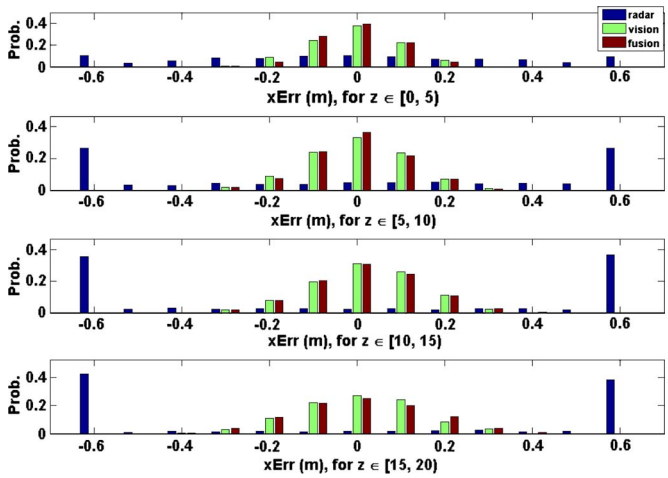

Fig. 8. Error distributions of the x-element of the closest point at range intervals of [0, 5), [5, 10), [10, 15), and [15, 20) m with 500 MCRs.

where $\hat{x}$ and $\bar{x}$ are the estimation and the ground truth of one element of the state vector, respectively, N is the total number of Monte Carlo runs (MCRs), and $j =$ radar, vision, and fusion. Fig. 7 shows the RMS errors of x_c, z_c, r_L, r_R, and θ defined in (11). It indicates that the fused errors are smaller than those from radar or vision.

We also calculate the normalized histograms from RMS errors for the left-edge, right-edge, and closest points in the range intervals [0, 5), [5, 10), [10, 15), and [15, 20) m, respectively. As an example, the results of the closest point in scenario (a) are displayed in Figs. 8 and 9, respectively. These results demonstrate the following facts: 1) There is no significant difference for the x-errors between vision and fused data since the vision azimuth detection errors are already small enough (compared with radar), and the fusion algorithm can no longer improve it. However, compared with the radar x-errors, there is a big improvement from the fusion results.

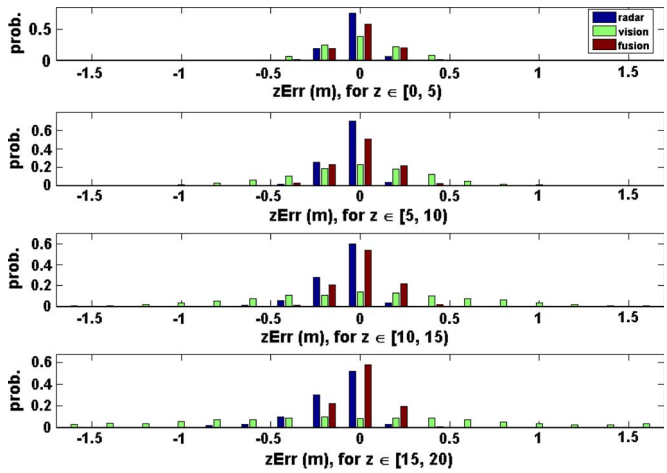

Fig. 9. Error distributions of the z-element of the closest point at range intervals of [0, 5), [5, 10), [10, 15), and [15, 20) m with 500 MCRs.

2) The z-errors in the fused result are much smaller than those from vision alone, particularly when the threat vehicles are far away from the host. This can be explained by the fact that the vision sensor at a larger range gives a larger observation error, and by fusing with the accurate radar observations, the overall range estimation accuracy is significantly improved.

B. Field Experiments

The proposed algorithm was integrated into our stereo-vision-based collision sensing system. It was tested by the in-vehicle stereo vision and radar test bed.

An extensive road test was conducted using two vehicles and driving 1500 mi. Driving conditions included day and night drive times in weather ranging from clear to moderate rain and moderate snowfall. Testing was conducted in heavy-traffic conditions using an aggressive driving style to challenge the crash-sensing algorithms.

During the driving tests, each sensor was configured with an object time-to-collision decision threshold so that objects could be tracked as they approached the test vehicle. The object location time-to-collision threshold was located at 250 ms from contact, as determined by each individual sensor's algorithms, as well as by the sensor fusion algorithm. As an object crossed the time threshold, raw data, algorithm decision results, and ground-truth data were recorded for 5 s prior to the threshold crossing and 5 s after each threshold crossing. This allowed aggressive maneuvers to result in 250-ms threshold crossings to happen from time to time during each test drive. The recorded data and algorithm outputs were then analyzed to determine the system performance in each of the close encounters that happened during the driving tests. During 1500 mi of testing, 307 objects triggered the 250-ms time-to-collision threshold of the radar detection algorithms, and 260 objects triggered the vision systems' 250-ms time-to-collision threshold. Eight objects triggered the fusion-algorithm-based time-to-collision threshold. The posttest data analysis determined that the eight fusion-algorithm-based objects detected were all actually 250 ms or closer to colliding with the test car, whereas the other detection was triggered from noise in the trajectory prediction of objects that were, upon analysis, found to be further away from the test vehicle when the threshold crossing was triggered.

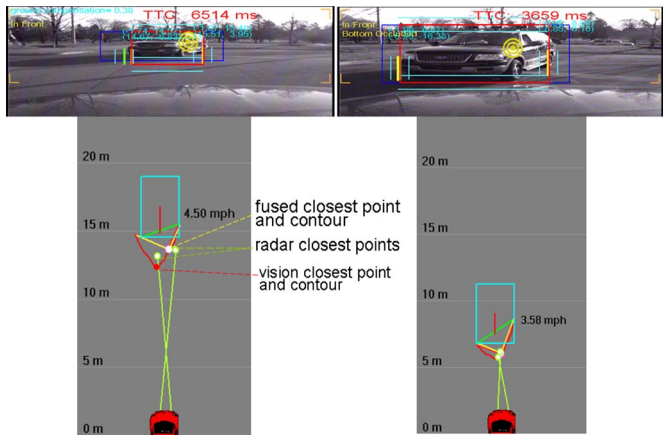

Fig. 10. Snapshots of the two scenes (left and right) and the corresponding synthetic threat vehicle in the bird's-eye-view representation with the vision closest point, the radar closest point, the fused closest point, the vision contour, and the fused contour displayed by different colors.

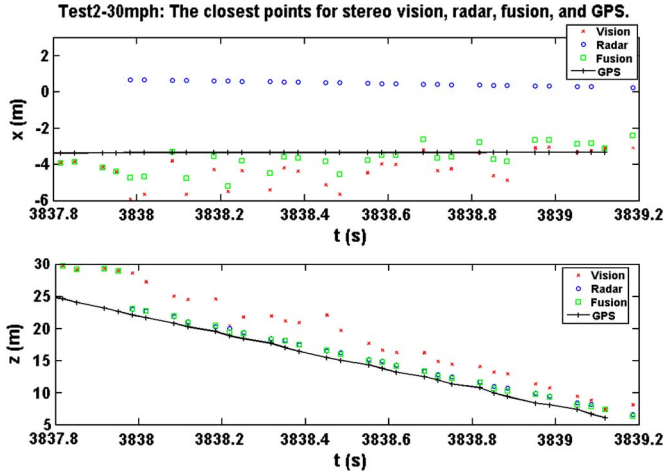

Fig. 11. Comparison of the closest points from stereo vision, radar, fusion, and high-resolution GPS.

Fig. 10 shows two snapshots of the video and a bird's eye view of the threat car with respect to the host vehicle. Fig. 11 compares the closest points from vision, radar, and fusion with GPS. The scenario of this plot is that the threat vehicle was parked in the left front of the host car when the host car was driving straightly forward at a speed of about 30 mi/h.

V. Conclusion and Discussion

In summary, we have proposed a novel algorithm for accurately estimating the location, size, pose, and motion information of a threat vehicle in the host vehicle's coordinate system by fusing the information from both stereo-camera and millimeter-wave radar sensors. Simulation and field experiment results have demonstrated that the proposed fusion algorithm can take the advantages of both sensors to improve the collision-sensing accuracy.

Finally, as pointed out by anonymous reviewers, there are several issues worth discussing in this prototype system. One is about line fitting: Since the vision range profile should have different error magnitudes of x and z, and the range errors are larger than those of x, it would be better to use weighted fitting in this case to utilize the *a priori* knowledge of the error magnitude. Another one is about the three-point contour modeling: In a scenario where a threat vehicle is moving from an almost perpendicular manner to a totally perpendicular manner to the host one, and the center point "C" may jump from one side to the middle of the vehicle. This discontinuity could generate large estimation errors in the tracking algorithm. Additionally, the line fitting in the bird's-eye-viewing plane may oversimplify the problem; instead, the plane fitting in the 3-D space may be more reliable.

ACKNOWLEDGMENT

The authors would like to thank the detailed valuable comments of anonymous reviewers.

REFERENCES

[1] E. Dickmanns, "The development of machine vision for road vehicles in the last decade," in *Proc. IEEE Intell. Vehicles Symp.*, 2002, pp. 268–281.
[2] S. Tokoro, "Automotive application systems of a millimeter-wave radar," in *Proc. IEEE Intell. Vehicles Symp.*, 1996, pp. 260–265.
[3] S. Oshima, Y. Asano, T. Harada, N. Yamada, M. Usui, H. Hayashi, T. Watanabe, and H. Iizuka, "Phase-comparison monopulse radar with switched transmit beams for automotive application," in *IEEE MTT-S Conf. Dig.*, 1999, pp. 1493–1496.
[4] S. S. Tsugawa, "Vision-based vehicles in Japan: Machine vision system and driving control systems," *IEEE Trans. Ind. Electron.*, vol. 41, no. 4, pp. 398–405, Aug. 1994.
[5] K. Saneyoshi, "Drive assist system using stereo image recognition," in *Proc. IEEE Intell. Vehicles Symp.*, 1996, pp. 230–235.
[6] T. Williamson and C. Thorpe, "Detection of small obstacles at long range using multibaseline stereo," in *Proc. IEEE Intell. Vehicles Symp.*, 1998, pp. 311–316.
[7] K. C. J. Dietmayer, J. Sparbert, and D. Streller, "Model based object classification and object tracking in traffic scenes from range images," in *Proc. IEEE Intell. Vehicles Symp.*, 2001, pp. 25–30.
[8] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, Jan. 1997.
[9] R. Grover, G. Brooker, and H. F. Durrant-Whyte, "A low level fusion of millimeter wave radar and night-vision imaging for enhanced characterization of a cluttered environment," in *Proc. Aust. Conf. Robot. Autom.*, 2001, pp. 73–80.
[10] B. Steux, C. Laurgeau, L. Salesse, and D. Wautier, "Fade: A vehicle detection and tracking system featuring monocular color vision and radar data fusion," in *Proc. IEEE Intell. Vehicles Symp.*, 2002, pp. 632–639.
[11] A. Gern, U. Franke, and P. Levi, "Advanced lane recognition-fusing vision and radar," in *Proc. IEEE Intell. Vehicles Symp.*, 2000, pp. 45–51.
[12] A. Gern, U. Franke, and P. Levi, "Robust vehicle tracking fusing radar and vision," in *Proc. Int. Conf. Multisensor Fusion Integr. Intell. Syst.*, 2001, pp. 323–328.
[13] A. Sole, O. Mano, G. Stein, H. Kumon, Y. Tamatsu, and A. Shashua, "Solid or not solid: Vision for radar target validation," in *Proc. IEEE Intell. Vehicles Symp.*, 2004, pp. 819–824.
[14] Y. Tan, F. Han, and F. Ibrahim, "A radar guided vision system for vehicle validation and vehicle motion characterization," in *Proc. IEEE Intell. Vehicles Symp.*, 2007, pp. 1059–1066.
[15] 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.
[16] N. Srinivasa, Y. Chen, and C. Daniell, "A fusion system for real-time forward collision warning in automobiles," in *Proc. IEEE Intell. Vehicles Symp.*, 2003, pp. 457–462.
[17] 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.
[18] N. Kawasaki and U. Kiencke, "Standard platform for sensor fusion on advanced driver assistance system using Bayesian network," in *Proc. IEEE Intell. Vehicles Symp.*, 2004, pp. 250–255.
[19] P. Chang, T. Camus, and R. Mandelbaum, "Stereo-based vision systems for automotive imminent collision detection," in *Proc. IEEE Intell. Vehicles Symp.*, 2004, pp. 274–279.
[20] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking System*. Boston, MA: Artech House, 1999.
[21] E. W. Weisstein, "Least squares fitting–perpendicular offsets," MathWorld—A Wolfram Web Resource. [Online]. Available: http://mathworld.wolfram.com/LeastSquaresFittingPerpendicularOffsets.html

Shunguang Wu (M'03) received the B.Sc. degree in physics from Hanzhong Teacher's College, Hanzhong, China, in 1987, the M.S. and Ph.D. degrees in theoretical physics from Northwestern University (NWU), Xi'an, China, in 1992 and 1998, respectively, and the M.S. degree in computer engineering and the Ph.D. degree in electrical engineering from Wright State University (WSU), Dayton, OH, in 2005.

From 1992 to 1998, he was an Instructor and an Associate Professor with the Department of Physics, NWU. From 1998 to 2000, he was a Research Associate with the Institute of Low Energy Nuclear Physics, Beijing Normal University, Beijing, China. From 2002 to 2005, he was a Research Associate with the Department of Electrical Engineering, WSU. He is currently a Member of the Technical Staff with Vision Technologies, Sarnoff Corporation, Princeton, NJ. His research interests include Bayesian estimation, multisensor data fusion, system modeling, computer vision and image processing, nonlinear dynamics, and nuclear physics.

Dr. Wu is a member of the International Society for Optical Engineers.

Stephen Decker (M'08) received the B.S. degree in electronics from California Polytechnic University, Pomona, in 1983. He continued to do graduate-level course work in sensor fusion, algorithm development, and artificial intelligence.

He developed sensor fusion algorithms, along with target detection, classification, and tracking algorithms for application to missile seekers and smart munitions before joining an automotive safety supplier. While working in the automotive sector, he developed precrash-sensing systems and smart restraint control sensors, including a neural-network-based occupant position and classification sensing system. He was with the Advanced Sensor Development, Autoliv Inc., Southfield, MI.

Peng Chang (M'03) received the B.S. degree from Tsinghua University, Beijing, China, in 1994 and the Ph.D. degree from Carnegie Mellon University, Pittsburgh, PA, in 2002.

From 2002 to 2007, he was a Member of the Technical Staff with Vision Technologies, Sarnoff Corporation, Princeton, NJ. He is the Founder of General Vision LLC, Princeton. His research interests include real-time object recognition with mono or stereo vision, sensor fusion, and statistical decision theory. He has explored applications in the area of autonomous robot navigation and automotive safety.

Theodore Camus (SM'00) received the B.S. degree from Rensselaer Polytechnic Institute, Troy, NY, in 1988 and the M.S. and Ph.D. degrees in computer science from Brown University, Providence, RI, in 1991 and 1995, respectively.

He was a National Research Council Postdoctoral Research Associate with the Perception Systems Group, National Institute of Standards and Technology, until 1996, working on real-time optical flow for robotic collision avoidance. He was a Principal Member of the Technical Staff with Sensar until 2000, working on real-time automatic iris recognition. He is currently a Senior Member of the Technical Staff with Vision Technologies, Sarnoff Corporation, Princeton, NJ, working on real-time vision systems for automotive and robotic applications. He is the author of more than 20 publications. He is the holder of nine patents in the areas of real-time optical flow, stereo processing, iris finding, automotive vision, and pedestrian detection.

Jayan Eledath (M'06) received the B.E. degree in electrical engineering from the University of Bombay, Mumbai, India, in 1994 and the M.S. degree in electrical and computer engineering from the University of Texas, Austin, in 1997.

Since 1997, he has been with Vision Technologies, Sarnoff Corporation, Princeton, NJ, where he has worked on aerial video surveillance, automotive safety and driver-awareness systems, robotic perception, medical image analysis, and embedded vision processing. He is currently the Head of the Applied Vision Group. More recently, he has been the Principal Investigator or the Technical Manager of multiple government and commercial programs in long-term change detection, multisensory fusion for vehicle safety, human/pedestrian detection, and multirobot 3-D mapping and structure characterization. He is the author or a coauthor of more than ten technical publications. He is the holder of three patents, with several others pending.

Mr. Eledath has received four Sarnoff Achievement Awards and a Best Paper (Novel Engineering Application) award at a leading neural network conference.