
A target tracking system using sensors of multiple modalities

Shuqing Zeng*

Electrical and Controls Integration Laboratory,
GM Research and Development Centre,
30500 Mound Road,
Warren, Michigan 48090, USA
E-mail: shuqing.zeng@gm.com

*Corresponding author

Abstract: Radar sensors and camera based vision systems will be used to provide object data in many Advanced Driver Assistance Systems (ADAS). In this paper, a multi-sensor target tracking system that combines data from a Frequency-Modulated Continuous Wave (FMCW) radar, multiple Short-Range Radars (SRR), a camera-based object detection system and vehicular dynamic sensors is described. Each object sensor individually measures the range, range rate and azimuth angle information of all objects within the observation region. The proposed system 1 groups objects from different sensors in overlapped observation region; 2 tracks an object across different sensor field of views; 3 reports the Cartesian coordinates of objects with improved accuracy and reduced rates of false detection and missed detection. The proposed target tracking system was implemented in a retrofitted vehicle. Only about two-percent CPU usage is needed for an 800 MHz embedded processor. The output data was directly used by several vehicle features such as low-speed collision avoidance and Adaptive Cruise Control (ACC) systems.

Keywords: vehicle safety; advanced driver assistance systems; multi-target tracking; sensors of multiple modalities.

Reference to this paper should be made as follows: Zeng, S. (2013) 'A target tracking system using sensors of multiple modalities', *Int. J. Vehicle Autonomous Systems*, Vol. 11, No. 4, pp.384–404.

Biographical notes: Shuqing Zeng received his PhD degree in Computer Science from the Michigan State University, East Lansing, Michigan, in 2004. Since 2004, he has been with the Research and Development Center, General Motors Corporate, Warren, MI, where he currently holds the position of Senior Research Scientist. From 2005–2007, he served as the newsletter Editor of Autonomous Mental Development TC, IEEE Computational Intelligence Society. He is currently an Associate

Editor of the *International Journal of Humanoid Robotics*. His research interests include computer vision, sensor fusion, autonomous driving and active-safety applications on vehicle. He served as a judge to the Intelligent Ground Vehicle Competition. He is a member of the Tartan Racing team who won the first place of the Defense Advanced Research Projects Agency Urban Challenge on November 3, 2007.

1 Introduction

Combining results of multiple sensors can provide more reliable and accurate information than using a single sensor (Hall, 1992; Smith and Singh, 2006; Luo et al., 2002; Gunes and Piccardi, 2005; Kaempchen et al., 2005); this allows either improved performance from existing sensors or the same performance from smaller and less expensive sensors. Consequently, there has been an increased interest in fusing data from multiple sensors to determine the location, moving direction, relative velocity and acceleration of critical objects around the vehicle for Advanced Driver Assistance Systems (ADAS). For example, (Mobus and Kolbe, 2004) shows that the Adaptive Cruise Control (ACC) is enhanced by fusing a 77 GHz radar and an infrared lidar. Good system performance is obtained in Becker (2000) and Redmill et al. (2006) for an autonomous vehicle application by using radar, stereo vision and lidar. A pre-crash system based on a lidar and two short range 24 GHz radars is presented in Skutek et al. (2005). A collision mitigation system using lidar and stereo vision is recommended in Labayrade et al. (2005).

The Field-Of-View (FOV) of forward-looking ACC Long-Range Radar (LRR) is limited to 15°. For ADAS systems beyond ACC, LRR alone might not be sufficient (Mobus and Kolbe, 2004). To improve ACC two extra sensor types were added: A forward target tracking system with a camera with object detection capability and multiple low cost 24 GHz Short Range Radars (SRR). The 45° FOV of vision system and 60° FOV of SRR cover a large area left uncovered by the LRR radar.

The measurements of these low-cost sensors are noisy, have errors and are only a partial view of the world. In addition, we have to evaluate the reliability of sensor data. Reliability represents how much confidence we have in the data reported by the sensor. All of these aspects contribute to an increase in the uncertainty in the multi-sensor tracking system. To address the aforementioned issues, the work reported here contributes in the following aspects:

- The architecture in this work is designed in a way so that new generations of sensors can be integrated. In this architecture, we can have the system integration in different configurations. We can use the tracking strategy in one sensor (time series), redundant sensors, or sensors of multiple modalities.
- A weighted-least-squares based method is proposed to fuse data from different sources. Different sensor configurations are evaluated to find the optimal setting.

- In-vehicle tests were conducted to demonstrate the proposed system can track the same object across multiple sensors.

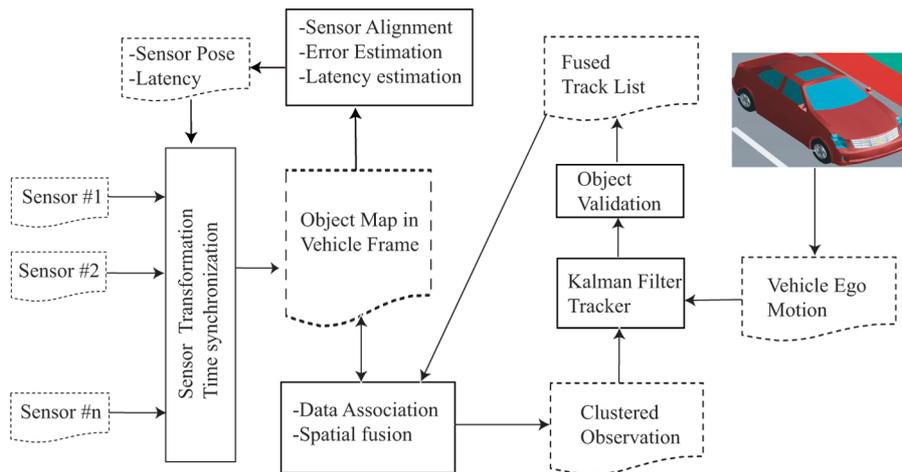
The remainder of this paper is organised as follows: Section 2 is focused on the architecture and algorithm, especially the filters for object tracking, data association and automatic sensor registration. Several simulation results are discussed in Section 3. Section 4 describes the results of in-vehicle tests.

2 Algorithm

2.1 Architecture aspect

Figure 1 outlines the block flow of the proposed system; the details of those blocks will be presented later. The multi-step process assumes that observations are processed sequentially and begins with the acquisition of the observations from the individual sensors. There are two independent threads.

Figure 1 Flow diagram of the proposed system (see online version for colours)



First, *Sensor Transformation* transforms the object maps from individual sensors into a unified object map in the vehicle frame based on the estimated pose and the latency of each sensor. *Data Association* compares the unified object map against known entities represented by *Fused Track List*. The observations may represent the observed position of an entity (e.g., range, azimuth and range rate), identity information and parameters that can be related to identify (e.g., confidence level, tracking maturity, and geometric information of the entity). As illustrated in Figure 1, the *Data Association* block systematically compares observation against the known fused tracks and determines whether or not these observation-track are related. The block *Spatial fusion* groups the observations that are associated with the same fused track and outputs *Clustered Observation*. The *Kalman Filter Tracker* updates the fused tracks by using *Clustered Observation* and the vehicle's ego motion.

In the second thread, starting from the *Data Association* block, we retrieve the candidate pairs from the observation-track pairs from a particular sensor and then select the pairs with good matching scores to estimate the position and pose of the sensor. The *Latency Estimation* block uses the synchronising clock (see Figure 1) as the time reference to find out the latency in each measurement cycle.

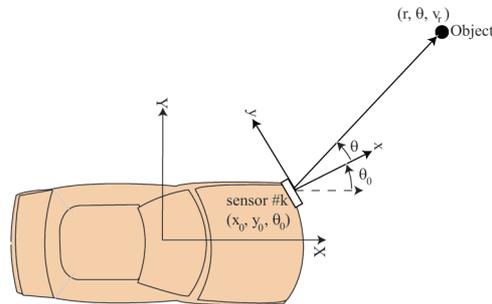
2.2 Error model of the sensor

We consider a sensing system with multiple sensors on the ego vehicle. Let's first define the system of reference as shown in Figure 2. The sensor k is mounted at the pose $\mathbf{m} = (x_0, y_0, \theta_0)$ with respect to the vehicle frame, where θ_0 denotes the orientation of the sensor's bore-sight. A measurement of an object is a three dimensional vector $\mathbf{o} = (r, \theta, v_r)$, where r and θ are the range and azimuth angle measurements in the sensor frame, respectively; v_r denotes the range-rate along the azimuth axis. With random error in measurement, the observation in the vehicle frame determined from \mathbf{o} becomes a probability distribution whose extent can be characterised by the sensor's error variances. The error variances $(\sigma_r^2, \sigma_\theta^2, \sigma_{v_r}^2)$ found in the sensor's specification determine the accuracy of the sensor measurement. Therefore, the covariance matrix Γ_o of a measurement in the sensor frame can be represented by

$$\Gamma_o = \begin{pmatrix} \sigma_r^2 & 0 & 0 & 0 \\ 0 & \sigma_\theta^2 & 0 & 0 \\ 0 & 0 & \sigma_{v_r}^2 & 0 \\ 0 & 0 & 0 & \sigma_{v_t}^2 \end{pmatrix} \quad (1)$$

where, σ_{v_t} is an extremely large number or infinity, corresponding to the unobservable tangent velocity v_t . By using a covariance matrix also including the component of tangent velocity v_t , we can treat the sensors with complementary performance characteristics and different orientation in a unified way. Although our analysis is applicable to both polar or cartesian representation of sensor measurement, we only focus on the polar representation in this section.

Figure 2 Reference system (see online version for colours)



We investigate the error in determining the 2D positions (in the vehicle frame) of a set of points visible by a sensor. These points denoted by \mathbf{y} correspond to a set

of sensor measured points \mathbf{o}^1 . Assume the estimated pose of the sensor to be \mathbf{m} . The estimated 2D position of points \mathbf{y} in the above reference is then a function $\mathbf{y} = g(\mathbf{o}, \mathbf{m})$. From Figure 2, it is not difficult to derive the equations to calculate the components of \mathbf{y} as

$$\begin{aligned} x &= x_0 + r \cos(\theta + \theta_0) \\ y &= y_0 + r \sin(\theta + \theta_0) \\ v_x &= v_r \cos(\theta + \theta_0) - v_t \sin(\theta + \theta_0) \\ v_y &= v_r \sin(\theta + \theta_0) + v_t \cos(\theta + \theta_0) \end{aligned} \tag{2}$$

and conversely $\mathbf{o} = g^{-1}(\mathbf{y}, \mathbf{m})$, that is

$$\begin{aligned} r &= \sqrt{(x - x_0)^2 + (y - y_0)^2} \\ \theta &= \tan^{-1} \left(\frac{y - y_0}{x - x_0} \right) - \theta_0 \\ v_r &= v_x \cos(\theta + \theta_0) + v_y \sin(\theta + \theta_0) \\ v_t &= -v_x \sin(\theta + \theta_0) + v_y \cos(\theta + \theta_0). \end{aligned} \tag{3}$$

To set up notation for later optimisation, we write the equations (2) and (3) in a generic form:

$$\mathbf{g}(\mathbf{o}, \mathbf{y}) = 0 \tag{4}$$

given sensor pose \mathbf{m} is known. We write small perturbations of \mathbf{o} by $\delta\mathbf{o}$ and of \mathbf{y} by $\delta\mathbf{y}$. Then the errors are related each other as expressed by the following equation:

$$\frac{\partial \mathbf{g}}{\partial \mathbf{o}} \delta\mathbf{o} + \frac{\partial \mathbf{g}}{\partial \mathbf{y}} \delta\mathbf{y} = 0 \tag{5}$$

where

$$\frac{\partial \mathbf{g}}{\partial \mathbf{o}} = \begin{pmatrix} b & -ra & 0 & 0 \\ a & rb & 0 & 0 \\ 0 & -v_r a - v_t b & b & -a \\ 0 & v_r b - v_t a & a & b \end{pmatrix} \tag{6}$$

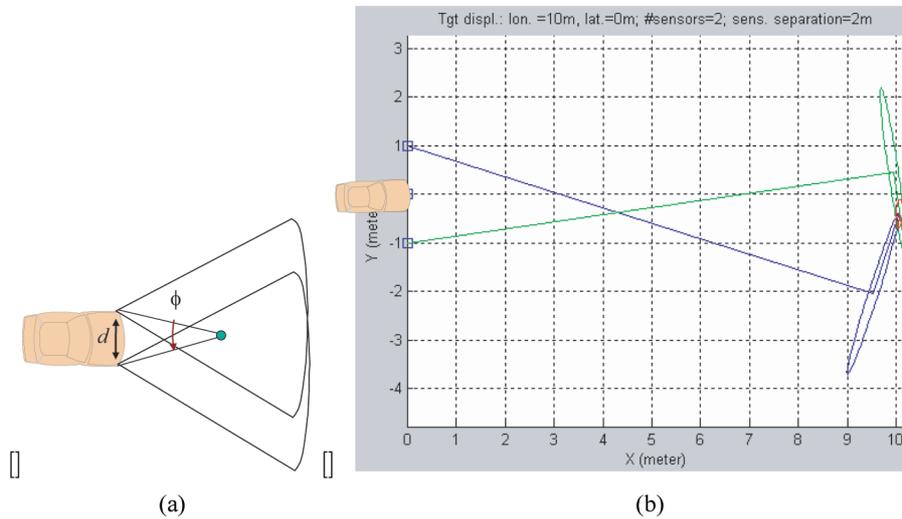
$$\frac{\partial \mathbf{g}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{x-x_0}{r} & \frac{y-y_0}{r} & 0 & 0 \\ -\frac{y-y_0}{r^2} & \frac{x-x_0}{r^2} & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & a \end{pmatrix} \tag{7}$$

$$a = \sin(\theta + \theta_0)$$

$$b = \cos(\theta + \theta_0)$$

Finally, as an illustrative example, Figure 3(b) shows the covariance matrices of the observations reported by two sensors with poor azimuth accuracies (e.g., $\pm 10^\circ$). The true object is located at 10 meters and 0° azimuth, and the sensors are mounted 2 meters apart from each other. The ellipses denote the uncertain measure of the observations.

Figure 3 (a) A simulated vehicle with two sensors and (b) The covariance matrices of the two observations given by two sensors with poor azimuth accuracy (e.g., $\pm 10^\circ$) (see online version for colours)



2.3 Data association

Data association answers the question: given N observations, \mathbf{y}_i from one or more sensors, how do we determine which observations belong together, representing observations of the same entity? That is, we seek to determine whether observation \mathbf{y}_{A_i} (the i th observation from sensor A) results from the same entity as observation \mathbf{y}_{B_j} (the j th observation from sensor B).

Six steps are needed to determine the association between observation and track:

1. retrieve candidate entities from the track database
2. update candidate entities to the observation time t
3. compute predicted observations for each candidate track
4. compute association matrix
5. perform gating to eliminate unlikely observation-track pairs
6. implement assignment logic.

Given an observation \mathbf{o}_i , the first step involves querying candidate entities from the track database containing previously determined state vector (\mathbf{x}_j) of the tracks, which denote the current estimate of the entities' position, velocity and classification. Such queries retrieve all entities from the database within specific geometric boundaries (e.g., the sensor's field-of-view), or having specified attributes (e.g., stationary vs. moving entities).

For dynamic situations the second step required for association is to update the state vectors of the tracks to the observation time t . Thus, for each previous determined track, $\mathbf{x}_j(t-1)$, the predicted value of $\mathbf{x}_j(t)$ is computed by

$$\mathbf{x}_j(t) = \Phi(t, t-1)\mathbf{x}_j(t-1) + \mathbf{k} \quad (8)$$

where $\Phi(t, t-1)$ represents a transformation that updates the state vector from time $t-1$ to time t and the vector \mathbf{k} denotes unknown noise modelled as: $\mathbf{k} \sim N(0, \mathbf{Q})$.

The third step is required in association unless our sensors are able to observe directly an entity's state. This step utilises an observation equation to predict observation $\tilde{\mathbf{o}}_j(t)$, which results from entity \mathbf{x}_j . That is, $\tilde{\mathbf{o}}_j(t) = g^{-1}(\mathbf{x}_j) + \mathbf{n}_j$ where $\mathbf{n}_j \sim N(0, \mathbf{R}_j)$ denotes a white Gaussian random process.

The fourth association step is to calculate an association matrix. The (i, j) th component of the matrix is the similarity measure that compares the closeness of observation $\mathbf{o}_i(t)$ and the predicted observation $\tilde{\mathbf{o}}_j(t)$. Many measures of similarity are proposed in literature. In this paper the Mahalanobis distance (Hall, 1992) is used, that is

$$\mathbf{d}(\mathbf{o}_i, \tilde{\mathbf{o}}_j) = (\mathbf{o}_i - \tilde{\mathbf{o}}_j)^T (\mathbf{P}_i + \mathbf{P}_j)^{-1} (\mathbf{o}_i - \tilde{\mathbf{o}}_j) \quad (9)$$

where \mathbf{P}_i and \mathbf{P}_j denote the covariance matrices of the given observation \mathbf{o}_i and the predicted quantity $\tilde{\mathbf{o}}_j$, respectively.

The fifth step is termed gating where the number of possible combinations of observation-track pairs is reduced. In other words, the gating performs a screen to eliminate the unlikely pairs, via heuristic knowledge or statistical hypothesis testing. A straightforward approach is to use the association matrix. All pairs not satisfying $\mathbf{d}(\mathbf{o}_i, \tilde{\mathbf{o}}_j) < T$ will be removed, where T denotes a threshold.

The final step in the association process is the actual assignment of observation to tracks. This assignment step is the invocation of decision logic to declare the association that relates an observation to a candidate track. A hard decision approach is employed in our proposed system; that is an observation is assumed to belong to a single target. In this paper, the assignment logic assigns the observation to the nearest adjacent track (namely, the nearest neighbour approach), i.e., $j = \arg \min_j \mathbf{d}(\mathbf{o}_i, \tilde{\mathbf{o}}_j)$.

2.4 Optimisation

Having established the association that relates a state vector to predicted observations, a key issue will be addressed in this subsection: how to determine a value of a state vector $\mathbf{x}(t)$ that best fits the observed data.

To illustrate the formulation and processing flow for the optimisation process, we consider the weighted-least-squares method to group related observations into a clustered observation \mathbf{y} in the vehicle frame.

One or more sensors observe an object, reporting multiple observations related to the target position \mathbf{x} . The (unknown) fused observation in the vehicle frame is represented by a vector \mathbf{y} , determined by a time invariant observation

equation $\mathbf{g}(\mathbf{o}, \mathbf{y}) = 0$. With actual observation \mathbf{o}^* and estimate \mathbf{y}^* , the first order approximation of $\mathbf{g}(\mathbf{o}, \mathbf{y})$ can be written as

$$\mathbf{g}(\mathbf{y}^*, \mathbf{o}^*) + \frac{\partial \mathbf{g}}{\partial \mathbf{y}} \Big|_{(\mathbf{y}^*, \mathbf{o}^*)} (\mathbf{y} - \mathbf{y}^*) + \frac{\partial \mathbf{g}}{\partial \mathbf{o}} \Big|_{(\mathbf{y}^*, \mathbf{o}^*)} (\mathbf{o} - \mathbf{o}^*) \approx 0 \quad (10)$$

Writing $\mathbf{A} = \frac{\partial \mathbf{g}}{\partial \mathbf{y}} \Big|_{(\mathbf{y}^*, \mathbf{o}^*)}$, $\mathbf{B} = \frac{\partial \mathbf{g}}{\partial \mathbf{o}} \Big|_{(\mathbf{y}^*, \mathbf{o}^*)}$, $\mathbf{l} = -\mathbf{g}(\mathbf{y}^*, \mathbf{o}^*)$, and $\varepsilon = -\mathbf{B}(\mathbf{o} - \mathbf{o}^*)$, equation (10) becomes a linearised form as

$$\mathbf{A}(\mathbf{y} - \mathbf{y}^*) = \mathbf{l} + \varepsilon \quad (11)$$

The residue $\mathbf{o} - \mathbf{o}^*$ gives the difference between the noise-free observation \mathbf{o} and actual observation \mathbf{o}^* . Hence quantity $\mathbf{o} - \mathbf{o}^*$ can be treated as observation noise. Letting Γ_o denote the observation noise, the covariance matrix (Γ_ε) of the residue ε in equation (11) becomes

$$\Gamma_\varepsilon = \mathbf{B}\Gamma_o\mathbf{B}^T. \quad (12)$$

Now, we assume a total of K independent observations, $\{\mathbf{o}_k | k = 1, \dots, K\}$, related to the fused quantity \mathbf{y} . Hence equation 11 can be extended to be

$$\begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \dots \\ \mathbf{A}_K \end{pmatrix} (\mathbf{y} - \mathbf{y}^*) = \begin{pmatrix} \mathbf{l}_1 \\ \mathbf{l}_2 \\ \dots \\ \mathbf{l}_K \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_K \end{pmatrix}. \quad (13)$$

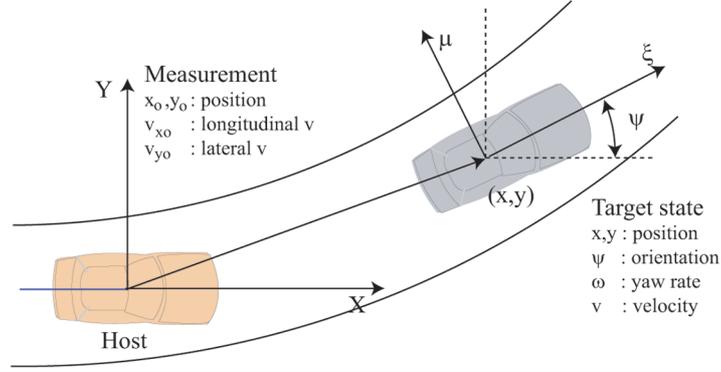
By the Gauss-Markov Theorem (Weng et al., 1993), obtaining the linear minimum variance estimate of $\mathbf{y} - \mathbf{y}^*$ in equation (13) yields

$$\hat{\mathbf{y}} = \mathbf{y}^* + \left(\sum_{k=1}^K \mathbf{A}_k^T \Gamma_{\varepsilon k}^{-1} \mathbf{A}_k \right)^{-1} \sum_{k=1}^K \mathbf{A}_k^T \Gamma_{\varepsilon k}^{-1} \mathbf{l}_k \quad (14)$$

2.5 Dynamic models

We assume the target executes a manoeuvre under constant speed along a circular path. This type of motion is common in ground vehicle traffic.

Consider a two-vehicle scenario, as shown in Figure 4. The host vehicle follows a target vehicle in a lane with constant curvature κ . As described previously, the measurement \mathbf{y} (in vehicle frame) includes x_o , y_o , v_{xo} , and v_{yo} . The target vehicle dynamic state is represented by $\mathbf{x} = (x, y, \psi, \omega, v)$, where the quantities x , y , and ψ denote the pose of the target, and ω and v denote the target's kinematic state.

Figure 4 Coordinated turn model (see online version for colours)

The kinematic state of the host, modelled as a bicycle model, is represented by yaw rate (ω_H), longitudinal speed (v_{xH}) and lateral speed (v_{yH}). Let ΔT denotes the sampling interval from previous cycle \mathbf{x} to new cycle \mathbf{x}' . Hence, we can write the dynamic equation $\mathbf{x}' = \mathbf{f}(\mathbf{x})$ of the target state by

$$\begin{aligned}
 x' &= x + (v \cos \psi + y\omega_H - v_{xH})\Delta T + k_1 \\
 y' &= y + (v \sin \psi - x\omega_H - v_{yH})\Delta T + k_2 \\
 \psi' &= \psi + (\omega - \omega_H)\Delta T + k_3 \\
 \omega' &= \omega + k_4 \\
 v' &= v + k_5
 \end{aligned} \tag{15}$$

and the observation function $\mathbf{y} = \mathbf{h}(\mathbf{x})$ by

$$\begin{aligned}
 x_o &= x \\
 y_o &= y \\
 v_{x_o} &= v \cos \psi + y\omega_H - v_{xH}
 \end{aligned} \tag{16}$$

Additionally, the Jacobians of equations (15) and (16) are

$$\mathbf{F} = \begin{pmatrix} 1 & \omega_H \Delta T & -v \Delta T \sin \phi & 0 & \Delta T \cos \phi \\ -\omega_H \Delta T & 1 & v \Delta T \cos \phi & 0 & \Delta T \sin \phi \\ 0 & 0 & 1 & \Delta T & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{17}$$

and

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & \omega_H & -v \sin \phi & 0 & \cos \phi \end{pmatrix} \tag{18}$$

respectively.

After establishing the observation equations that relate a state vector to predicted observations, and also the motion equations for the dynamic system, we can then write the tracking algorithm, a version of Extended Kalman Filter (EKF), as

Measurement update:

$$\mathbf{K}(t) = \mathbf{P}(t, t-1) \mathbf{H}^T (\mathbf{H} \mathbf{P}(t, t-1) \mathbf{H}^T + \mathbf{C})^{-1} \quad (19)$$

$$\mathbf{x}(t) = \mathbf{x}(t, t-1) + \mathbf{K}(t) (\mathbf{y}(t) - \mathbf{h}(\mathbf{x}(t, t-1))) \quad (20)$$

$$\mathbf{P}(t) = \mathbf{P}(t, t-1) - \mathbf{K}(t) \mathbf{F} \mathbf{P}(t, t-1). \quad (21)$$

Time update:

$$\mathbf{x}(t, t-1) = \mathbf{f}(\mathbf{x}(t-1)) \quad (22)$$

$$\mathbf{P}(t, t-1) = \mathbf{F} \mathbf{P}(t-1) \mathbf{F}^T + \mathbf{Q} \quad (23)$$

2.6 Algorithm summary

Inputs to the algorithm include: 1) an initial estimate of the state vectors $\{\mathbf{x}_j(0)\}$ and initial state uncertainties $\{\mathbf{P}_j(0)\}$ at an epoch $t = 0$ and 2) I_t observations, $\{\mathbf{o}_i(t) | i = 1, 2, \dots, I_t\}$, at time t with associated uncertainties $\Gamma_{oi}(t)$.

At each time t , the system performs a series of calculations:

- 1 Retrieve the observations, $\{\mathbf{o}_i(t) | i = 1, 2, \dots, I_t\}$, measured at time t , and its associated observational uncertainty, $\Gamma_{oi}(t)$.
- 2 For each previous determined track in $\{\mathbf{x}_j(t-1) | j = 1, 2, \dots, J_t\}$, calculate the predicted quantities $\tilde{\mathbf{x}}_j(t)$ by equation (15) and $\tilde{\mathbf{o}}_j(t)$ by equation (3).
- 3 Utilising the data association algorithm shown in Section 2.3, find the associated observation-track pairs $\mathcal{P} = \{(\mathbf{o}_i, \mathbf{x}_j)\}$.
- 4 For each unmatched observation in set $\{\mathbf{o}_{i_k} | \nexists \mathbf{x} : (\mathbf{o}_{i_k}, \mathbf{x}) \in \mathcal{P}\}$, create a new track \mathbf{x}_{j_k} .
- 5 Remove the unmatched tracks in set $\{\mathbf{x}_{i_k} | \nexists \mathbf{o} : (\mathbf{o}, \mathbf{x}_{i_k}) \in \mathcal{P}\}$ from the track database.
- 6 For each track \mathbf{x}_j , retrieve the associated observations $\{\mathbf{o}_{i_k} | (\mathbf{o}_{i_k}, \mathbf{x}_j) \in \mathcal{P}\}$ from the observation-track pairs, and perform the following steps
 - Letting \mathbf{y}_j^* be the predicted state vector $\tilde{\mathbf{x}}_j(t)$ and \mathbf{o}_j^* be predicted sensor measurement $\tilde{\mathbf{o}}_j(t)$, linearise the observation equation 10 at the point $(\mathbf{o}_j^*, \mathbf{y}_j^*)$.
 - Compute the fused observation $\hat{\mathbf{y}}_j$ in the vehicle frame by equation (14)
 - Apply equations (19)–(23) to sequentially estimate the state vector \mathbf{x}_j .

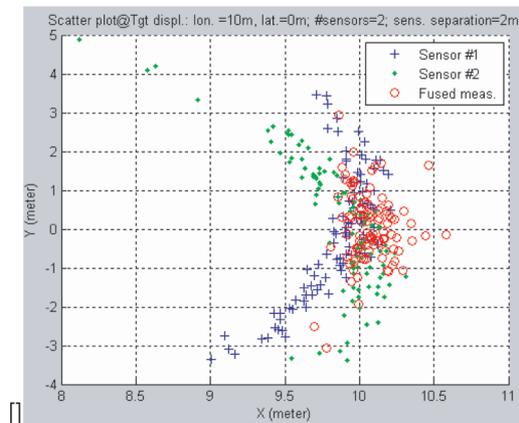
3 Simulation results

To further verify the analysis presented above and validate the proposed algorithms, simulations have been conducted under different conditions. As shown

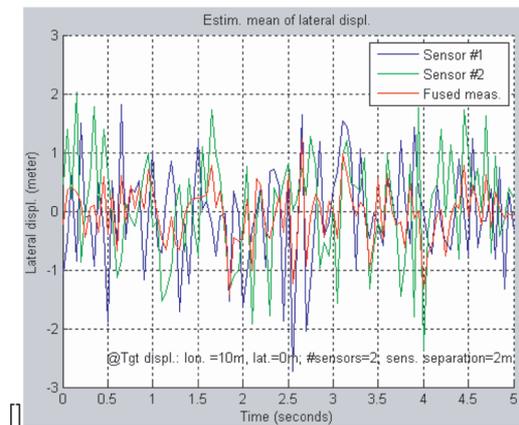
in Figure 3(a), two sensors with poor azimuth accuracy (e.g., $\pm 10^\circ$) are placed in the front bumper 2 m apart.

In the first simulated scenario, both the host and target vehicles are stationary. The target is located at 10 m and 0° azimuth. Hence, the initial state vector $\mathbf{x}(0) \sim N(\mu_0, P_0)$ with $\mu_0 = (10, 0, 0, 0, 0)$ and $P_0 = \text{diag}[100, 100, (1\pi/180)^2, (5\pi/180)^2, 25]$. Assume the motion noise parameters are $\sigma_v^2 = 1$ and $\sigma_\omega^2 = (1\pi/180)^2$ (cf. equation (15)). The result is shown in Figure 5. Observe that the error in lateral displacement measurement is reduced by about 50%.

Figure 5 (a) Shows scatter plot of raw sensor measurements and fused tracks. The crosses and dots denote the measurements of sensor 1 and sensor 2, respectively. The circles represent the fused tracks and (b) shows the measure of the target’s lateral displacement (see online version for colours)



(a)

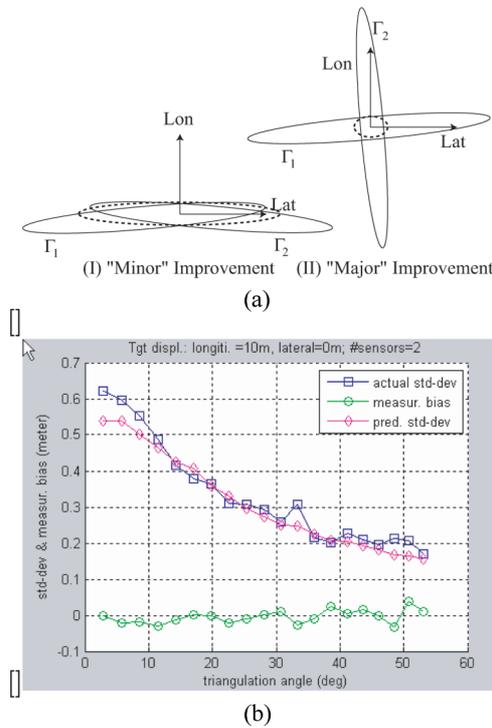


(b)

It is interesting to observe that the estimated error of lateral displacement is a function of angle ϕ shown in Figure 3(a). The plot (a) in Figure 6 illustrates the basic idea. We assume that estimation uncertainty can be approximated by a

Gaussian distribution, illustrated by an uncertainty ellipsoid in the state space. In the plot, the solid ellipses indicate measurement uncertainties, and the dash ellipses represent the fused estimates described in Section 2.4. Although in both cases, the uncertain areas are reduced, the fused uncertainty in Case (I) along the long principal axis of the ellipse is still big. This is directly caused by the sensor's poor accuracy in measuring azimuth angle, and angle ϕ between the sensors is small. A big improvement is observed in Case (II) when the orientations of the sensors are perpendicular. We can observe that the uncertainty is reduced tremendously along both axes of the fused ellipse. The plot (b) in Figure 6 confirms the above discussion by showing that the standard deviation of the fused estimate decreases as ϕ approaches $\pi/2$.

Figure 6 (a) Accuracy improvement varies with different sensor configurations and (b) The standard deviation curve of lateral displacement measure with increasing angle ϕ (see online version for colours)



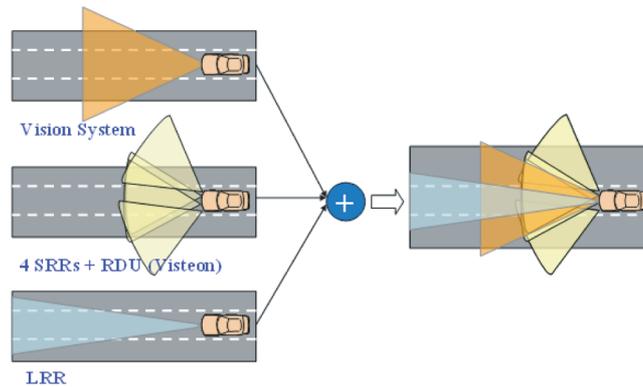
4 In-vehicle testing results

The proposed algorithm is implemented in a prototype embedded system with 800 MHz PowerPC processor, CPU usage of two percents is reported in tracking 20 objects.

As shown in Figure 7, the FOV of Frequency-Modulated Continuous Wave (FMCW) radar is limited to 15° . For ADAS systems beyond ACC, this long

range radar alone might not be sufficient (Mobus and Kolbe, 2004). To improved ACC, two extra sensor types was added: A forward multi-sensor tracking system with a camera with vision and 24GHz short range radars. Figure 7 illustrates how the 45° FOV vision system and 60° FOV short range radars cover a large area uncovered by the ACC radar. In addition to objects from LRR, the system utilises the objects from the Vision System (VIS) and from Short Range Radars (SRR).

Figure 7 Forward sensor suite (see online version for colours)

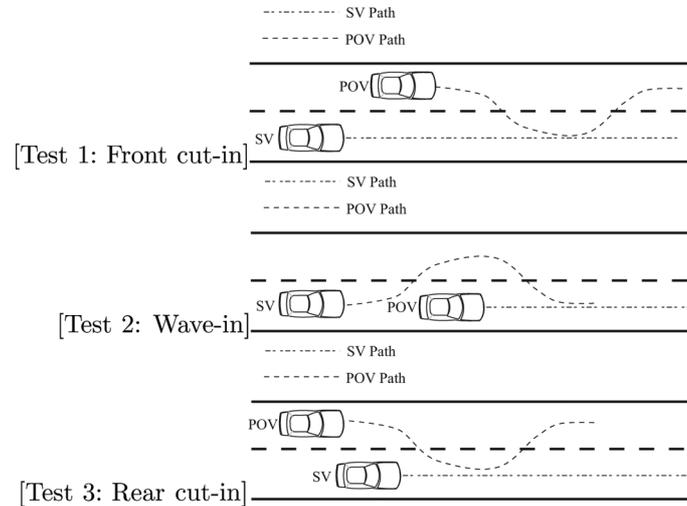


4.1 Cut-in vehicle tracking

We report both results of the proposed system and a commercial ACC system with single LRR sensor in several system-level tests. These tests compare the latencies of the primary object acquisition modules of the both implemented systems.

Figure 8 shows cut-in vehicle detection scenarios. The scenarios begin with the SV travelling on a flat paved road at speed 30mph. Next to the SV in the adjacent lane is a single Principal Other Vehicle (POV) travelling at a different speed than that of the SV. The POV passes the SV and then changes to the same lane as the SV at a distance approximately 10–20 m. The POV keeps the lane for a while, switches back to its original lane in front of the SV.

As illustrative examples, the results of a run of Test 3 are presented in Figure 9. The screen snapshot of a data logging software is shown in (a), including a top down view object map and a synchronised video. The three blue lines in the top down view denote the instantaneous vehicle path; and red cross, red square, and blue circle denotes the detected objects by LRR, VIS, and SRR, respectively. The top and bottom curves in plot (b) presents the longitudinal measurements: displacement (x) and velocity (v_x), respectively. As can be seen from these curves, the measurements from vision (denoted by VIS) are very crude. The POV changes lane at a range about 20 m away. Plot (c) illustrates the result of object validation of both systems: the commercial system (denoted by DELPHI) and the proposed system. The red dash-dot curves represent the boolean variables that determine whether the object is selected as the Closest In-Path Vehicle (CIPV) for the ACC feature. In the two diagrams of plot (c), as the POV completes the Lane Change

Figure 8 Manoeuvre diagrams for ACC cut-in vehicle detection

(LC), the lateral displacements approach zero. The commercial system selects the object as the CIPV when its lateral distance is less than about 1.0 m. However, the proposed system can claim the object as the CIPV at about 2.0 meters. The commercial system has to be conservative because of the low lateral measurement accuracy in order to reduce the rate false object acceptance. On the other hand, the proposed system can validate the object earlier because

- 1 The lateral estimate in the proposed system is more accurate with the help of vision.
- 2 The merged field-of-view in the proposed system is much larger than that of the commercial system. Hence, the object was tracked much longer in the proposed system than in the commercial system before being validated as the CIPV. As can be seen from plot (c), the POV was observed by the commercial system only after time 5 seconds. The proposed system, in meanwhile, had tracked the object 3 seconds earlier than what the commercial system does.
- 3 Independent sources of information (i.e., from different sensors) are acquired for the same object.

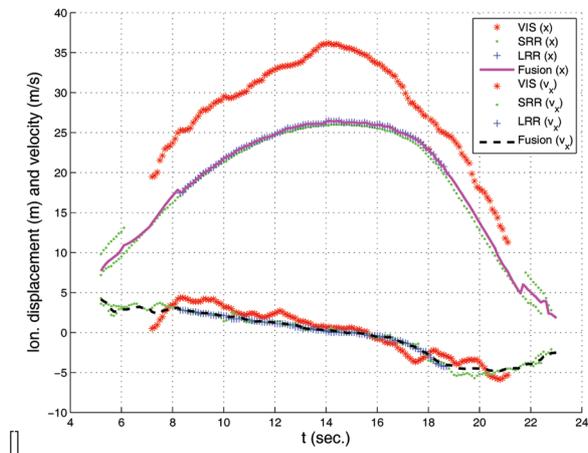
From plot (c), the POV was validated as the CIPV at time 5 seconds by the proposed system and at time 8 seconds by the commercial system. About 3 seconds are gained by the proposed system, which is significant as it enables building more sophisticated *feedforward* vehicle control systems.

As we have discussed, the measurement covariance matrices are introduced to handle sensors with complementary performance characteristics. It is interesting to observe how the results of the proposed system are affected by the sensors.

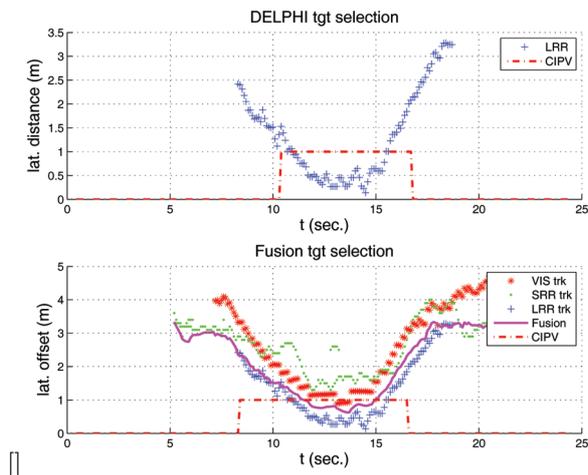
Figure 9 The result of a run of Test 3 (see online version for colours)



(a)



(b)



(c)

For example, longitudinal measurements from vision system (denoted by red symbol ‘*’) are very crude, and hence, as shown in plot (b) of Figure 9, the fused estimates (denoted by a magenta line) are not correlated with the vision measurements. However, as shown in plot (c), the lateral displacement measurements (denoted by ‘*’) from vision contribute significantly to the estimates (denoted by a dark red solid line) of the proposed system.

Moreover, Figure 9 also shows the proposed system can track the same object across different sensors. As shown in plots (c) and (d) of Figure 9, in the process of the POV’s LC manoeuvre, the target vehicle was

- Detected by SRR (measurements denoted by green dots) at $t = 5$ seconds
- Detected by VIS (measurements denoted by symbol ‘*’) at $t = 7$ seconds
- Detected by LRR (measurements denoted by cross ‘+’) at $t = 8$ seconds
- LRR lost the track at $t = 19$ seconds
- VIS lost the track at $t = 21$ seconds
- SRR lost the track at $t = 23$ seconds

During the whole LC manoeuvre, in the proposed system the target transits from one-sensor, two-sensor, three-sensor, two-sensor and finally to one-sensor configurations smoothly without having significant discontinuity in estimation.

In summary each test in Figure 8 was executed about 20 runs including combinations of left/right sides and straight/curved lanes. Figure 10 demonstrates the proposed system can validate and output the object about 1 second earlier than LRR-based commercial system.

4.2 Never-seen stopped vehicle detection

Never-seen stopped vehicle detection is an unsolved problem in current commercial ACC systems. Stationary objects are intentionally not registered for the reasons of avoiding possible hazards caused by events of false braking during normal operation. Recent ACC systems approach stop and go scenarios by tracking the objects’ historical kinematic status (i.e., objects moved before). However, none of the ACC systems can demonstrate intended system behaviour by using never-seen stopped vehicles as targets (Bishop, 2005).

The proposed ACC system select a front object as a CIPV if and only the object is detected by both vision and radar systems. This cross-validation strategy helps to reduce the possibility of false braking.

As shown in Figure 11, this test begins with the Subject Vehicle (SV) travelling on a straight flat paved road at a speed of 35mph. Ahead of the SV, in the same lane, is the first Principal Other Vehicle (POV1) travelling at the same speed as the SV. The SV is following at a moderate distance. Ahead of the POV1 is the second stationary POV (POV2). The POV1 executes a LC manoeuvre prior to colliding with POV2.

Figure 10 (a) Number of runs in each test. We split runs of each test into two categories: left and right and (b) Max-min-average plot of detection time difference (see online version for colours)

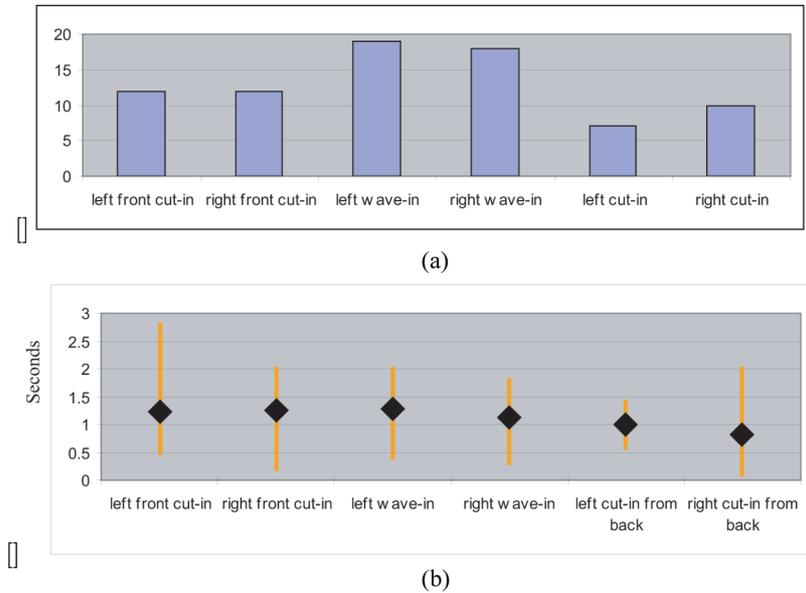
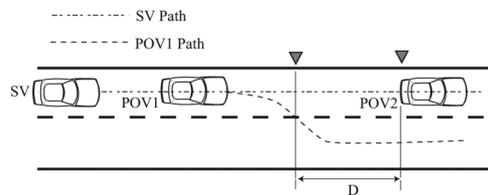


Figure 11 Manoeuvre diagram for never-seen stopped vehicle detection. Quantity D denotes the distance between the stopped vehicle and the position when the POV1 completes the LC manoeuvre



This test especially explores the ability of the proposed system to accurately report stationary in-path targets and, then, determines whether the braking request occurs at a range that is consistent with the collision and ACC onset braking requirement.

In LRR based ACC the never-seen stationary objects are not registered in order to reduce the false object acceptance rate. However, with the vision system, the proposed system can discriminate a stopped in-lane vehicle among stationary roadside objects.

Ten runs of the test with different quantity D were executed at an automotive proving ground. We first show the fifth run as an illustrative example. Then the key parameter—the range when the observations of the POV2 was validated first by the proposed system is analysed.

Plot (a) in Figure 12 shows the screen snapshot of the logging software at time 5.854 seconds in the log file when the POV2 was first detected by LRR. Plot (b)

shows the screen snapshot when the POV2 was validated by the proposed system at 8.004 seconds.

Figure 12 (a) The snapshot when the POV2 was first detected by LRR at 5.854 seconds and (b) The snapshot when the POV2 was validated and reported by the proposed system at 8.004 seconds. The green squares in the left window are detected LRR tracks, blue squares denotes fused tracks, and red large circle denotes vision track. The video was taken by a camera mounted on windshield. The left vehicle in the video is POV2 (see online version for colours)

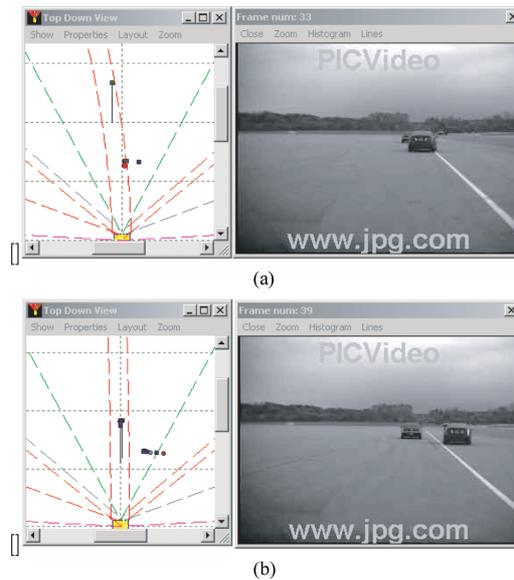


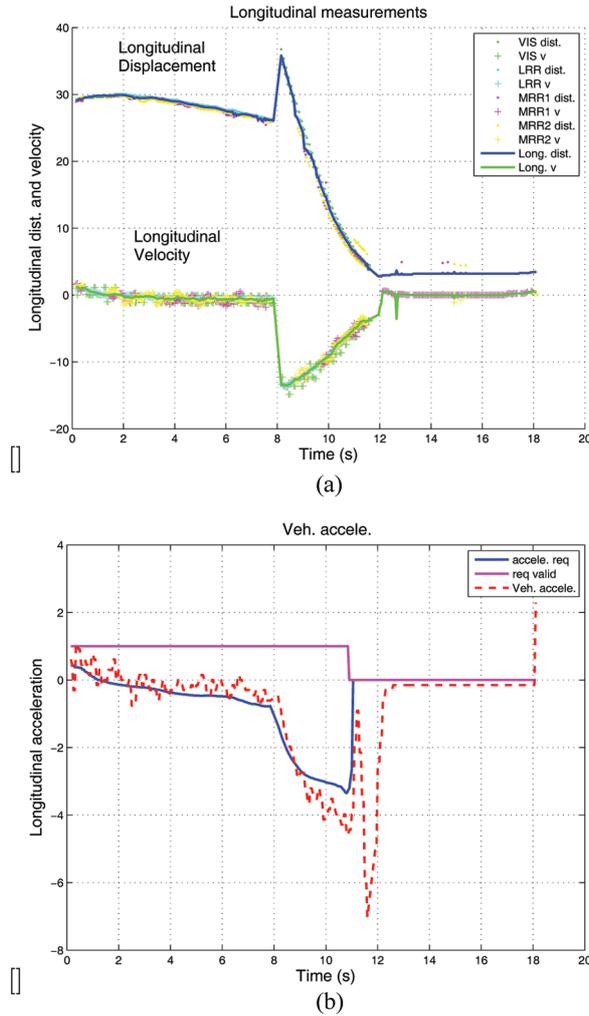
Figure 13 illustrates the behaviour of the SV in the fifth run. Plot (a) shows the longitudinal measurements of the ACC object. The vehicular control and response signals are shown in plot (b). The fused target was not selected as a valid object until the vision system detected the stopped POV2 at time 8 seconds. The SV starts to brake harshly at time 8 seconds, and increases brake energy until the vehicle is full stopped at time 12 seconds.

In order to have a smooth braking profile, it is important to detect and validate the stopped POV2 object as early as possible. In the proposed system, this relies on how early the vision system detects the POV2. As shown in Figure 14, although LRR detects the stopped POV2 about 15 m farther than the vision system does, the proposed system has to defer the decision until the object is confirmed by the vision system, which usually comes later.

5 Summary and conclusions

This paper describes the developed multi-sensor tracking algorithm, associated software modules and tests that have been executed. Based on the results of the

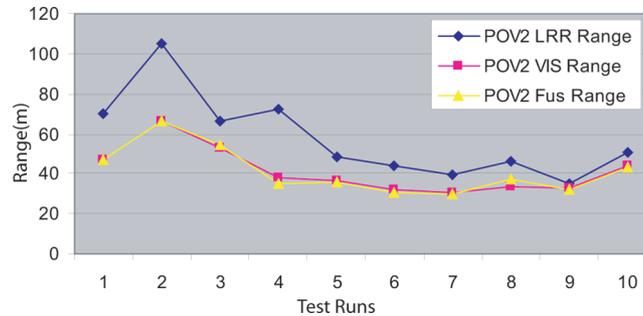
Figure 13 The result of a run of Test 4. (a) The curves of longitudinal displacement and velocity measurement and (b) Vehicle deceleration request. The solid curve with label ‘acce req’ denotes the ACC request acceleration, the solid curve with label ‘req valid’ indicate the validity of ‘acce req’, and the dash curve is the actual vehicle acceleration (see online version for colours)



work, the author offers the following conclusions:

- i The proposed system reduces the rates of missed detection and false detection in the multiple sensor overlapping region, by using the strengths and the independent sources of information from different sensors.
- ii Integrating data from multiple sensors extends the coverage of ACC. By tracking the same object smoothly across different sensors, the proposed system shortens the needed cycles to complete track initialisation.

Figure 14 The ranges when the POV2 were first reported by LRR, by the vision system, and by the proposed system (see online version for colours)



- iii For the ACC application, the proposed system validates objects about one second earlier than LRR-only system and hence provides an opportunity for the vehicle control system to improve smoothness and responsiveness.

Although promising results have been achieved using the proposed system, we observe performance degradation in one or two scenarios where the assumption of Gaussian distribution is inappropriate for modelling the noise of the measurement and system dynamic processes. In response to this issue, future work may use alternative filtering techniques (e.g., *switching Kalman filter* or *particle filter*) but, in the meantime, requiring more computing resource.

References

- Becker, J.C. (2000) 'Fusion of heterogeneous sensors for the guidance of an autonomous vehicle', *The Third International Conference on Information Fusion*, Paris, France, pp.11–18.
- Bishop, R. (2005) *Intelligent Vehicle Technology and Trends*, Artech House, Norwood, MA.
- Gunes, H. and Piccardi, M. (2005) 'Affect recognition from face and body: Early fusion vs. late fusion', *IEEE International Conference on Systems, Man and Cybernetics*, Hawaii, USA, Vol. 4, pp.3437–3443.
- Hall, D.L. (1992) *Mathematical Techniques in Multisensor Data Fusion*, Artech House, Norwood, MA.
- Kaempchen, N., Buehler, M. and Dietmayer, K. (2005) 'Feature-level fusion for free-form object tracking using laserscanner and video', *IEEE Intelligent Vehicles Symposium*, Las Vegas, USA, pp.453–458.
- Labayrade, R., Royere, C. and Aubert, D. (2005) 'A collision mitigation system using laser scanner and stereovision fusion and its assessment', *IEEE Intelligent Vehicles Symposium*, Las Vegas, USA, pp.441–446.
- Luo, R.C., Yih, C. and Su, K.L. (2002) 'Multisensor fusion and integration: Approaches, applications, and future research directions', *IEEE Sensors Journal*, Vol. 2, No. 2, pp.107–119.
- Mobus, R. and Kolbe, U. (2004) 'Multi-target multi-object tracking, sensor fusion of radar and infrared', *IEEE Intelligent Vehicles Symposium*, Parma, Italy, pp.732–737.

- Redmill, K.A., Martin, J.I. and Ozguner, U. (2006) 'Sensing and sensor fusion for the 2005 desert buckeyes darpa grand challenge offroad autonomous vehicle', *IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, pp.528–533.
- Skutek, M., Linzmeier, D.T., Appenrodt, N. and Wanielik, G. (2005) 'A precrash system based on sensor data fusion of laser scanner and short range radars', *2005 8th International Conference on Information Fusion*, Philadelphia, PA, Vol. 2, pp.1287–1294.
- Smith, D. and Singh, S. (2006) 'Approaches to multisensor data fusion in target tracking: A survey', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 18, No. 12, pp.1696–1710.
- Weng, J.J., Huang, T.S. and Ahuja, N. (1993) *Motion and Structure from Image Sequences*, Springer-Verlag, Berlin.

Note

¹Note that \mathbf{o} is now redefined to be $\mathbf{o} = (r, \theta, v_r, v_t)$.