Real-time Monocular Ranging by Bayesian Triangulation

Katsuyuki Nakamura, Kazuma Ishigaki, Takehito Ogata and Shoji Muramatsu

Abstract—A method for estimating the range between moving vehicles by using a monocular camera is proposed. Although most conventional methods use vertical triangulation, the proposed method uses both vertical and horizontal triangulation, which reduces errors due to vehicle’s own pitching in the far distance. Unknown vehicle width is estimated by measuring three ranging parameters associated with an image captured by the camera, and the following distance is then computed by horizontal triangulation. Both vehicle width and following distance are dynamically updated during the vehicle-tracking process by single filtering. The proposed method runs in real time and can produce highly accurate estimation of following distance. The key contribution of this study is the coupled estimation of unknown vehicle width and following distance by sequential Bayesian estimation.

I. INTRODUCTION

Advanced driver-assistance systems (ADAS) are rapidly becoming more popular. Typical ADAS applications, including forward-collision warning (FCW) and lane departure warning (LDW), have already been selected as standard evaluation items in the United States new car assessment program (US-NCAP). There are several other regulations expected to be implemented in the near future, including requiring vehicles to be equipped with a rear camera in the U.S. and requiring large commercial vehicles to be equipped with autonomous emergency braking (AEB) in Europe. Active safety via ADAS is expected to become more prominent in the future.

There are three approaches for implementing FCW: radar-based, laser-based, and vision-based. The vision-based approach is the most cost-effective because it can be used for both FCW and LDW applications using only the camera itself. There has been a lot of interest in and research on object detection from the stereo camera (e.g., [1]–[5]), but the monocular camera is still strongly advantageous in terms of its large sensing area, ease of installation, and low cost. There is one disadvantage, however, in that it is difficult for a monocular camera to achieve accurate range estimation.

In this paper, we propose a method for estimating the range between moving vehicles by using a monocular camera. Although most conventional works have used only vertical triangulation, the proposed method uses both vertical and horizontal triangulation based on an estimation of unknown vehicle width, thereby reducing errors due to a vehicle’s own pitching (a serious issue in conventional works). The primary contribution of this study is the innovative use of the coupled estimation of unknown vehicle width and following distance by sequential Bayesian estimation.

This paper is structured as follows. Section II reviews existing research on range estimation using a monocular camera. Section III-A describes an overview of our method. Section III-B shows an example of vehicle detection prior to range estimation. Sections III-C and III-D describe the method for monocular ranging based on the estimation of an unknown vehicle width. Section IV shows performance evaluations in real scenes.

II. RELATED WORK

Although there has been much research on vehicle detection using the monocular camera (e.g., [6]–[8]), there has been relatively less on methods for range estimation. As a pioneer work, Stein et al. [9] proposed calculating the following distance on the basis of vertical triangulation. They detected a contact position between a vehicle and the road and then computed the following distance on the basis of the planar road assumption. Dagan et al. [10] proposed directly computing time-to-collision (TTC) without a range estimation that used the enlargement ratio of a vehicle in an image sequence. Dellaert et al. [11] presented a real-time system for detecting and tracking vehicles with a range estimation. They combined a vehicle detector based on kernel regression and an extended Kalman filter for estimating a vehicle’s position. Yin-Yu et al. [12] achieved monocular ranging at night using a vehicle’s taillight and license plate. There have been several other studies for range estimation, including projective invariant [13] and utilization of a known 3D model [14] etc.

The above examples show that although there have been many methods using a monocular camera proposed, it remains a challenge to achieve robust ranging. For example, it is difficult to achieve accurate estimation in the far distance, which results in several difficulties such as lower resolution of a target, small object motion, bigger pitching effect, etc.

In contrast with the conventional methods, the proposed method computes the following distance by using both vertical and horizontal triangulation based on an estimation of unknown vehicle width. It is less susceptible to vehicle’s own pitching because the ranging parameters mainly consist of horizontal factors. Furthermore, the method does not require prior knowledge of a target vehicle (e.g., real width, real height, and size of license plate) for range estimation. Unknown vehicle width and following distance are effectively computed in a practical manner.
III. PROPOSED METHOD

A. Overview

The proposed method, shown in Fig. 1, consists of two parts: real width estimation and Bayesian triangulation.

The real width estimation is to compute unknown vehicle width based on the camera geometry. Because there are so many different types of vehicles – sedans, coupes, wagons, trucks, etc. – vehicle width is unknown and varies from about 1.5 m to 3.0 m [9]. We found that unknown vehicle width can be computed by an image measurement of three parameters \((w, y_v, \text{ and } y_b)\). The vehicle width for range estimation is then computed. After that, Bayesian triangulation is performed to update vehicle width and following distance by sequential Bayesian estimation.

These processes are described in detail in the following sections.

B. Vehicle Detection

In this work, we used a classifier trained by AdaBoost with Haar-like features [15] for the vehicle detection.

\[
H(x) = \text{sign} \left( \sum_{t=1}^{T} \alpha_t h_t(x) \right),
\]

where \(H(x)\) represents a strong classifier, \(\alpha_t\) is a weighting factor, \(h_t(x)\) is a weak classifier, \(T\) is the number of weak classifiers, and \(x\) represents part areas in an image. The sign function returns +1 if an input variable is positive and −1 if it is negative. The weak classifier is defined by

\[
h_t(x) = \begin{cases} 
+1 & \text{if } f_t(x) > \theta_t \\
-1 & \text{otherwise} 
\end{cases},
\]

where \(f_t(x)\) and \(\theta_t\) represent a \(t\)-th Haar-like feature and its threshold, respectively.

Generally speaking, this classifier can obtain only a rough region around a vehicle. To extract a more accurate region, therefore, edge-based fitting is conducted. To achieve this fitting, edge detection is applied to the neighboring area around the rough region, and then a histogram of edge intensity in each row and column of the region is created. After that, by computing the maximums of the two histograms, the optimal boundary of the vehicle can be determined.

Figure 2 shows an example of an initial detection and a corrected result. By applying the fitting, we can obtain more accurate parameters for use in the next step.

C. Estimation of Unknown Vehicle Width

In this section, we describe a method of estimating unknown vehicle width on the basis of camera geometry. The pin-hole camera model enables us to express the geometric relationship between world and camera coordinate, as shown in Fig. 3. In the figure, \(H_c\) is camera mounting height, \(Z\) is following distance, \(W\) is the real width of a target vehicle, and \(h = y_b - y_v\) and \(w = x_R - x_L\) are projected camera height and vehicle width, respectively, in the imaging surface. Both parameters are corrected lens aberrations. According to this geometry, following distance can be calculated by two methods.

The first method is vertical triangulation [9], shown in Fig. 3(a).

\[
Z = \frac{f H_c}{y_b - y_v}
\]

This is a well-known method for monocular ranging and can easily compute the following distance. However, estimation errors tend to occur due to the vehicle’s own pitching.

The second method is horizontal triangulation, shown in
Unlike vertical triangulation, horizontal triangulation is less susceptible to vehicle’s own pitching because parameters $W$ and $w$ do not relate to vertical parameters. However, the real width of a vehicle is unknown, so ranging errors would occur if we were to use a statistical mode value of vehicle width (e.g., 1.7 m).

In this study, we combined these methods to tackle each problem. According to equations (3) and (4), we can calculate the real width of a vehicle ($W$) by

$$W = \frac{wH_c}{y_b - y_v}. \quad (5)$$

In other words, unknown vehicle width can be calculated by using both the reference length $H_c$ and the three ranging parameters ($w$, $y_v$, and $y_b$).

However, this equation consists of vertical parameters, so estimation errors caused by pitching still tend to occur. These errors are reduced by eliminating outliers during the tracking process, thereby achieving robust ranging. In the next section, we describe the proposed method combined with Bayesian estimation.

**D. Monocular Ranging by Bayesian Triangulation**

In this step, we first conduct monocular ranging based on the estimated vehicle width. Next, detected vehicles are tracked in the world coordinate based on the monocular ranging. Finally, the estimated distance and vehicle width are updated based on the tracking result. We implement this feedback loop by sequential Bayesian estimation. There are several well-known approaches to this type of loop, including the Kalman filter, unscented Kalman filter, particle filter, and others [16]. We chose to use the Kalman filter [17] because of its quick computational time and low memory consumption.

Figure 4 shows a conceptual diagram of the proposed Bayesian triangulation. First, the real width of a vehicle is computed by equation (5). Second, monocular ranging is conducted to estimate the $XZ$ positions of the target vehicles in the world coordinate. Third, vehicle positions in the next frame are predicted by assuming a uniform accelerated motion for the target vehicles. We extract the ranging parameters ($w$ and $y_b$) by image fitting of a projected rectangle. Based on these measurements, the estimated distance and vehicle width are dynamically updated.

This framework has two distinct advantages: (1) estimation errors related to following distance and vehicle width can be reduced during the tracking process, and (2) a fast response can be obtained due to the use of single filtering.

1) **State definition:** We define the state vector using positions $X, Z$, velocities $vX, vZ$, and accelerations $aX, aZ$ in the world coordinate:

$$x_t = [X, Z, vX, vZ, aX, aZ]^T \quad (6)$$

$X$ and $Z$ axes represent the horizontal and vertical directions, respectively (see Fig. 4). Posterior expected values were sequentially computed by algorithm 1.

**Algorithm 1 Bayesian triangulation**

1: **Step 1: Motion prediction**
2: $x_t^− ← Fx_{t−1}^−$
3: **Step 2: Observation**
4: $w, y_b$ ← computed by image fitting of $x_t^−$
5: $y_v$ ← computed by camera angle
6: if Initialization phase then
7: $W_t ← \frac{wH_c}{y_b - y_v}$
8: else
9: $W_t ← W_{t−1}^+$
10: **endif**
11: $Z_t ← \frac{W_t}{w}$
12: $z_t ← [Z_t \tan \theta, Z_t]^T$
13: **Step 3: Filtering**
14: $x_t^+ ← x_t^− + K_t(z_t - Hx_t^−)$
15: $y_b^+ ←$ computed by $x_t^+$
16: $W_t^+ ← \frac{wH_c}{y_b^+ - y_v}$

2) **Motion Prediction:** We predict the current state vector $x_t^+$ based on the previous state vector $x_{t−1}^−$ by using state transition matrix $F$. We assume a uniform accelerated motion for target vehicles.

$$F = \begin{bmatrix} 1 & 0 & \Delta t & 0 & 0 & \frac{1}{2} \Delta t^2 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & \frac{1}{2} \Delta t^2 & 0 \\ 0 & 0 & 1 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (7)$$

3) **Observation:** To obtain accurate parameters, edge-based fitting is performed using the image projection of $x_t^−$ as an initial rectangle. The corrected rectangle can then produce accurate $w$ and $y_b$. Vanishing point $y_v$ is computed by using the camera angle of depression. We compute the vehicle width $W_t$ on the basis of these parameters. We can then calculate the following distance $Z_t$ as

$$Z_t = \frac{fW_t}{w}. \quad (8)$$

Measurement vector $z_t$ is set up on the basis of angle $\theta$.
between an image center and vehicle position:
\[
z_t = [Z_t \tan \theta, Z_t]^T \tag{9}
\]

4) Filtering: Posterior expected values \(x_t^+\) are computed by using the innovation between \(z_t\) and \(x_t^-\), where \(H\) represents the measurement matrix and \(K_t\) is the Kalman gain.
\[
H = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0
\end{bmatrix} \tag{10}
\]

\(y_t^+\) is computed by projecting an updated \(x_t^+\) to the image. We can obtain maximum a posteriori (MAP) estimates of vehicle width \(W_t^+\) by using range estimation after an initialization phase. This enables us to stably update the vehicle width while eliminating outliers caused by vehicle pitching.

IV. Experiment

A. Experimental Setup

We used experimentation to evaluate the effectiveness of the proposed method. A test vehicle was equipped with a monocular camera (viewing angle of 37.2°, VGA resolution with 10 fps) and laser radar. The camera mounting height was 1.35 m and the angle of depression was 4.3°. The camera was setup on the inside of a wind shield while the laser radar was mounted on front of the vehicle.

We implemented the proposed algorithm on a system on a chip (Renesas SH series with 300-MHz CPU and 256-KB RAM). This system ran in real time, and the computational time for ranging was roughly 1 ms.

B. Results

First, we evaluated the system performance on a straight track under a rainy condition. The host vehicle approached a stopped vehicle at a speed of 32 km/h. The implemented system then computed the vehicle width and the following distance. Figure 5 shows the estimation result. The upper graph shows the estimated vehicle width. The average estimation error was just 2% compared with the real width of the target vehicle (1.85 m). However, these results also show some bias in the initialization phase. The main reason for this bias is difficulties in measuring parameters \(w\) and \(y_b\) because the region around a captured vehicle is small and the image is slightly blurred by rain. The lower graph shows the estimated distance. The distance estimates obtained by the camera were in good agreement with the estimates by the laser radar.

Second, we tested the robustness of the proposed method in consecutive experiments. The experimental condition was based on the FCW-NCAP Test 1 [18], in which a host vehicle approaches a stopped vehicle at a speed of 72 km/h. Seven consecutive experiments were performed under a rainy condition. Figure 6 shows the ranging accuracy obtained by all experiments. The horizontal axis shows the true distance measured by the laser radar and the vertical axis shows the distance estimated using the monocular camera. There is a strong correlation between the estimation and the ground truth. Average ranging error was 2.67 m and standard deviation was 1.27 m. We found excellent coefficients of determination \(R^2 = 0.99\).

Third, we compared the proposed method with the vertical triangulation (Equation 3) combined with Kalman filter. Table I shows a comparison of the ranging performance. The proposed method can reduce both the average errors \(\mu_Z\) and the standard deviation \(\sigma_Z\). Improvement in the far distance was greater than that in the near distance. The average error ratio of over 50 m was reduced from 6.3% to 4.5%. Standard deviation was also reduced by about 47%. These results show that the proposed method can reduce errors due to vehicle pitching.

Figure 7 shows examples of monocular ranging on public roads. \(Z\) and \(W\) represent the following distance and the vehicle width, respectively. We were able to achieve good range estimates with various vehicle types. However, estimation errors increased in the ramp excursion and the adjacent lanes. Our future work will focus on improving the robustness.
Fig. 7. Examples of monocular ranging on public roads. Z is following distance and W is estimated vehicle width.

### TABLE I

<table>
<thead>
<tr>
<th>Distance</th>
<th>Vertical triangulation</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_Z$</td>
<td>$\sigma_Z$</td>
</tr>
<tr>
<td>0–50 m</td>
<td>2.53 m</td>
<td>1.28 m</td>
</tr>
<tr>
<td>50–100 m</td>
<td>4.00 m</td>
<td>2.67 m</td>
</tr>
<tr>
<td>Total</td>
<td>2.99 m</td>
<td>1.94 m</td>
</tr>
</tbody>
</table>

**V. SUMMARY**

A method for monocular ranging from a moving vehicle was proposed. It relies on a Bayesian framework based on an estimation of unknown vehicle width, which is estimated by using three parameters, and following distance. A sequential Bayesian framework dynamically updates the following distance and vehicle width. The proposed method reduces errors due to vehicle pitching and can therefore deliver robust estimation of following distance even at long range. It was demonstrated that an FCW system using this method can run in real time and provide excellent ranging performance.

**VI. ACKNOWLEDGMENTS**

The authors thank Hitachi Automotive Systems, Ltd. for supporting the experiments.
REFERENCES


