Vision-based ACC with a Single Camera: Bounds on Range and Range Rate Accuracy

Gideon P. Stein
MobileEye Vision Technologies Ltd.
Jerusalem, Israel
gideon.stein@mobileye.com

Ofer Mano
MobileEye Vision Technologies Ltd.
Jerusalem, Israel
ofer.mano@mobileye.com

Amnon Shashua
Hebrew University
Jerusalem, Israel
shashua@cs.huji.ac.il

Abstract

This paper describes a Vision-based Adaptive Cruise Control (ACC) system which uses a single camera as input. In particular we discuss how to compute range and range-rate from a single camera and discuss how the imaging geometry affects the range and range rate accuracy. We determine the bound on the accuracy given a particular configuration. These bounds in turn determine what steps must be made to achieve good performance. The system has been implemented on a test vehicle and driven on various highways over thousands of miles.

1 Introduction

The Adaptive Cruise Control (ACC) application is the most basic system in the evolutionary line of features where sensors in the vehicle assist the driver to increase driving safety and convenience. The ACC is a longitudinal distance control designed to find targets (other vehicles), determine their path position (primary target determination), measure range and range-rate to the primary target vehicle and perform appropriate brakes and throttle actuation to maintain safe distance to the primary target and to resume the preset cruising speed when no such targets are detected by the system. The basic ACC feature is offered today (as a convenience feature) in serial production models by an increasing number of car manufacturers.

The underlying range measurement technology of existing systems falls into the category we call “direct range” sensors which include millimeter wave radars (77GHZ radars mostly)[1], Laser Radars (LIDAR) and Stereo Imaging (introduced in Japan on the Subaru Legacy Lancaster[2]). These sensors provide an explicit range measurement per feature point in the scene. The range map provides strong cues for segmenting the target from the background scene and, more importantly to this paper, explicit range is then being used for distance control.

In this paper we investigate the possibility of performing distance control, to an accuracy level sufficient for a serial production ACC product, using a monocular imaging device (a single video camera) which provides only “indirect range” using the laws of perspective (to be described below). This investigation is motivated by two sources: first is biological vision and second is practical. In the human visual system the stereo base-line is designed for hand-reaching distances and for very rough approximate range measurements at farther distances. Distance control in an ACC application requires range measurements of distances reaching 100m where a human observer cannot possibly make accurate absolute range estimations at that range. Moreover, many people suffer from stereo deficiency without any noticeable effect on the daily visual navigation (and driving) abilities. On the other hand, based on retinal divergence (scale change of the target) the human visual system can make very accurate “time to contact” assessments. Therefore, the question that arises in this context is what are the *necessary* measurement accuracies required for a distance control? clearly, the accuracies of range provided by Radar and LIDAR are *sufficient* for distance control, but the example of human vision indicate that perhaps one can achieve satisfactory actuation control using only the laws of perspective. The second motivation is practical and is borne out of the desire to introduce low-cost solutions for the ACC application. A stereo design not only includes the cost of the additional camera and processing power for dense disparity but also the problem of maintaining calibration of the system (relative coordinate frames between the two cameras) is somewhat challenging for a serial production product[3, 4]. A monocular visual processing system would be easier to mass produce and would cost less as an end product.

The challenges of a monocular visual system are twofold. On the one hand, the system lacks the depth cues[1] used for target segmentation and instead pattern recognition techniques should be heavily relied on to compensate for the lack of depth. The question that arises there is whether pattern recognition can be sufficiently robust to meet the stringent detection accuracy requirements for a serial production product? On the other hand, and this is the focus of this paper, once the target is detected can the laws of perspective and retinal divergence meet the required accuracies for actuation control?

[1]At short distances one can rely on some weak motion parallax measurements but those are not available at ranges beyond 20-30m.

We have built a monocular visual processing system targeted for mass production over a wide range of applications. Fig. 1 shows the camera mounted near the rear-view mirror. The prototype processing box used at this stage of development is based on the PPC7410 processor.[2]. The system runs at 10 frames per second and performs target detection (vehicles and motorcycles) in the lane and adjacent lanes, lane mark detection and following, lane departure warning and cut-in calculation using optic-flow analysis. Finally, to be used for control, it determines the range and range rate information about the target vehicles.

Of the many task required by the vision sensor this paper will focus on the issue of determining the range and range rate. After discussing the methods for computing range and range rate we provide an analysis of the accuracy of those measurements. Finally we show some results taken during closed loop operation and we compare those results to measurements using Radar.

2 Range

Since we have only a single camera we must estimate the range using perspective. There are two cues which can be used: size of the vehicle in the image and position of the bottom of the vehicle in the image. Since the width of a vehicle of unknown type (car, van, truck etc) can vary anywhere between 1.5*m* and 3*m* a range estimate based on width will only be about 30% accurate. It can be used as a sanity check and possibly to check *out-of-calibration* but it is not good enough for actuation control.

A much better estimate can be achieved using the road geometry and the point of contact of the vehicle and the road. We will at first assume a planar road surface and a camera mounted so that the optical axis is parallel to the road surface. A point on the road at a distance Z in front of the camera will project to the image at a height y, where y is given by the equation:

$$y = \frac{fH}{Z} \tag{1}$$

where H is the camera height in meters.

Figure 2 shows a diagram of a schematic pinhole camera comprised of a pinhole (P) and an imaging plane (I) placed at a focal distance (f) from the pinhole. The camera is mounted on vehicle (A) at a height (H). The rear of vehicle (B) is at a distance (Z_1) from the camera. The point of contact between the vehicle and the road projects onto the image plane at a position (y_1). The focal distance (f) and the image coordinates (y) are typically in *mm* and are drawn here not to scale.

[2]The serial production hardware is based on a system-on-chip called EyeQ — more details in http://www.mobileye.com

Figure 1: *Two pictures of the compact monocular camera mounted near the rear-view mirror.*

Equation 1 can be derived directly from the similarity of triangles: $\frac{y}{f} = \frac{H}{Z}$. The point of contact between the and a more distant vehicle (C) projects onto the image plane at a position (y_2) which is smaller than (y_1).

The camera electronics converts the image coordinates from *mm* to pixels and inverts the image back to the upright position for processing. Figure 3 shows an example sequence of a truck at various distances. The distance from the horizon line to the bottom of the truck is smaller when the truck is more distant (a) than when it is close (b and c).

To determine the distance to a vehicle we must first detect the point of contact between the vehicle and the road (i.e. the wheels) and then we can compute the distance to the vehicle:

$$Z = \frac{fH}{y}. \tag{2}$$

In practice the camera optical axis is not aligned parallel to the road surface. Thus the *horizon line* is not in the center

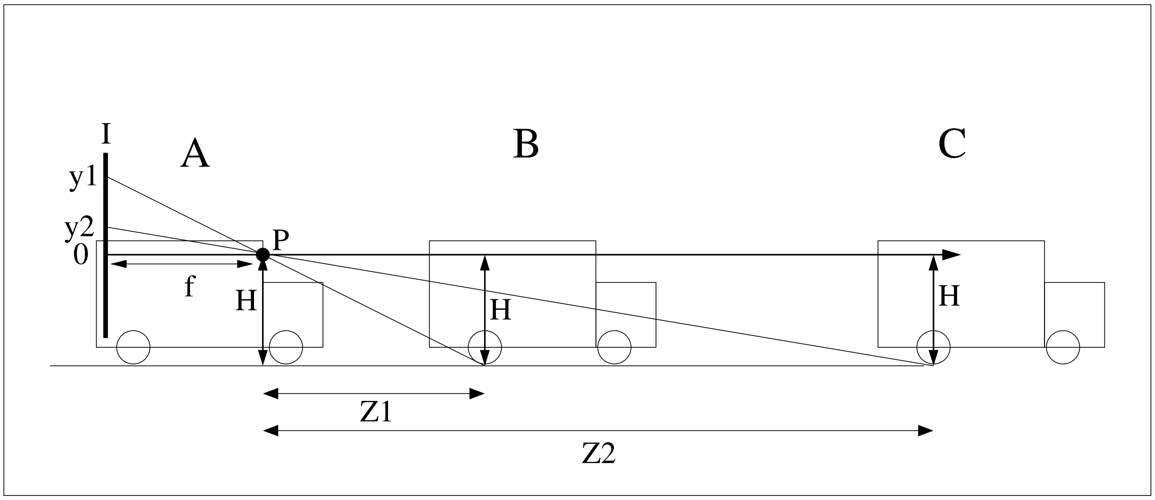


Figure 2: *Schematic diagram of the imaging geometry (see text).*

of the image. Both the mounting angle and the change in pitch angle due to vehicle motion can be determined by the method described in [5]. The non-planarity can be compensated for to a first order by analyzing the lane markings [6]. After compensating for the camera angle and road slope the main source of error is in determining the image coordinates of contact point between the vehicle and the road. In practice, this point can be found to within 1 pixel.

The error in range Z_{err} due to an error of n pixels in location of the contact point is:

$$Z_{err} = Z_n - Z = \frac{fH}{y+n} - Z = \frac{fH}{\frac{fH}{Z}+n} = \frac{nZ^2}{fH+nZ} \tag{3}$$

Typically $n \approx 1$ and $fH >> nZ$ so we get:

$$Z_{err} \approx \frac{nZ^2}{fH} \tag{4}$$

We see that the error increases quadratically and as the distance to the target increases the percentage error in depth ($\frac{Z_{err}}{Z}$) increases linearly.

Example: In our case a $640x480$ image with a horizontal FOV of 47^o gives $f = 740 pixels$. The camera height at $H = 1.2m$. Thus assuming 1 pixel error, a 5% error in depth is expected at a distance of:

$$Z = \frac{Z_{err}}{Z} fH = 0.05 * 740 * 1.2 = 44m. \tag{5}$$

The error at $90m$ will be around 10%. These values are sufficient for ACC. The headway distance depends on the vehicle speed and the driver *comfort setting* and is typically less than $45m$. A 5% error at $45m$ is well below the error of a human driver and is not significant. What is important is the range rate or relative velocity. In layman's terms, it is not important whether the target is at $45m$ or $42m$. It is important to know whether we are maintaining a constant distance.

3 Range Rate

With a radar system range rate or relative velocity can be measured using doppler effect. With a vision system it is computed from the discrete differencing:

$$v = \frac{\Delta Z}{\Delta t}. \tag{6}$$

Subtracting two noisy values for Z at two different time points cannot produce accurate measurements. We will first show (in sec. 3.1) how ΔZ can be computed from the scale change (s), that is the change in image size, and the range (Z). We will then show (in sec. 3.2) how the discrete difference introduces an error for non infinitesimal Δt which limits our accuracy. We also show how to achieve the optimal value for Δt.

3.1 Computing Range Rate from Scale Change

Let W be the width (or height) of the target vehicle in meters, and let w and w' be the width (or height) in the image in pixels when the target vehicle is at distances Z and Z' respectively. As in equation (1):

$$\begin{aligned} w &= \frac{fW}{Z} \\ w' &= \frac{fW}{Z'}. \end{aligned} \tag{7}$$

Then:

$$v = \frac{\Delta Z}{\Delta t} = \frac{Z'-Z}{\Delta t} = \frac{\frac{fH}{Z'} - \frac{fH}{Z}}{\Delta t} = \frac{fH\frac{w-w'}{w'w}}{\Delta t} = \frac{Z\frac{w-w'}{w'}}{\Delta t} \tag{8}$$

Let us define:

$$s = \frac{w-w'}{w'} \tag{9}$$

and we get:

$$v = \frac{Zs}{\Delta t} \tag{10}$$

3.2 Range Rate Error

The scale change can be computed by alignment of the image of the vehicle from images taken at two points in time t, t'. Various techniques for alignment of images can be used ([7, 8]) and alignment errors (s_{err}) of $0.1 pixels$ are possible if the image patch has a few hundred pixels. This is the case for example, of a small car at $75m$ which will be a $15x15$ rectangle in the image using a 47^o FOV lens.

The effect of an alignment error of $0.1 pixels$ depends on the size of the target in the image. Let us therefore define the scale error (s_{acc}) as the alignment error (s_{err}) divided by vehicle image width:

$$\begin{aligned} s_{acc} &= \frac{s_{err}}{w} \qquad (11) \\ &= \frac{s_{err} Z}{fW} \end{aligned}$$

If we assume the range (Z) is accurate the error in the relative velocity is:

$$\begin{aligned} v_{err} &= \frac{Z s_{acc}}{\Delta t} \qquad (12) \\ &= \frac{Z^2 s_{err}}{fW \Delta t} \qquad (13) \end{aligned}$$

Notes:

1. The relative velocity error is independent of the relative velocity.
2. The relative velocity error increases with the distance squared.
3. The relative velocity error is inversely proportional to the time window Δt. We can get more accurate relative velocity if we use images that are further apart in time.
4. Having a narrow field of view camera (i.e. increasing f) will reduce the error and increase accuracy linearly.

From eq. (10) we see that an error in range will have a corresponding error in relative velocity. Substituting eq. (4) into eq. (10) we can compute the velocity error (v_{zerr}) due to error in range:

$$v_{zerr} = \frac{Z_{err} s}{\Delta t} = \frac{nZ^2}{fH} \frac{s}{\Delta t} = \frac{nZv}{fH} \qquad (14)$$

Taking the range error (eq. 4) into account the velocity err becomes:

$$v_{err} = \frac{Z^2 s_{err}}{fW \Delta t} + \frac{nZv}{fH} \qquad (15)$$

Example: Following a small car at $30m$: $Z = 30m$, $f = 740 pixels$, $W = 1.5$, $h = 1.2m$ and $v = 0m/s$. We will use $\Delta t = 0.1s$.

$$\begin{aligned} v_{err} &= \frac{30^2 x 0.1}{740 x 1.2 x 0.1} + \frac{30 x 0}{740 x 1.2} \qquad (16) \\ &= 1m/s \qquad (17) \end{aligned}$$

It followt from eq. (15) that the accuracy of velocity estimation can be improved significantly (i.e. v_{err} reduced) by increasing the time window Δt. Tracking the vehicle over a few seconds is quite feasible and this is a good solution for *follow mode*.

However if we increase Δt we no longer have infinitesimal motion and computing velocity using finite differencing (eq. 8) is not accurate. In case of constant acceleration the range and velocity are:

$$\begin{aligned} Z(\Delta t) &= \frac{1}{2} a \Delta t^2 + v \Delta t + Z_0 \qquad (18) \\ v(\Delta t) &= a \Delta t + v_0 \qquad (19) \end{aligned}$$

where a is the relative acceleration. As we know, if we compute the velocity by taking the range difference at two time points:

$$\Delta Z = Z(\Delta t) - Z_0 = \frac{1}{2} a \Delta t^2 + v_0 \Delta t \qquad (20)$$

and dividing by Δt:

$$\frac{\Delta Z}{\Delta t} = \frac{1}{2} a \Delta t + v_0 \qquad (21)$$

we get a different result from eq. (19) by a term:

$$v_{FDerr} = \frac{1}{2} a \Delta t. \qquad (22)$$

Thus a longer Δt adds to the inaccuracy or error in v. If we add this term to eq. (15) we get:

$$v_{err} = \frac{Z^2 s_{err}}{fW \Delta t} + \frac{nZv}{fH} + \frac{1}{2} a \Delta t. \qquad (23)$$

There are two terms which depend on Δt. The first term is inversely proportional to Δt and third term proportional. Therefore we can find an optimal Δt which reduces the value of v_{err} by differentiating eq. (23) and setting to zero:

$$-\frac{Z^2 s_{err}}{fW \Delta t^2} + \frac{1}{2} a = 0. \qquad (24)$$

We can now solve for Δt:

$$\Delta t = \sqrt{\frac{2 Z^2 s_{err}}{fWa}} \qquad (25)$$

and substitute this value back into eq. (23):

$$v_{err} = Z \sqrt{\frac{2 a s_{err}}{fW}} + \frac{nZv}{fH}. \qquad (26)$$

Notes:

1. The optimal Δt gives a velocity error that is linear with range Z.

2. For zero acceleration the optimal Δt is infinity. In practice the system has been limited to $\Delta t = 2s$.

4 Experiments and Results

The system has been installed in a production vehicle equipped with radar based ACC. For these experiments the output from the vision system is sent to the controller in place of the output from the radar. The radar output is logged and used for *ground truth*. The Figure 3 shows a few frames from a typical sequence where the host vehicle approaches a slower vehicle and the vision based ACC decelerates to match the speed and keep a safe headway distance.

Figure 4 shows the range and range rate results. From the figure 4b we can see that the vehicle is decelerating relative to the target vehicle from a speed of $-6m/s$ to $0m/s$ during 60 frames (i.e. 6 secs) at a near constant rate of $1m/s^2$. This is a typical deceleration for an ACC system ($2m/s^2$ or 0.2G is the typical upper limit allowed in a *comfort system*).

Figure 5 shows the optimal value for Δt using eq. (25) assuming the relative acceleration of $a = -1m/s^2$ and the range to the vehicle. The values range from $0.65s$ for the vehicle at $57m$ to $0.27s$ when the vehicle is at $24m$. The truck width was taken to be $2m$. Using eq. (26) one can compute the theoretical error bounds on the relative velocity error. This is shown in figure 6 together with the actual velocity error. As one can see, the actual error lies mostly within the theoretical bounds.

5 Summary

Both range and range rate can be estimated from a single camera using the laws of perspective. This can be done because we are dealing with a constrained environment: the camera is at a known height from a near planar surface and the objects of interest (the other vehicles) lie on that plane.

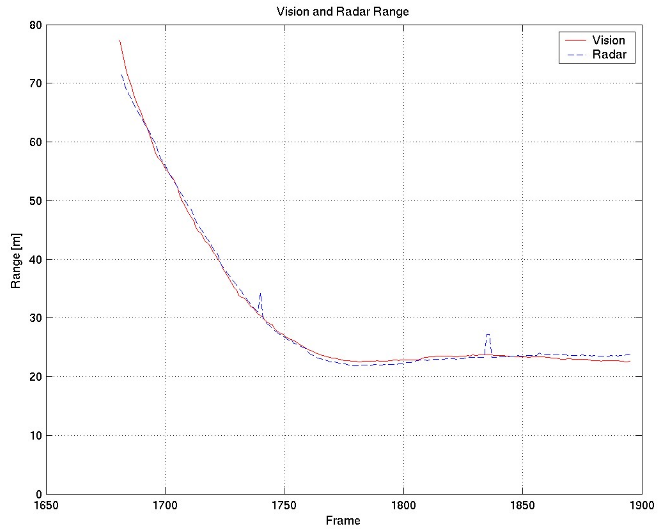
We have shown how various parameters affect the accuracy of the estimates. In particular we have discussed the effect of the length of time Δt over which we compute the scale change of the target vehicle. Using the optimum values produces errors in relative velocity which are small enough to allow for ACC control. The system has been implemented in a real-time embedded system and has be driven on highways on both flat and hilly terrains. Being able to perform ACC using a single camera opens the way to truly Low Cost ACC.

(a)

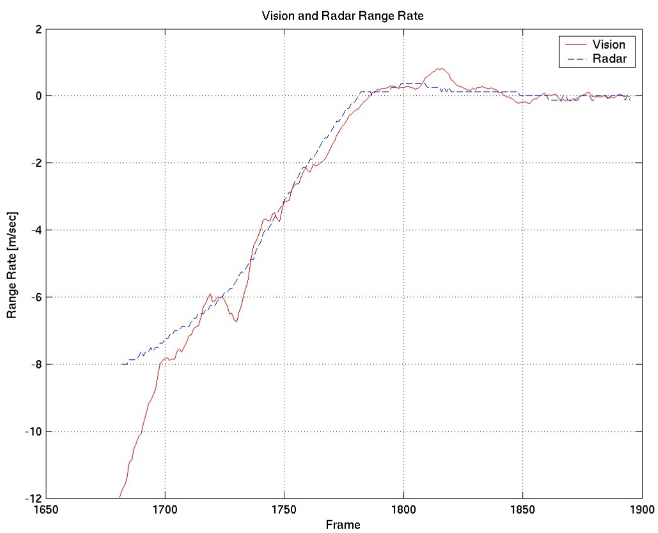
(b)

(c)

Figure 3: *A typical sequence where the host vehicle decelerates so as to keep a safe headway distance from the detected vehicle. The detected target vehicle (the truck) is marked by a white rectangle. As the distance to the target vehicle decreases the size of the target vehicle in the image increases.*



(a) Range



(b) Range Rate

Figure 4: *(a) Range and (b) Range Rate for a typical sequence where the host vehicle decelerates so as to keep a safe headway distance of 24m from the detected vehicle. The lead vehicle was traveling at 80KPH.*

References

[1] G. R. Widman, W. A. Bauson and S. W. Alland Development of collision avoidance systems at Delphi Automotive Systems, In *Proceedings of the Int. Conf. Intelligent Vehicles*, pages 353 -358, 1998

[2] K. Saneyoshi Drive assist system using stereo image recognition, In *Proceedings of the Intelligent Vehicles. Symposium*, pages 230-235, Tokyo, 1996.

[3] A. Broggi et al, Automatic Vehicle Guidance: The experience of the ARGO Autonomous Vehicle, World Scientific, 1999.

[4] T. Williamson and C. Thorpe Detection of small obstacles at long range using multibaseline stereo, In *Proceedings of the Int. Conf. Intelligent Vehicles*, pages 230-235, 1998

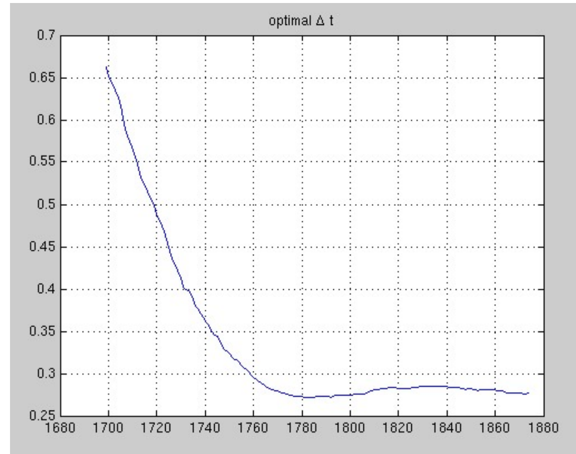


Figure 5: *Optimal Δt computed from eq. (25) assuming an acceleration of* $a = -1m/s^2$.

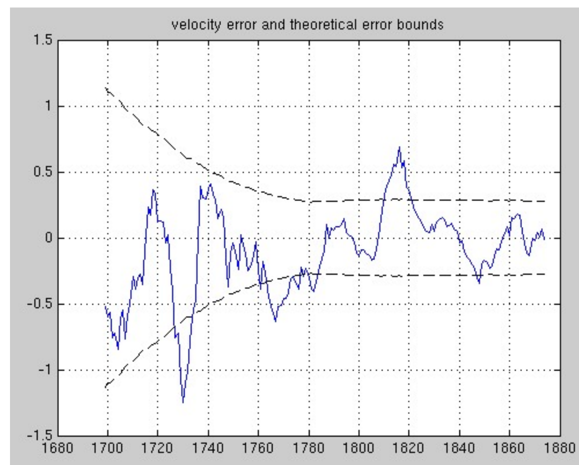


Figure 6: *The relative velocity error and error bounds (dashed lines) computed from eq. (26) assuming a relative acceleration of* $a = 1m/s^2$.

[5] G. Stein, O. Mano and A. Shashua. A robust method for computing vehicle ego-motion, In *IEEE Intelligent Vehicles Symposium (IV2000)*, Oct. 2000, Dearborn, MI

[6] C. Kreucher, S. Lakshmanan and K. Kluge A driver warning system based on the LOIS lane detection algorithm, In *IEEE Intelligent Vehicles Symposium (IV1998)*, Oct. 1998, Stuttgart

[7] M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using image stabilization, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 454–460, Seattle, Washington, June 1994.

[8] M. J. Black and P. Anandan. Robust dynamic motion estimation over time, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 296– 302, June 1991.