SENSOR SIGNAL PROCESSING FOR IVHS APPLICATIONS

W. Y. Kan, J. V. Krogmeier, and P. C. Doerschuk

School of Electrical Engineering, Purdue University
West Lafayette, IN 47907-1285 USA

ABSTRACT

Accurate and wide-area estimates of vehicle velocity and traffic spatial and temporal densities will be essential components of future algorithms for freeway and arterial street control, for incident prediction and detection, and for optimization in route selection. Algorithms like these figure prominently in the current research and development of Intelligent Vehicle-Highway Systems (IVHS). This paper presents an approach to a class of vehicle monitoring problems which is based upon a video backbone sensor and multiple target tracking (MTT). The method allows the integration of measurements made from other sensors like inductive loops, microwave radars, and laser range profilers.

1. INTRODUCTION

The Intelligent Vehicle-Highway System (IVHS) is envisioned to be an integrated group of technologies and services providing fundamental improvement in the efficiency and safety of the nation's surface transportation infrastructure. Central to the achievement of these goals is a traffic management system which will be required to process distributed traffic observations for the purposes of state estimation and prediction. The results of estimation and prediction algorithms are used for traveler information systems, traffic control (via ramp metering [1] and signal coordination), and incident response.

In this paper we describe our application of video image processing and multiple target tracking (MTT) to the problem of traffic monitoring. The basic idea is to use low frame-rate video to detect cars and then drive an MTT with the detections in order to estimate vehicle tracks. Video-based tracking for non-IVHS purposes is discussed in Refs. [2, 3, 4]. From the tracks it is straightforward to compute basic traffic parameters such as flow, speed, and concentration. If accurate tracks can be maintained over distance, then they can provide additional information; for example, origin-destination pairs of interest to traffic planners and transit time estimates which cannot be accurately derived from snapshots of position and velocity. In addition, tracks can provide information about lane change activity per unit time or distance which could be very useful in incident detection and prediction.

In the remainder of this paper we describe vehicle detection based on video (Section 2), modeling of vehicle dynamics and road geometry for MTT (Section 3), the MTT algorithm (Section 4), and experimental results on real highway video data (Section 5). Due to space limitations we provide greater detail for the detector than for the MTT.

This work was partially supported by a graduate fellowship from the Purdue Research Foundation.

2. VEHICLE DETECTION

One of the central goals of this work is efficient computational realization of the entire algorithm, including detection and tracking. Therefore, we restrict the detection problem to the individual lanes themselves (call these the lane images), we use only single frames for detection, and where feasible we avoid pixel or region based image processing. The primary goal of the detection work is the development of detection thresholds computed from estimated lane image statistics which will automatically adapt to changing ambient illumination. A secondary goal is good approximate expressions for vehicle detection probability and false alarm rate which may be used in the development of subsequent tracking algorithms.

Examination of histograms of pixel intensities in typical lane images suggests that a three component Gaussian mixture may be a reasonable statistical model for a lane including background and vehicles. We have obtained a good fit to such a model using a common variance for the three terms, a low mean to represent shadows under and alongside vehicles, a high mean to represent bright regions associated with vehicles, and an intermediate mean representing the road surface. The mixing proportions depend upon the number, size, and position of vehicles in the lane. In addition, the image sequences examined in this project have exhibited a considerable longitudinal and lateral variation in the apparent mean of background pixel intensity. These variations, due to the changing viewing angle in the longitudinal direction and to surface irregularities in the lateral direction, are modeled by assuming that the pixels associated with the road surface in the lane are independent Gaussian random variables with a common variance and spatially dependent mean. To remove the effect of the spatially varying mean we form a pixel by pixel estimate of the lane background. The estimate is allowed to slowly forget the past in order to track temporal variations in ambient illumination and detections are fed back to remove vehicles from the estimate.

2.1. The Per Pixel Detection Model

The background-subtracted lane image at pixel location ij and frame t is denoted $\phi_{ij}(t)$ and is modeled as a three component Gaussian mixture. The central peak in the mixture

0-7803-2431-5/95 $4.00

is centered at zero and represents the background and the two other peaks (typically much smaller than the central one) are due to vehicles. The component of positive mean corresponds to bright points in a vehicle and the component of negative mean corresponds to shadows.

These considerations lead to the following simple classification of individual pixels in the (background subtracted) lane image:

$$\begin{array}{lll} H_0: & X \sim \mathcal{N}(0, \sigma^2) & \text{[background]} \\ H_b: & X \sim \mathcal{N}(\mu_b, \sigma^2) & \text{[bright]} \\ H_d: & X \sim \mathcal{N}(\mu_d, \sigma^2) & \text{[dark]} \end{array}$$

where X denotes a generic pixel (i.e., one of the $\phi_{ij}(t)$). If the relative proportions of bright and dark pixels were known it would be possible to formulate a two class test for H_0 versus a combination pixel distribution representing a vehicle. The difficulty in this approach is that these proportions are unknown and will typically vary from frame to frame as vehicles leave and enter the field of view or change in apparent size.

Instead, the approach taken here is to separately test H_0 versus H_b and H_0 versus H_d via Neyman–Pearson detectors and to combine the results so as to avoid using knowledge of the mixing proportions. Separately, these problems amount to Gaussian location testing for which the Neyman–Pearson detector is easy to find. This leads to a pixel by pixel classification via a two sided thresholding of the value of each pixel in the image $\phi_{ij}(t)$. If the value fell between the two thresholds the pixel would be declared background and otherwise vehicle. Such a method would obviously require some spatial filtering because the performance of the pixel by pixel classification would not typically be acceptable. An alternate approach would be to incorporate prior knowledge of the shape of vehicles into the model. However, even retaining the assumption of independent pixels, such would entail a composite testing problem with three degrees of freedom (location and size of vehicle) and the computation requirements of the resulting search would be prohibitive. In the next section, we make a further refinement of the vehicle model which concentrates only on lines of the lane image to simplify considerably the computational effort involved.

2.2. The Per Line Detection Model

To simplify the complexity of the proposed vehicle detector we have modified the model slightly to concentrate on individual lines in the lane image. Lines will be placed into one of three classes: $\bar{H}_0$ (line intersects background), $\bar{H}_b$ (line intersects bright part of vehicle), or $\bar{H}_d$ (line intersects dark part of vehicle). Classification will be done by first testing $\bar{H}_0$ versus $\bar{H}_b$. If the bright hypothesis is selected the line is declared to have intersected a vehicle. Otherwise, we test $\bar{H}_0$ versus $\bar{H}_d$. If the dark hypothesis is selected the line is then declared to have intersected a vehicle. If both the bright and the dark hypotheses fail, the line is declared to be a part of the background. For the description to follow we will concentrate solely on the $\bar{H}_0$ versus $\bar{H}_b$ test. Under all of the above hypotheses, the pixels in the selected line are modeled as independent. Fix a frame number t and a particular line i and suppose that the lane is W pixels wide

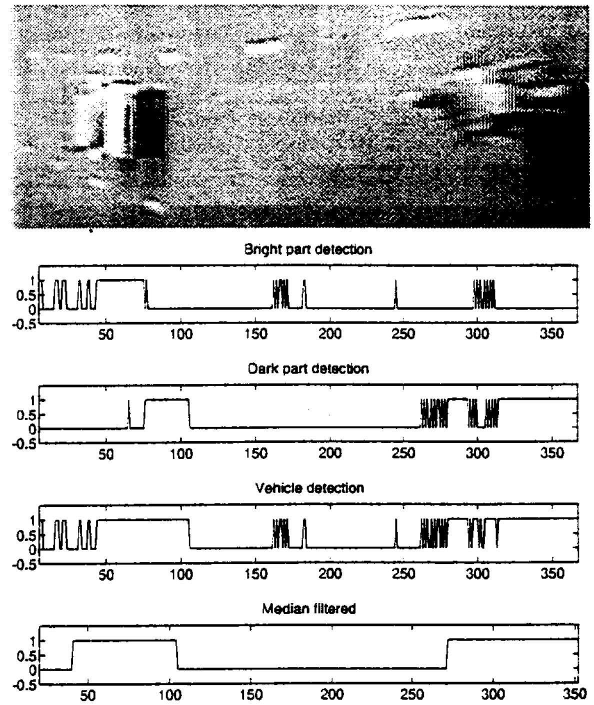


Figure 1: The Detector: Intermediate and Final Results.

at line i. The random variables observed in this line of the image are $\{\phi_{ij}(t) : 1 \le j \le W\}$ but they will be denoted $\{X_j : 1 \le j \le W\}$ for simplicity. Then the density for hypothesis $\bar{H}_0$ is

$$\bar{p}_0(\underline{x}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^W \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{W} x_k^2\right\}$$

Hypothesis $\bar{H}_b$ is composite in that both the vehicle width V and position k are unknown. Thus it consists of the family of densities

$$\bar{p}_k^V(\underline{x}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^W \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{W}(x_j - \mu_b w_k^V(j))^2\right\}$$

where $w_k^V(j) = 1$ if $k \le j \le k + V - 1$ and $= 0$ otherwise, k ranges from 1 to $W - V$, and V ranges over allowable vehicle widths.

A generalized likelihood ratio test for this problem uses a test statistic of the form:

$$\max_{V,k}\left\{2\sum_{j=1}^{W} w_k^V(j)X_j - \mu_b V\right\}. \tag{1}$$

This test is unacceptable for two reasons. First, it requires that μ_b be known in order to set the threshold. However, the means of the two vehicle distributions (bright and dark) may be difficult to estimate given the relative paucity of vehicle pixels in comparison to background pixels. Second, the distribution of the statistic in Equation (1) will be difficult to calculate or approximate directly.

Hence, we further approximate by using a fixed and typical value for V which we assume is an integer fraction of the lane width, i.e., we take $W/V = N$. Further, we restrict the possible vehicle locations in the lane line to be of the form $k = (n-1)V+1$ where $1 \leq n \leq N$. If we define

$$Y_n = \sum_{j=(n-1)V+1}^{nV} X_j$$

then we may use the follow approximate statistic in place of (1)

$$Z_b = \max_{1 \leq n \leq N} Y_n. \tag{2}$$

This removes the dependence upon the bright mean μ_b and allows the computation of the distribution of the test statistic since the variables Y_n are independent under either hypothesis.

Under $\bar{H}_0$ the Y_n are iid and distributed individually as $\mathcal{N}(0, V\sigma^2)$. It is easy to show then that the distribution[1] of Z_b under $\bar{H}_0$ is $\Phi^N(z_b/\sqrt{V}\sigma)$. Hence, for a specified false alarm rate $\bar{\alpha}_b$ the Neyman–Pearson test for $\bar{H}_0$ versus $\bar{H}_b$ is

$$Z_b \begin{array}{c} \geq \\ \\ < \end{array} \tau_b = \sqrt{V}\sigma\Phi^{-1}((1-\bar{\alpha}_b)^{1/N}) \begin{array}{c} \text{decide } \bar{H}_b. \\ \\ \text{decide } \bar{H}_0. \end{array} \tag{3}$$

The approximation made in replacing statistic (1) with (2) amounts to replacing the original composite hypothesis $\bar{H}_b$ with a family of distributions for the Y_n (one for each $1 \leq m \leq N$), specifically, $Y_m \sim \mathcal{N}(V\mu_b, V\sigma^2)$, $Y_n \sim \mathcal{N}(0, V\sigma^2)$ for $n \neq m$. Since Z_b is the maximum of the Y_n we easily see that its distribution will not depend upon the particular incarnation of the approximate $\bar{H}_b$ hypothesis above. Arguing along these lines it is easy to derive the approximate performance of the detector (3)

$$\bar{P}_b\{Z_b > \tau_b\} = 1 - \Phi\left(\frac{\tau_b - V\mu_b}{\sqrt{V}\sigma}\right)\Phi^{N-1}\left(\frac{\tau_b}{\sqrt{V}\sigma}\right)$$

where τ_b is given in (3). Since the minimum of the Y_l's is the maximum of the $-Y_l$'s, the test of $\bar{H}_0$ versus $\bar{H}_d$ is similar.

The combined line hypothesis test can be summarized as follows. If the minimum Z_d is less than τ_d or the maximum Z_b is greater than τ_b declare that the line includes vehicle pixels. Otherwise, declare the line to be part of the background. The final step of the detection process is to combine the line-by-line results into vehicle detections. Recall that the result of a line test is 0 (background) or 1 (vehicle, either bright or dark). As shown in Figure 1, the line-by-line results have both false positive and missed detection errors. However, after processing by a 1-dimensional 21-point median filter, the errors are removed.

[1] The cumulative distribution function of Gaussian random variable X with mean zero and variance one is denoted $\Phi(x)$.

3. DYNAMIC MODEL AND OBSERVATION GEOMETRY

3.1. Vehicle Dynamics

In most multiple target tracking problems where computational complexity is an issue, target dynamics are handled with relatively simple (two or three state) models. The motivation for this is twofold. First, often the measurement to track association problem is quite difficult and requires considerable computational resource. Second, simple dynamic models frequently suffice particularly when good coordinate systems have been chosen and/or it is inadvisable to overmodel imprecisely known dynamics.

Our approach is to model the dynamics of individual vehicles separately, i.e., there is assumed to be no coupling between vehicles. Furthermore, each vehicle is assumed to remain in its lane requiring that lane changes be handled by a supervisory track–observation association function. Within a lane we assume that the vehicle follows a known path and that its dynamics along the path are given by the standard constant velocity model which, when sampled, yields the discrete time model

$$\begin{bmatrix} x_{k+1} \\ \dot{x}_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ \dot{x}_k \end{bmatrix} + w_k$$

where T is the observation sampling interval and the process noise w_k is assumed to be zero mean, white, with covariance matrix

$$Q = \mathrm{E}\{w_k w_k^T\} = \begin{bmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{bmatrix} q, \quad q > 0.$$

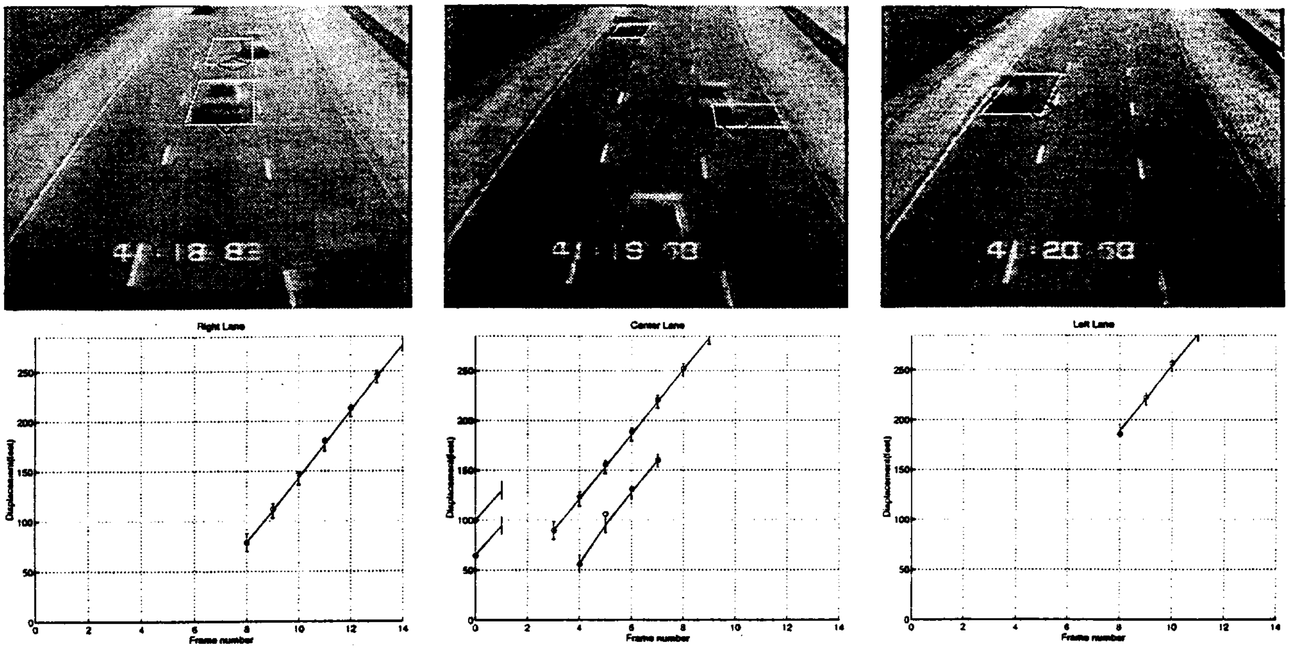
More detailed dynamic models based on human factors are available [5] which include interactions between vehicles in one lane and the work described here could, at the cost of additional computational complexity, be extended to use such models.

3.2. Road–Camera Geometry

The fundamental input to the MTT is the detection of a vehicle and its location. The image coordinates $\vec{x}^I \in \mathcal{R}^2$ of a location $\vec{x} \in \mathcal{R}^3$ within the field of view of the camera are $\vec{x}^I = h(\vec{x})$ where h contains a linear rotation and translation and a nonlinear projection. Given that the vehicle is constrained to lie on the road, h can be inverted to compute road location $\vec{x}$ from image location $\vec{x}^I$. In the MTT we use road coordinates in the tracking filters and an observation that is h^{-1} applied to the return provided by the detection algorithm. This results in an observation equation with a state-dependent observation noise covariance: $z_k = Hx_k + \epsilon_k$ where $H = [1\ 0]$, $\mathrm{Cov}[\epsilon_k] = R_k(x_k)$, and $z_k = h^{-1}(\vec{x}_k^I)$.

4. MULTIPLE TARGET TRACKING

The fundamental issue in MTT is data association—which return is due to which target, including the possibility that some returns do not correspond to any target (clutter) and that the return from some targets at some times may be missed. Our approach has 2 steps. First, we select out a set



(a) (b) (c)

Figure 2: Tracking Results: Three Selected Images and Per-Lane Space-Time Trajectories for the (a) Right, (b) Center, and (c) Left Lanes.

of candidate returns based on a gate determined from the standard deviation of the predicted location provided by the dynamic system model. Second, since most of the clutter in this problem is due to end-of-vehicle effects, we use a nearest neighbor filter which selects the return closest to the predicted location. This is equivalent to a Generalized Likelihood Ratio test in which the parameter is the identity of the return.

We deal with lane changes in a post-processing stage. After performing per-lane data association for each lane, those returns which have not been associated with any track are determined. For this set of returns, a second round of gating and association are performed where the predicted locations are generated by tracks which, at the previous time-step, were in adjacent lanes. If the association is successful, then the vehicle is declared to have changed lanes.

In addition to performing track continuation, an MTT must also perform track initialization, confirmation, and deletion. We use simple algorithms: A track is initialized whenever an return cannot be associated with any existing track. A track is confirmed if an observation is associated with it at the sample time following its initialization. A track is deleted if for 2 consecutive sample times there has been no return associated with the track.

5. RESULTS

Figure 2 shows 3 frames from a video sequence. Rectangular (diamond) boxs mark the detected (filtered) position. The height of the diamond corresponds to the standard deviation of the position estimate from the Kalman filter. Below the 3 images are 3 space-time plots, one plot for each lane. Tracks and detections are represented by solid lines and circles respectively. Error bars correspond to the standard deviation of the tracks' positions. Notice the one vehicle which changes lanes so that it's track terminates in the center-lane space-time plot and continues in the left-lane plot.

6. REFERENCES

[1] D. P. Looze, P. K. Houpt, Jr. N. R. Sandall, and M. Athens. On decentralized estimation and control with application to freeway ramp metering. *IEEE Trans. Automat. Contr.*, AC-23(2):268–275, April 1978.

[2] P. Bouthemy and P. Lalande. Recovery of moving object masks in an image sequence using local spatiotemporal contextual information. *Opt. Eng.*, 32(6):1205–1211, June 1993.

[3] A. L. Gilbert, M. K. Giles, G. M. Flachs, R. B. Rogers, and Yee Hsun U. A real-time video tracking system. *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-2(1):47–56, January 1980.

[4] D. D. Sworder and R. G. Hutchins. Image-enhanced tracking. *IEEE Trans. Aero. Elec. Sys.*, AES-25(5):701–709, September 1989.

[5] R. E. Chandler, R. Herman, and E. W. Montroll. Traffic dynamics: Studies in car following. *Operations Research*, pages 165–184, March 1958.