

# Motion Based Vehicle Identification in Car Video

A. Jazayeri, H. Cai, J. Y. Zheng, *Senior member, IEEE* and M. Tuceryan, *Senior member, IEEE*

**Abstract**—This work aims at detecting and tracking vehicles in in-car video. Rather than enhancing shape analysis of various vehicle types and road situations, this work focuses on vehicle and background motions because they are more general than shapes and colors of cars in various road environments. Basic features are tracked stably using corners, intensity peaks, and horizontal line segments. We use the HMM in the temporal domain to separate background and moving vehicles in the video. To realize this, we model the image motion of vehicles and background probabilistically according to the scene characteristic and vehicle driving mechanism, as well as the joint distribution of horizontal position and velocity of scenes. The identification and tracking are robust to various illumination and environments and the processing is performed in real time. The identification results based on motion only is good and a better result can be achieved further by fusing the motion result with the results from shape analysis.

## I. INTRODUCTION

SENSING traffic during driving is an important aspect for safety driving, accident avoidance, and automatic car pursuit (Fig.1). We implement this function with a vehicle borne camcorder. The goal further includes localizing target cars continuously in front of the observer car. The fundamental problem here is to identify and track dynamic vehicles in a constantly changing environment (street and side scenes) and illumination (due to shadow, day and night, weather, etc.). Although there have been numerous works on object recognition and tracking, and the combination of them, not many could be successfully applied to the in-car video robustly in real time.

The objective of this work is to detect vehicles ahead from a moving camera and track them continuously in the video, which is not easy without stereo or other sensors' assistance [25]. The main difficulty is the variation of vehicles in color, shape, and type, which is hard to model or learn. Although there have been many works on tracking, most of them assume detectable targets or known initial positions [11]. Moreover, the in-car video taken on roads may confront drastic change of environment and illumination [9]. For example, quick transition through shadow and sunny locations in urban areas, dim lighting at night, loss of color on a cloudy day, shining highlight on vehicle bodies, large scale changes due to varied depth, occlusions between vehicles and background, and so on all make the feature extraction, tracking, and recognition unstable.

Our novel method first selects and detects the most common low-level features on vehicles that are robust to changes of illumination, shape, and occlusion. This avoids high-level scene analysis and learning. Second, we focus on the horizontal scene movement for fast processing based on the camera configuration and vehicle driving mechanism. One dimensional profiles are condensed from video frames to show the horizontal motion directly in traces. We track feature trajectories in such a domain of reduced dimensionality. Third, we model the motion behavior of the vehicles and background in probability and identify targets using the Hidden Markov Model (HMM) [24,26]. This generates probability continuously following the feature movement and yields robust vehicle detection.

Related works in identifying cars [1,2,3,5,6,9,23,27] are mostly shape based methods that usually suffer from brittleness of variety of cars and backgrounds. Several works have employed probability [12,20,21,22], but they are mainly applied to the appearance and occurrence of vehicles or for static cameras. In contrast, the significance of our detection-via-tracking approach lies in exploring the temporal behavior of scenes rather than their shapes. The success of this method will extend the processing to long video sequences and facilitate the processing of many vehicle related tasks. The modeling of motion behavior with the HMM brings the vehicle detection with certainty.

In the following, we will describe the motion model of camera/vehicles, targets and background in Sect. 2. And address feature detection, selection, and tracking in Sect. 3. Probability based identification of cars by the HMM is given in Sect. 4. Experimental results are given in Sect. 5.

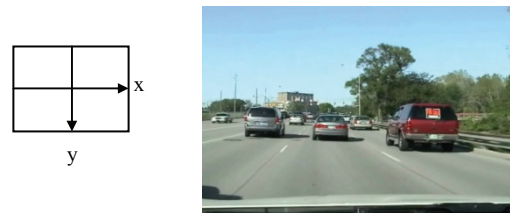


Fig. 1. A frame of in-car video in right hand coordinate system.

## II. PROBABILITY MODEL OF VEHICLE MOTION

### A. Motion Model of Vehicle Borne Camera

For a vehicle-borne camera, a rotation without significant translation, e.g., turning at a street corner, does not provide sufficient information for separating targets and background motion, as it generates almost the same optical flow in entire view. Therefore, a translation of the observer vehicle is required in our assumption. Another property of vehicle is continuity in motion guaranteed by the mechanism of four-

wheeled vehicles.

Denote the camera system  $O-XYZ$  with the  $X$  axis toward right, the  $Y$  axis downward and the  $Z$  axis forward in the vehicle moving direction, respectively. A 3D point  $(X,Y,Z)$  in  $O-XYZ$  is projected to image  $I(x,y)$ . An FOE (focus of expansion) is located in the center part of  $I(x,y)$  during translation, usually overlapped with the vanishing point of the road. On a curved road, a non-zero steering angle causes rotation of the camera. On the other hand, a target car moves on the same road may change its horizontal position, image velocity  $\mathbf{u}(x,y)$  and scale drastically when it changes lane and speed. Background points, however, have motion depending on their 3D distances and the motion of the observer vehicle. They cause the flow spreading out from FOE gradually towards the sides of frame. The image velocity increases as the scene is close to the camera. These cues are sufficient for humans to classify vehicles and background even without knowing the shape of objects. We will model the motion coherence of scenes for background and vehicle separation.

According to the perspective projection of the camera, the image position of an object point is

$$x(t) = \frac{fX(t)}{Z(t)} \quad y(t) = \frac{fY(t)}{Z(t)} \quad (1)$$

where  $f$  is the camera focal length. Denote its relative translation to the camera by  $(T_x(t), T_y(t), T_z(t))$  in  $O-XYZ$ , and the rotation of the observer vehicle around the  $Y$  axis by  $R_y(t)$ . The pitch and roll of the observer vehicle are  $R_x(t) \approx 0$  and  $R_z(t) \approx 0$  on a flat road. Then, the relative speed of the point to the camera,  $(V_x(t), V_y(t), V_z(t))$ , is then [29]

$$(V_x(t), V_y(t), V_z(t)) = (T_x(t), T_y(t), T_z(t)) + (X, Y, Z) \times (0, R_y(t), 0) \quad (2)$$

By differentiating Eq. 1 with respect to  $t$ , and replacing terms in the result using Eq. 1, the horizontal image velocity at the point becomes

$$v(t) = \frac{\partial x(t)}{\partial t} = \frac{fV_x(t) - x(t)V_z(t)}{Z(t)} \quad (3)$$

Replacing  $V_x(t)$  and  $V_z(t)$  with Eq. 2, we get

$$v(t) = \frac{fT_x(t) - x(t)T_z(t)}{Z(t)} - \frac{x^2(t) + f^2}{f} R_y(t) = vt(t) + vr(t) \quad (4)$$

where  $vt$  and  $vr$  summarize components of horizontal image velocity components from translation and rotation,

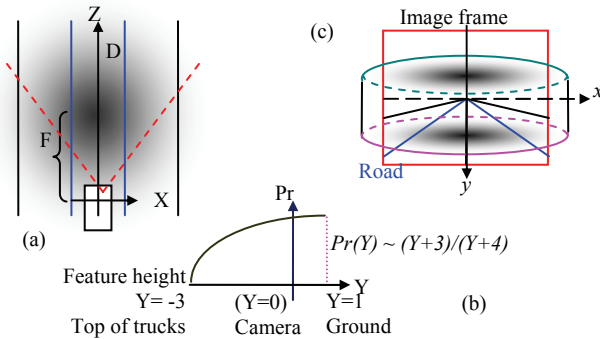


Fig. 2. Probability density of target vehicle positions shown as the intensity. The darker the intensity, the higher the probability is. (a) Target vehicle distribution in top view, (b) Vehicle feature distribution in height, which is from the ground to the top of highest vehicles, (c) Projection of the vehicle feature distribution onto the image frame.

respectively. If the observer vehicle moves straight,  $R_y(t)=0$ ,  $V_x(t)=T_x(t)$ , and  $V_z(t)=T_z(t)$ .

### B. Occurrence of Target Vehicles in Camera Frame

We examine vehicle position  $(X(t), Z(t))$ , as shown in Fig. 2, appearing with a 2D normal distribution  $G((0, F), (D^2, (2F)^2))$ . Here  $F$  is the average distance of targets from the camera.  $X(t)$  and  $Z(t)$  are independent.  $D$  (related to road width with multiple lanes) and  $|2F|$  are used as the standard deviations. The projection of the distribution onto the image frame is depicted in Fig. 2c determined from a probability distribution,  $H(Y)$ , of vehicle features in height.

In  $Y$  direction, the vehicle features are easily detected near the ground due to the uniformity in vehicle shadow, tire, and bumper against homogeneous road surfaces. However, features may not be detected reliably at higher positions due to reasons such as a low vehicle height, highlights on a metal top, and mirror reflection on the back window, etc. As depicted in Fig. 2b, we design a vehicle feature distribution function  $H(Y) \sim (Y+3)/(Y+4)$ , where  $Y \in [-3m, 1m]$ , with the camera position ( $Y=0$ ) at about 1m above the ground. Moreover, we can set the relative speed of a target vehicle to the camera,  $(T_x, T_z)$ , as normal distribution  $G((0, 0), (\sigma_x^2, \sigma_z^2))$  during a stable pursuit, where  $T_x$  and  $T_z$  are independent.

Based on above definition, the probability of image position of a target car,  $p(x, y|C)$ , is computed. Since the mapping from  $X, Y, Z$  to  $x, y$  is not one-to-one relation, we use Bayes theorem and dependent variable [10] to have

$$\begin{aligned} p(x, y|C) &= \int_Z p(Z) \times p(x, y|Z) dZ \\ &= \int_Z p(Z) \times p\left(X = \frac{xZ}{f}, Y = \frac{yZ}{f} \mid Z\right) \times \left|\frac{Z}{f}\right|^2 dZ \\ &= \int_Z p(Z) \times p\left(X = \frac{xZ}{f} \mid Z\right) \times p\left(Y = \frac{yZ}{f} \mid Z\right) \times \left|\frac{Z}{f}\right|^2 dZ \\ &= C_1 \int_Z e^{-\frac{(Z-F)^2}{2(2F)^2}} \times e^{-\frac{\left(\frac{xZ}{f}\right)^2}{2D^2}} \times \frac{yZ+3f}{yZ+4f} \times \left|\frac{Z}{f}\right|^2 dZ \end{aligned} \quad (5)$$

where  $C_1$  is a constant for normalization  $\int_y \int_x p(x, y|C) dx dy = 1$ . A result is shown in Fig. 3, which is consistent to the vehicle occurrence in Fig. 1.



Fig. 3. The probability of vehicle features  $p(x, y|C)$  in the image frame displayed in gray level.

### C. Motion Model of Target Vehicles in Car Video

The motion properties of scenes provide the key evidence in our vehicle identification. Determined by vehicle speed and scene distance, the motion properties are robust and general as compared to car shapes and colors that

are influenced from illumination, specular reflection, occlusion, etc. Our motion model of a feature uses its horizontal image position  $x(t)$  and horizontal image velocity  $v(t)$ , denoted by  $(x,v)$ , from its continuous trajectory.  $(x,v)$  are not independent, neither one can determine the object identity alone. We deal with two events,  $C$  and  $B$ , for car and background respectively. By identifying these events, we can alarm the driver to avoid obstacles in the background and follow the target cars at a proper speed.

For a target car, we compute its image and motion behavior  $p(x,v|C)$ . According to Eqs. 1, and 3, the probability of a point on a target vehicle to have motion  $(x,v)$  is

$$p(x,v|C) = p(x,v|(X,Z,T_x,T_z) \in C) = p\left(x = \frac{fX}{Z}, v = \frac{fT_x - xT_z}{Z}\right) \quad (6)$$

$$\begin{aligned} &= \int_Z p(Z) \times p\left(x = \frac{fX}{Z}, v = \frac{fT_x - xT_z}{Z} \mid Z\right) dZ \\ &= \int_Z p(Z) \times p\left(X = \frac{xZ}{f}, T_x, T_z = \frac{Zv + xT_z}{f} \mid Z\right) \left|\frac{Z}{f}\right|^2 dZ \\ &= \int_Z p(Z) \times p\left(X = \frac{xZ}{f} \mid Z\right) \times p\left(T_x, T_z = \frac{Zv + xT_z}{f} \mid Z\right) \left|\frac{Z}{f}\right|^2 dZ \\ &= \int_Z p(Z) p\left(X = \frac{xZ}{f} \mid Z\right) \left\{ \int_{T_z} p(T_x) p\left(T_x = \frac{Zv + xT_z}{f} \mid T_z, Z\right) \left|\frac{Z}{f}\right|^2 dT_x \right\} dZ \end{aligned}$$

using Bayesian and Jacobi [10]. Filling in probability distributions of  $p(T_x)$ ,  $p(Z)$ , and  $p(T_z)$ , we obtained

$$p(x,v|C) = C_2 \iint_{T_z} e^{-\frac{(Z-F)^2}{2(2F)^2}} e^{-\frac{\left(\frac{xZ}{f}\right)^2}{2D^2}} e^{-\frac{T_z^2}{2\sigma_z^2}} e^{-\frac{\left(\frac{Zv+xT_z}{f}\right)^2}{2\sigma_x^2}} \left|\frac{Z}{f}\right|^2 dT_z dZ \quad (7)$$

where  $C_2$  is a constant for normalization. Figure 4 shows the computed motion probability distribution of vehicles.

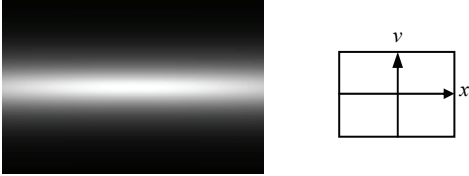


Fig. 4. Probability distribution  $p(x,v)$  of the vehicle motion appearing in continuous space  $(x,v)$ . The center has value  $(0,0)$ .

#### D. Background Motion Model in Car Video

Background motion in the video relies on the motion of observer vehicle and its distance to the background. We describe background distribution in  $O-XYZ$  as depicted in Fig. 5 for the observer vehicle on road. The background features may uniformly distribute in height to include high buildings and trees. During the translation of the observer vehicle ( $R_x \approx 0$ ), the absolute speed  $V$  follows a normal distribution, i.e.,  $p(V) \sim G(S, \sigma^2)$  for  $V > 0$ , where  $S$  can be set at a proper value (e.g., 50km/hr) for pursuing targets, and otherwise  $p(V) = 0$ . A 3D point on the background then has  $T_x = 0$  and  $T_z = -V$  approaching to the camera. Its image velocity becomes

$$v(t) = \frac{fXV}{Z^2(t)} = \frac{Vx^2(t)}{fX} \quad \text{for } V > 0 \quad (8)$$

according to in Eq. 3. For background scenes with various  $X$ ,

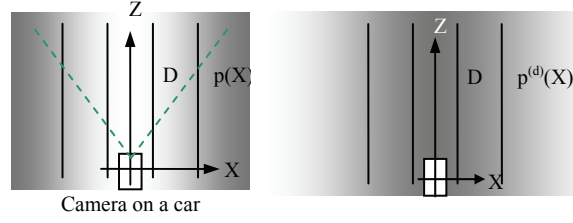


Fig. 5. Probability distribution of background beside a road and its sensibility/visibility by camera displayed in intensity. The higher the probability, the darker the intensity is. (a) Background feature distribution on roadsides, (b) Visibility of features on street due to the occlusion of close scenes over distant ones.

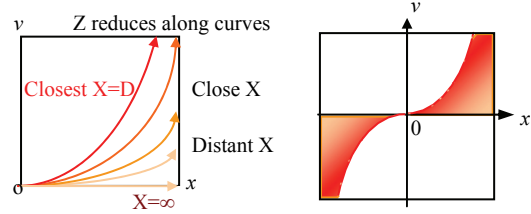


Fig. 6. Relation of the horizontal image velocity and the image position on background scenes. (Left) Motion traces of background points on right side of road. As depth  $Z$  reduces, background moves fast in outward direction. (Right) Traces for background on both sides of the road. The colors correspond to different  $X$  from the camera axis (or simply road center).

they draws dense curves in  $(x(t), v(t))$  space as shown in Fig. 6a. If the observer vehicle has roughly a constant speed, the background point trace  $x(t) = fX/(Z_0 - Vt)$  is a hyperbola from an initial depth  $Z_0$  according to Eq. 1. In Fig. 6a, an object at a far distance from road (large  $|X|$ ) has a flat trajectory, while an object close to road has a strongly curved trajectory.

To describe the background scene along the  $X$  direction, we use a flipped Gaussian distribution, i.e.,  $p(X) \sim 1 - \exp(-X^2/2D^2)$ , as in Fig. 5a. Now, let us compute the motion probability  $p(x,v|B)$  from background point  $(X,Z) \in B$ . Here the variables are  $X, Z$ , and  $V$  in the 3D space and their functions are image parameters  $x$  and  $v$ , as described by Eqs. 1~4. Because the inverse mapping from  $x, v$  to  $X, Z, V$  is not unique, we use conditional probability to include all the possible values of  $X, Z$ , and  $V$ . For given  $v$  and  $x$ , we loop variable  $X$  to determine  $Z$  and  $V$  in the 3D space for  $p(x,v|B)$ .

$$\begin{aligned} p(x,v|B) &= p(x,v|(X,Z) \in B) \\ &= \int_X p(X) p(x,v|X, (X,Z) \in B) dX \\ &= \int_X p(X) p(Z(x,v), V(x,v)|X) \left|\frac{fX}{x^2}\right|^2 dX \\ &= \int_X p(X) p(Z(x,v)|X) p(V(x,v)|X) \left|\frac{fX}{x^2}\right|^2 dX \\ &= \int_X p(X) \times p\left(Z = \frac{fX}{x} \mid X\right) \times p\left(V = \frac{v fX}{x^2} \mid X\right) \times \left|\frac{fX}{x^2}\right|^2 dX \quad (9) \end{aligned}$$

Input original probability distribution of  $Z$  and  $V$ , the background probability  $p(x,v)$  becomes

$$p(x,v|B) = C_3 \int_X \left(1 - e^{-\frac{X^2}{2D^2}}\right) \times e^{-\frac{\left(\frac{v fX}{x^2} - S\right)^2}{2\sigma^2}} \times \left|\frac{fX}{x^2}\right|^2 dX \quad (10)$$

where  $C_3$  is a constant and  $|\cdot|$  is from a Jacobian [10]. Because the background distribution  $p(Z)$  is uniformed for  $Z$  according to Fig. 5a,  $p(Z)$  is treated as a constant included in  $C_3$  in Eq.10.

In real situations, we further consider the horizontal background scene visibility defined by  $p^{(d)}(X)$ , as depicted in Fig. 6b, as

$$p^{(d)}(X) \propto 1/(|X|+1) \quad (11)$$

where objects on roadsides have the highest visibility and scenes far off the road (large  $|X|$ ) have more chances to be occluded. The PDF for background is then computed by

$$\begin{aligned} p(x, v | B) &= \int_x p^{(d)}(X) p(X) p\left(Z = \frac{fX}{x} | X\right) p\left(V = \frac{vfX}{x^2} | X\right) \left|\frac{fX}{x^2}\right|^2 dX \\ &= C_4 \int_x \frac{1 - e^{-\frac{x^2}{2D^2}}}{|X|+1} \times e^{-\frac{(\frac{vfX}{x^2} - s)^2}{2\sigma^2}} \times \left|\frac{fX}{x^2}\right|^2 dX \end{aligned} \quad (12)$$

where  $C_4$  is a constant for normalization. Figure 7a shows a result of background distribution.

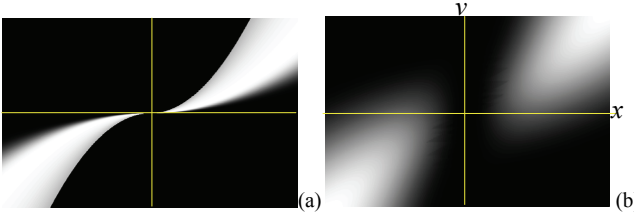


Fig. 7 Probability distribution  $p(x, v)$  of background in motion. The intensity corresponding to probability is scaled for display. (a)  $p(x, v)$  of background during camera translation. (b) Motion probability distribution for background when the observer vehicle has rotation velocity in Gaussian distribution  $G(0, 5^\circ)$ .

During smooth driving, the observer vehicle may have a small steering change along a straight or mildly curved path. We thus describe rotation velocity in a normal distribution with a small variance, e.g.,  $R_y(t) \sim G(0, 5^\circ)$ . According to Eq. 4, we estimate a general  $p(x, v | B)$  by adding rotation component  $vr(t)$  to the image velocity  $vt(t)$  from translation. This will vertically shift  $p(x, v)$  from  $vt(t)$  (Fig. 7a) in a term  $vr(t)$ . If the rotation parameter  $R_y(t)$  is not provided from the encoder of the observer vehicle, we have to include all its possible values for the probability distribution, i.e.,

$$\begin{aligned} p(x, v | B) &= \int_{R_y} p(R_y) \times p(x, v | R_y) dR_y \quad (13) \\ &= \int_{R_y} p(R_y) \times p(x, v | v = \frac{x^2 V}{fX} - \frac{x^2 + f^2}{f} R_y) dR_y \\ &= \iint_{R_y, X} p(R_y) p^{(d)}(X) p(X) p\left(Z = \frac{fX}{x} | X\right) p\left(V = \left(v + \frac{x^2 + f^2}{f} R_y\right) \frac{fX}{x^2} | X\right) dXdR_y \\ &= C_{4r} \int_{R_y} \int_{X} e^{-\frac{R_y^2}{2\sigma_r^2}} \times \frac{1 - e^{-\frac{x^2}{2D^2}}}{|X|+1} \times e^{-\frac{\left(\left(v + \frac{x^2 + f^2}{f} R_y\right) \frac{fX}{x^2} - s\right)^2}{2\sigma^2}} \times \left|\frac{fX}{x^2}\right|^2 dX dR_y \end{aligned}$$

where  $C_{4r}$  is a constant in normalization of  $p(x, v | B)$ . The result is obtained in Figure 7b, which is basically a blur of the PDF in Fig. 7a vertically. By comparing Figures 4 and 7, we can realize the major differences between motion behaviors of target vehicles and background.

### III. ROBUST MOTION DETECTION IN CAR VIDEO

#### A. Feature Extraction of Target Vehicles

The segmentation of vehicles from background is difficult due to the complex nature of the scenes such as constant occlusion over background or by other vehicles, variations in shapes and textures. We select several types of low-level features such as corners, intensity peaks, line segments, and intensity. We have noticed that vehicle shapes typically contain many horizontal edges formed by car tops, windows, bumpers, shadows, etc. Most of these structural lines are visible during daylight and even in nighttime. Vertical lines, however, are not assured in detection due to a curved vehicle body, and frequent occlusion by other cars or over changing background during its motion.

To extract the horizontal segments, we convolve each video frame with vertical differential operator,  $\partial I(x, y) / \partial y$ , resulting differential image  $I'_y(x, y)$ . Then, an edge following algorithm searches peaks in  $I'_y(x, y)$  for horizontal segments with contrasts above threshold  $\delta_l$ . The horizontal search of edge points has a vertical tolerance of  $\pm 2$  pixels and selects point candidates with the same sign of edge. It also uses another threshold  $\delta_2 < \delta_l$  to bridge segments at weak-contrast points, which allows a 3-pixel trial after reaching an end point using  $\delta_l$ . The tracked edge points form a line segment if they satisfy constraints on the minimum and maximum lengths, high contrast, and near horizontal orientation.

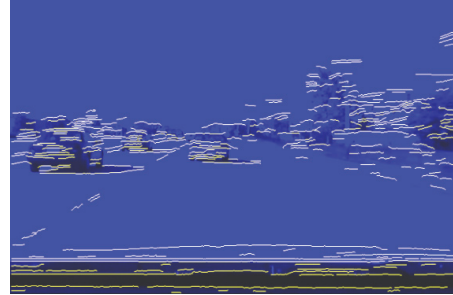


Fig. 8. Tracked edge points form horizontal line segments in video frames that can characterize vehicles.

Line tracking may break down due to highlights on horizontal structures, insufficient resolution on distant cars, and scale changes when the vehicle changes depth. The results may also contain random segments from static background such as long roofs, painted marks on road, and other building structure. Figure 8 shows a frame of line extraction overlapped with the intensity image. Different from many other methods on shape analysis in individual frames, this work relies on the motion properties of scenes, i.e., the scene positions corresponding motion behavior in the video. The continuity of vehicle's motion provides more stable observation than complex shapes of various vehicles, which will enhance the shape approaches.

#### B. Condensing Features to Temporal Profiles

To speed up the processing for real time target tracking and yield robust results from shaking vehicle and changeable road slope, we project the intensity/color  $I(x, y)$  and features in each video frame vertically to a 1D profile [7,8]. For the

color in each frame, we produce profile  $T(x)$  using the probability distribution calculated in Fig. 3. Consecutive profiles along the time axis generate a condensed spatial-temporal image,  $T(x,t)$ , used for analyzing the scene movement. The vertical projection of intensities by weight mask  $w(x,y)$  as  $p(x,y|C)$  is implemented as

$$T(x,t) = \sum_{y=-h/2}^{h/2} w(x,y)I(x,y,t) \quad (14)$$

where  $h$  is the image height. This profiling ignores most irrelevant features on the background. The traces in the condensed spatial-temporal image show movements of long or strong vertical lines in the video frames as in Figure 9a. The intensity profile shows the horizontal positions of vertical features and ignores the horizontal features at any object height. Slanted or short edges in individual frames will not be distinct in  $T(x,v)$  either. This produces a compact image for understanding the horizontal motions in the video and avoids mismatching in point tracking. The image changes caused by jitter in camera roll and tilt ( $R_z(t)$  and  $R_x(t)$  on uneven roads) are significantly reduced in the profiles. The background traces are in hyperbolas moving towards the sides of image.

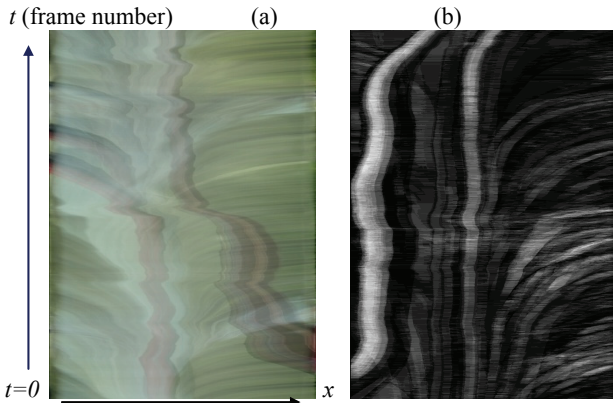


Fig. 9. Examples of profiles. (a) profile from intensity (of tree background and red cars), (b) profile from horizontal line segments.

In addition to the intensity profiling, we also profile lines extracted in the video to generate traces in the spatial-temporal images. We profile the number of horizontal line segments at each  $x$  position by

$$T_l(x,t) = \sum_{y=-h/2}^{h/2} w(x,y)C(x,y,t) \quad (15)$$

where  $C(x,y,t)$  takes value 1 on a horizontal line segment and 0 otherwise in frame  $t$ . Such a result can be found in Fig. 9b, where the bright stripes accumulated from many line segments show the motion of vehicles. Many long and horizontal lines in the images such as road paints and wires add equally to all  $x$  positions of the profile, which do not influence vehicle traces. Due to the existence of multiple horizontal lines on a vehicle, the vehicle position generally appears brighter than other positions of background.

*Instantaneous illumination changes* happen in cases such as entering a tunnel, penetrating shadows, and light up by other vehicles. Such changes alter the intensity of entire frame. In the spatial-temporal profiles, such illumination changes appear as an obvious horizontal line over the entire

image width. Its influence can be successfully filtered out by taking horizontal derivative of the profiles.

Compared to the intensity/color profiles with smooth traces, the profiling of horizontal line segments yields noisy traces; a line may fail to be extracted by fixed thresholds due to subtle intensity changes in consecutive frames. Because the lighting condition, background, and camera noise varied frame by frame, the line segments found in consecutive images do not always appear at the consistent positions and in the same length, which needs further processing according to their time coherence. Also, as is evident in Fig. 9b, many traces may not belong to a vehicle but to the background. Fortunately, the profiled numbers of horizontal lines is significantly less on background than on vehicles.

### C. Tracking Profiles for Motion Information

Tracking intensity profiles is done by checking  $\partial T(x,t)/\partial x$ . The edges are marked as  $E(x,t)$  and a maximum span for search is set for consecutive trace points as time goes on. At the same time,  $\partial T(x,t)/\partial t$  is also computed for confirming horizontal edges that connect fast moving traces in  $T(x,t)$ . For those very long horizontal segments in  $E(x,t)$ , they are mainly from the instantaneous illumination changes and are ignored in the result. This processing yields image position  $x(t)$  and horizontal image velocity  $v(t)=x(t)-x(t-1)$ . To preserve the motion continuity in tracking, we require  $|v(t)-v(t-1)| \rightarrow \min$ . Besides the requirement of high contrast, the sign of trace is also used as reference in tracking.

The horizontal line segment piles provide salient clue of vehicle presence. Tracking traces of line segments in  $T_l(x,t)$  is done by locating the center of each trace and following its movement continuously, because the margin points of each trace are usually unreliable. In implementation, we filter  $T_l(x,t)$  horizontally with a smoothing filter to obtain major traces as shown in Fig. 10. The traces are then marked at their peaks  $x(t)$  for the trace centers, if it is over a threshold determined adaptively. Line segments on background and instantaneous light changes over the entire frame (tail braking lights, police alarm lights, etc.) also yield long horizontal lines in  $T_l(x,t)$ . However, these lines are ignored in the processes of finding horizontal peaks (they have no

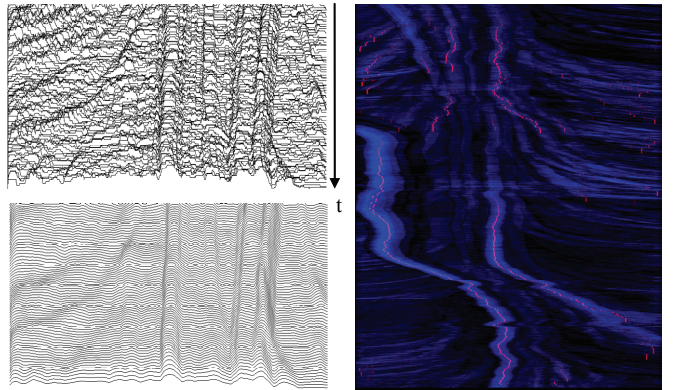


Fig. 10 Tracing centers of line segment clusters. (left) Profiled distributions of line segments before and after smoothing. (right) An example of extracted centers (in purple) overlapped with traces.

distinct peak) and tracking them over long periods of time.

#### IV. COMPUTING POSTERIOR PROBABILITY IN HMM

As we obtain observation  $(x(t), v(t))$  on each trace dynamically, we estimate the identities of features as cars or background using the HMM based on their motion behaviors. The posterior probabilities are denoted by  $P(C_t|x(t), v(t))$  and  $P(B_t|x(t), v(t))$  respectively, or  $P(C_t)$  and  $P(B_t)$  for short. As parameters of the HMM, we keep

$$P(C_t) + P(B_t) = 1 \quad (16)$$

at any time  $t$ . The probabilities of state transition from frame  $t-1$  to frame  $t$  are defined empirically as

$$\begin{aligned} P(C_t | B_{t-1}) &= 0.5 & P(B_t | B_{t-1}) &= 0.5 \\ P(C_t | C_{t-1}) &= 0.8 & P(B_t | C_{t-1}) &= 0.2 \end{aligned} \quad (17)$$

The transition from car to car is set as 0.8 to emphasize the continuity of car motion. A background may be identified as a car later, i.e.,  $P(C_t | B_{t-1})=0.5$ , since there may not have strong clue to classify a car near the FOE in the frame, where both background and vehicles have small image velocities. When a trajectory is initially detected ( $t=0$ ), its probabilities as a car and background are set empirically as  $P(C_0)=0.7$  and  $P(B_0)=0.3$  according to our ability to identify vehicles using lines in the weighted image area.

Using Viterbi algorithm [24,26], the probability of a trace to be assigned as car at time  $t$  is optimized as

$$P(C_t) = \max[ P(B_{t-1}) \times P(C_t | B_{t-1}) \times p(x(t), v(t) | C_t), \\ P(C_{t-1}) \times P(C_t | C_{t-1}) \times p(x(t), v(t) | C_t) ] \quad (18)$$

And the probability as background is

$$P(B_t) = \max[ P(B_{t-1}) \times P(B_t | B_{t-1}) \times p(x(t), v(t) | B_t), \\ P(C_{t-1}) \times P(B_t | C_{t-1}) \times p(x(t), v(t) | B_t) ] \quad (19)$$

using the likelihoods calculated in Section 2. If  $P(C_t) > P(B_t)$ , the trace is considered as a car at time  $t$ , and as background otherwise. The identity is reported after the trace is tracked over a minimum duration of time. Otherwise, such a short trace is removed as noise; we assume that a target vehicle will not vanish from the field of view rapidly.

As we track all the traces in the profiles during the vehicle motion, we apply the HMM on each trace to update its state, i.e., car or background. At every  $t$ , the obtained probabilities  $P(B|x, v)$  and  $P(C|x, v)$  are normalized by

$$P(C_t) \leftarrow \frac{P(C_t)}{P(C_t) + P(B_t)} \quad P(B_t) \leftarrow \frac{P(B_t)}{P(C_t) + P(B_t)} \quad (20)$$

in order to avoid a quick decreasing of  $P(C_t)$  and  $P(B_t)$  values to zero in Eq. 18 and 19. The processing is in real time as the vehicle moves. If a new trace is found, we assemble a new HMM. The calculated traces may have low certainty levels at beginning due to lack of evidence in short periods of time. The probability will get high as the trace is constantly tracked and updated.

#### V. EXPERIMENTS

Because of the probability distribution of scenes introduced to tolerate variations of targets, the precision of feature locations becomes less critical to the results. We therefore omit serious camera calibration by indicating the image position of forward direction, horizon, and the visible portion in the video frame in advance. These parameters are invariant to the dynamic scenes during vehicle travelling.

TABLE I  
PARAMETERS USED IN COMPUTING LIKELIHOODS

Symbol	Quantity	Symbol	QUANTITY
D	6m	f	900 pixel
F	10m	$\int_z$	0~200m
$\sigma_F$	20m	$\int_x$	-50~50m
$\sigma_x$	6m/s	$\int_{Tz}$	-40~40m/s
$\sigma_z$	10m/s	H	4m
$S, \sigma$	15m/s, 5m/s	$\sigma_r$	5 degree/s

We have examined our method on videos collected from vehicle borne cameras. The sample sets contain eight video shots lasting 1 hour and 20 minutes on rural and urban roads with various lighting conditions. The videos consist of both night-time and day-time clips. We implemented our method in Visual C++ environment and use OpenCV for input of AVI video and differentiation. The likelihood maps are initially computed once using MatLab.

The probability distributions defined for environment and vehicles are mostly Gaussian distributions with large variances so that they can work on almost all possible video shots and vehicle appearance covering three lanes in width. Table I gives several parameters involved in computing likelihoods. All the likelihood distributions are registered in look-up tables for real time referring by the HMM. Although it is derived from a horizontal road, it is also applicable to a surface segment on non-horizontal roads because any road has a mild slope change locally according to the road regulation.

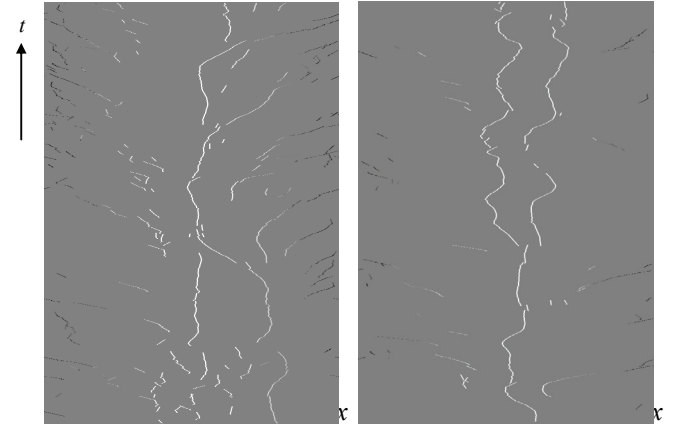


Fig. 11. Traces displayed with their probabilities in intensity. White and black indicate the possibilities of a trace to be vehicle or background respectively. Grey is uncertain trace with probability 0.5 or has no feature.

Figure 11 shows the identification of tracked traces in the profiles. The probability as vehicle or background is displayed in intensity. Traces become certain after they demonstrate their motion. The existing of vehicles are clearly displayed even they are occluded by other vehicles shortly. This result provides probability to the scene identification. Multiple cars are tracked with high probabilities. Background traces are also identified after they display their behaviors moving outward. Figure 12 shows examples of identified cars. After vehicles are detected, boxes are located onto them by searching vertically the car positions. The width of a box is refined with the profiles of intensity where clear sides are visible around the traces.

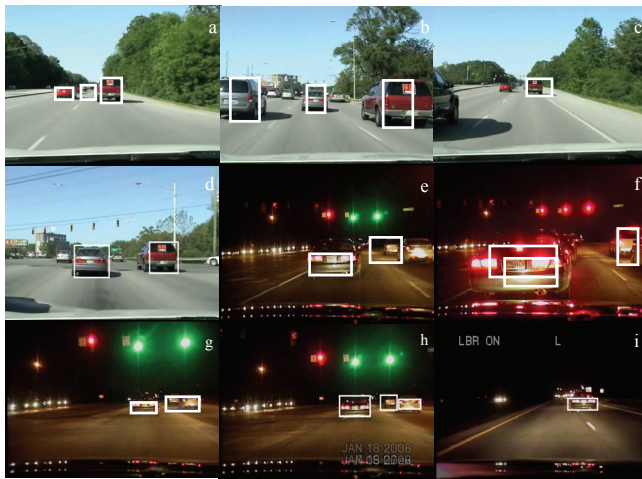


Fig. 12. Vehicle detection and tracking results in car videos.

The response of a target vehicle is after 5 frames, though the probability may not be certain at beginning. Tracking in profiles is more stable than tracking point in the video so that obtained probability is less bothered by matching mistakes and it is increasing as the trace gets long.

The confusion matrix tested from our video sequences is shown in Table II. In most cases of correct detection, the duration of the tracking lasts as long as the vehicle remains in the field of view. If a detected vehicle moves too far from the camera to be detected, the tracking is stopped due to its small size to be extracted as a trace. Very slow or stopped cars in the same direction and the vehicles on the opposite side of road are not classified as the target vehicles along the same direction, because their motion behaviors are close to that of background. Coming cars at opposite lanes are ignored if they are distant; their images are small and are not accumulated to traces. If a coming car is right at next lane of the camera, it is classified as an obstacle with a high speed.

TABLE II  
IDENTIFICATION OF VEHICLES IN CONFUSION MATRIX

Prediction	Car	Background
Car	True Positive 86.6%	False Positive 13.2%
Background	False Negative 14.1%	True Negative 85.9%

## VI. CONCLUSION

This work detects and tracks vehicles ahead with an in-car video camera. Our approach uses motion information of vehicles and background described by probability model, which is general for all vehicles and is more effective than shape analysis in identification. Features that characterize the target vehicles are extracted in each frame and their positions are profiled to yield a compact spatial-temporal image for fast vehicle tracking and robust identification. We use the HMM in identification during tracking to yield a probability framework that is less influenced by ad-hoc thresholds. Image positions and velocities of features are calculated in the HMM. Experiment results purely based on the motion of features show the effectiveness of the method. The computation is implemented in real time during the vehicle motion and is easy to be embedded into hardware.

## ACKNOWLEDGMENT

The authors would like to thank Mr. Herbert Blitzer at Indiana Forensic Institute for the support to this work.

## REFERENCES

- [1] C. Hoffman, T. Dang, C. Stiller, Vehicle detection fusing 2D visual features, IEEE Intelligent Vehicle Symposium, 280-285, 2004.
- [2] D. Alonso, L. Salgado, M. Nieto, Robust Vehicle detection through multidimensional classification for on board video based systems IEEE ICIP07, 4, 321-324, 2007.
- [3] P. Parodi, G. Piccioli, A feature-based recognition scheme for traffic scenes, IEEE IV 95, 229-234, 1995.
- [4] C. R. Wang, J. Lien, Automatic vehicle detection using local features — a statistical approach, IEEE Trans. ITS, 9(1), 83-96, 2008.
- [5] J. Dubuisson, S. Lakshmanan, A. K. Jain, Vehicle segmentation and classification using deformable templates, IEEE PAMI 18(3), 293-308, 1996.
- [6] G. D. Sullivan, et al, Model-based vehicle detection and classification using orthographic approximations. Image and Vision Computing, 15 (8), 649-654, 1997.
- [7] J. Y. Zheng, Y. Bhupalam, H. Tanaka, Understanding vehicle motion via spatial integration of intensities”, 19th ICPR, 1-5, 2008
- [8] G. Flora, J. Y. Zheng, Adjusting route panoramas with condensed image slices, ACM Conf. Multimedia 07, 815-818, 2007
- [9] M. Betke, H. Nguyen, Highway scene analysis from a moving vehicle under reduced visibility conditions, IEEE IV 98, 131-136, 1998.
- [10] David Stirzaker (2003). *Elementary Probability*. Cambridge.
- [11] R. Lakaemper, S. S. Li, M. Sobel, Correspondences of point sets using Particle Filters, ICPR08.
- [12] H. Schneiderman, T. Kanade, A statistical method for 3D object detection applied to faces and cars, IEEE CVPR, 1746-1759.
- [13] H. Bay, T. Tuytelaars, L.V. Gool, SURF: Speeded Up Robust Features, ICCV, May 2006.
- [14] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE PAMI, vol. 24, no. 5, 603-619, 2002.
- [15] O. R. Duda, E. H. Hart, G. D. Stork, (2000). Pattern Classification, second edition. New York, NY: Wiley- Interscience Publication.
- [16] C. Démonceaus, A. Potelle, D. Kachi-Akkouche, Obstacle detection in a road scene based on motion analysis, IEEE Trans. On Vehicular Technology, 53(6), 1649-1656, Nov. 2004.
- [17] Z. F. Zhu, H. Lu, J. Hu, K. Uchimura, Car detection based on multiscenes integration. 17<sup>th</sup> ICPR04, Vol. 2, 699 – 702, 2004.
- [18] T. N. Tan, K. D. Baker, Efficient image gradient based vehicle localization, IEEE Trans. Image Processing, 9(8), 1343-1356, 2000.
- [19] M. Kagesawa, S. Ueno, K. Ikeuchi, H. Kashiwagi, Recognizing vehicles in infrared images using IMAP parallel vision board, IEEE Trans. ITS, 2(1), 10-17, 2001
- [20] J. Kato, T. Watanabe, S. Joga, J. Rittscher, A. Blake, An HMM-based segmentation method for traffic monitoring movies, IEEE PAMI, 24(9), 1291- 1296, 2002
- [21] W. Zhang, X. Z. Fang, X. K. Yang, Moving vehicles segmentation based on Bayesian framework for Gaussian motion model, Pattern Recognition Letter, 27(9), 956-967, 2006.
- [22] J. Chu, L. Ji, L. Guo, R. Wang, Study on method of detecting preceding vehicle based on monocular camera, IEEE IV 04, 750- 755.
- [23] L. Gao, C. Li, T. Fang, Z. Xiong, Vehicle detection based on color and edge information, LNCS, Image Analysis and Recognition, Vol. 5112, 1611-3349, 2008.
- [24] X. Huang, A. Acero, and H.-W. Hon (2001). *Spoken Language Processing*. Prentice Hall. ISBN -013-022616-5.
- [25] H. Takizawa, K. Yamada, T. Ito, Vehicles detection using sensor fusion, IEEE Intelligent Vehicle Symposium, 238- 243, 2004.
- [26] G. D. Forney, The Viterbi algorithm. Proceedings of the IEEE 61 (3): 268-278, March 1973.
- [27] H. Cheng, N. Zheng, C. Sun, Boosted Gabor Features applied to vehicle detection, 18<sup>th</sup> ICPR, Vol. 2, 662-666, 2006.
- [28] A. Jazayeri, H. Cai, J. Y. Zheng, M. Tuceryan, and H. Blitzer, An intelligent video system for vehicle localization and tracking in police cars, ACM SAC 939-940, 2009.
- [29] A. Bruss, B. K. P. Horn, Passive navigation, CVGIP, 1983.